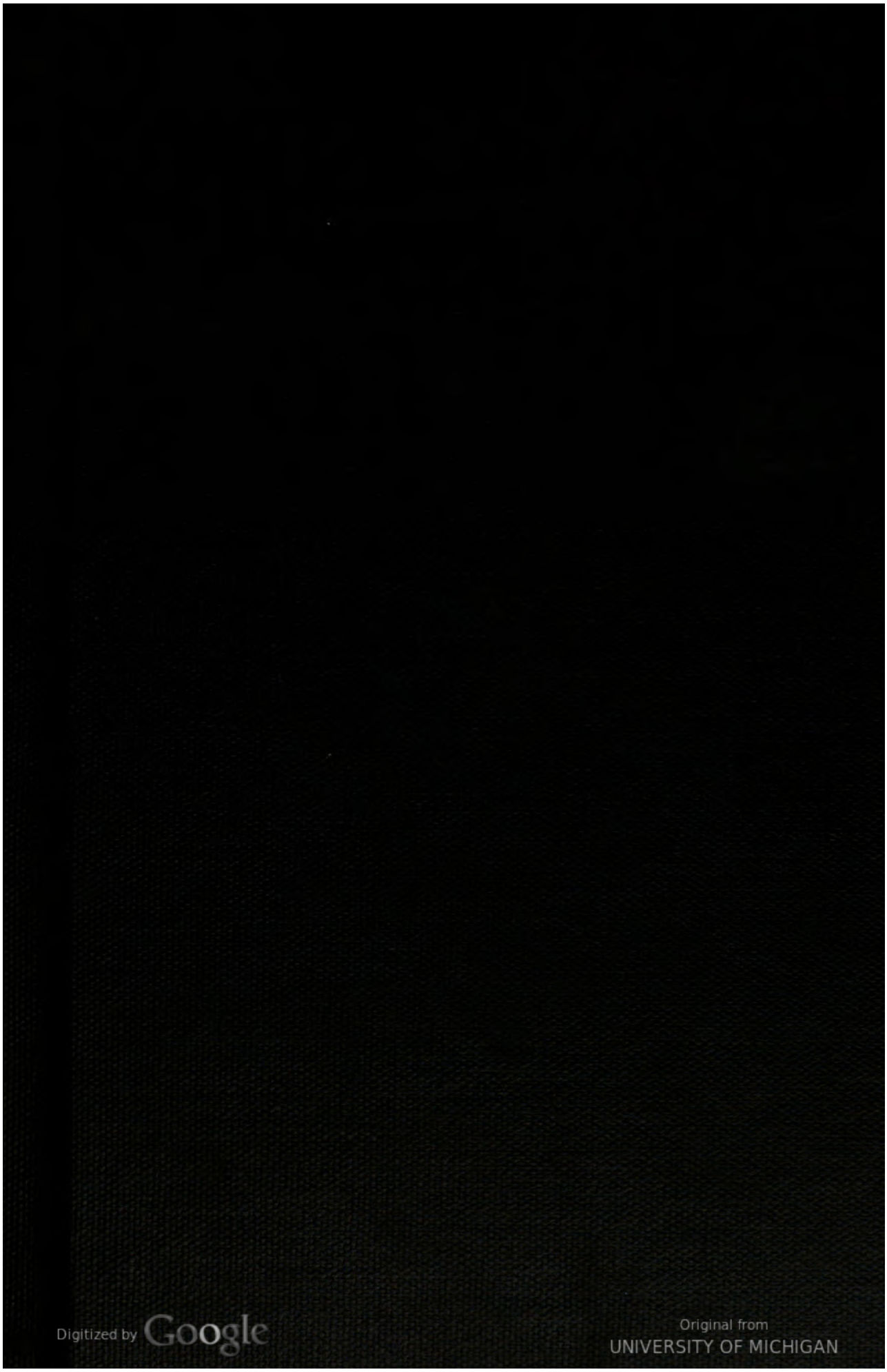
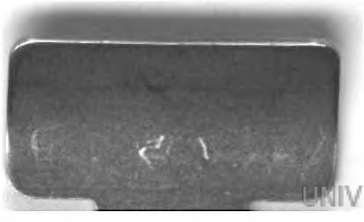




Groups, Rings, Modules

Harper's Series in Modern Mathematics
I. N. Herstein and Gian-Carlo Rota, Editors

Differential Geometry Louis Auslander
Groups, Rings, Modules Maurice Auslander and David A. Buchsbaum
Abstract Methods in Partial Differential Equations Robert W. Carroll
The Calculus: An Introduction Casper Goffman
Calculus of Several Variables Casper Goffman
Introduction to Real Analysis Casper Goffman
Systems of Ordinary Differential Equations: An Introduction
Jack L. Goldberg and Arthur J. Schwartz
Constructive Real Analysis A. A. Goldstein
Finite Groups Daniel Gorenstein
Introduction to Calculus Donald Greenspan
Elementary Probability Theory Melvin Hausner
Introduction to Linear Algebra Peter Kahn
A Course in Numerical Analysis H. Melvin Lieberstein
Cohomology Operations and Applications in Homotopy Theory
Robert E. Mosher and Martin C. Tangora
Ordered Topological Vector Spaces Anthony L. Peressini
Modern Probability Theory Rui Zong Yeh

GROUPS, RINGS, MODULES

MAURICE AUSLANDER
DAVID A. BUCHSBAUM
Brandeis University

HARPER & ROW, PUBLISHERS
New York, Evanston, San Francisco, London

Dedicated to our parents

Sponsoring Editor: George J. Telecki
Project Editor: Lois Wernick
Designer: T. R. Funderburk
Production Supervisor: Bernice Krawczyk

GROUPS, RINGS, MODULES

Library of Congress Cataloging in Publication Data

Auslander, Maurice.
Groups, rings, modules.
(Harper's series in modern mathematics)
1. Commutative rings. 2. Modules (Algebra)
3. Groups, Theory of. I. Buchsbaum, David Alvin
joint author. II. Title.
QA251.3.A93 512'.4 73-13199
ISBN 0-06-040387-X

CONTENTS

PREFACE

The main thrust of this book is easily described. It is to introduce the reader who already has some familiarity with the basic notions of sets, groups, rings, and vector spaces to the study of rings by means of their module theory. This program is carried out in a systematic way for the classically important semisimple rings, principal ideal domains, and Dedekind domains. The proofs of the well-known basic properties of these traditionally important rings have been designed to emphasize general concepts and techniques. Hopefully this will give the reader a good introduction to the unifying methods currently being developed in ring theory.

Part I is a potpourri of background material, much of which is undoubtedly familiar to the reader, some of which is probably new. In addition to the usual notions of sets, monoids, and groups, heavy emphasis is put on maps and morphisms of monoids and groups. This naturally leads to the notion of a category, which is briefly discussed in Chapter 3. In Chapter 4, the notions already developed for sets, monoids, and groups are applied to a preliminary discussion of the category of rings. Chapter 5 is far less formal. It is devoted to the study of unique factorization in arbitrary commutative domains. Here the principal novelties are the heavy use of localization in commutative domains and the introduction of chain conditions for ideals.

Part Two begins with a lengthy discussion of modules over general rings. Starting from the notion of a basis for vector spaces, we develop free modules as well as the general notion of sums and products in the category of modules over a ring. Among other things, it is shown that a ring R is a division ring if and only if

every R-module is free. This is the first step of our general program of studying rings by means of their modules. Although Chapter 6 is too long to describe in further detail, we caution the reader that familiarity with the contents of this chapter is essential to the understanding of the rest of the book.

The remainder of Part Two is devoted to the next step of our program of studying rings by means of their modules. Namely, it is shown that a ring is semisimple if and only if its modules are semisimple. In this context, projective modules arise naturally. So, also, does the notion of the radical of a ring. Although this part of the book is devoted mainly to semisimple rings, some fundamental facts are developed for general artin rings in the text as well as in the exercises.

The rest of the book is devoted almost exclusively to commutative rings. Since localization and tensor products play such important roles in this theory, Part Three starts with a discussion of these techniques. This is then followed by the study of principal ideal domains. These rings are characterized as commutative rings R with the property that submodules of free R-modules are free. Thus, they arise naturally as the next step in our program of studying rings by means of their modules. In describing the structure of finitely generated modules over principal ideal domains, injective modules are introduced. Part Three ends with applications of this structure theory to the study of endomorphisms of finite-dimensional vector spaces. Included are such standard items as canonical forms of matrices and determinants.

The final part of the book is devoted to algebraic extensions of fields and the study of integral extensions of noetherian domains. The major aim of Chapter 12 is to develop finite galois theory of fields. This theory is used to study integral extensions of noetherian domains which leads to the theory of Dedekind domains. As part of our general module theoretic point of view, we characterize Dedekind domains as those integral domains having the property that submodules of projective modules are projective. The book ends with a description of the ideals in Dedekind domains and the structure theorem for finitely generated modules over such domains.

We recommend that the reader have pencil and paper close at hand when reading the text. Proofs for many assertions have been omitted. The reader will be able to supply the missing steps or proofs either by himself or after consulting outlines given in the exercises. In addition to exercises explaining the text, there are exercises dealing with related but supplementary material.

The partitioning of the book was done on pedagogical as well as logical grounds. Part One can be used for a leisurely one-semester course on the fundamental structures of algebra. Parts Two and Three can serve as a one-semester introduction to general ring theory for more advanced students. For students familiar with Chapters 1, 2, and 4, the entire text should constitute a full year course in algebra.

We thank our publishers, Harper & Row, for their patience during the preparation of the manuscript.

M. A.
D. A. B.

PART ONE

Chapter 1 SETS AND MAPS

INTRODUCTION

This chapter and the next are devoted to a review of the basic concepts of set and group theory. Because we are assuming the reader already has some familiarity with these topics, our exposition is neither systematic nor complete. Only a brief description of the basic concepts and results that are needed in the rest of this book is presented.

This should serve to give the reader some idea of the mathematical background we are assuming as well as help fix conventions and notations for the rest of the book. Although few proofs are given, outlines of proofs of the less obvious results cited in the text are given in the exercises. It is hoped that the reader will find completing these outlines a useful way of familiarizing himself with any new concepts or results he may encounter in this or the next chapter.

1. SETS AND SUBSETS

We take a naive, nonaxiomatic view of set theory. We view a set as an actual collection of things called the elements of the set. We will often denote the fact that x is an element of the set X by writing $x \in X$. From this point of view it is obvious that two sets are the same if and only if they have the same elements. Or stated more precisely, two sets X and Y are the same if and only if both of the

following statements are true:

(a) If $x \in X$, then $x \in Y$.
(b) If $y \in Y$, then $y \in X$.

In this connection, we remind the reader that in mathematical usage, a statement of the form "If A, then B" is true unless A is true and B is false, in which case it is false. In particular, if A is false, then the statement "If A, then B" is true independent of whether B is true or false. To illustrate this point we show that there is only one empty set.

We recall that a set X is said to be empty if X has no elements; or more precisely, if the statement "$x \in X$" is always false. Suppose now that the sets X and Y are empty. Then both of the statements "$x \in X$" and "$y \in Y$" are always false. Hence, by our convention concerning sentences of the form "If A, then B," both of the statements

(a) If $x \in X$, then $x \in Y$;
(b) If $y \in Y$, then $y \in X$;

are true. This shows that if the sets X and Y are both empty, then $X = Y$. Following the usual conventions of set theory, we assume that there is an empty set. This uniquely determined set will usually be denoted by $\emptyset$.

An important set associated with a set X is the **power set** 2^X of X which we will define once we have recalled the notion of a subset of a set.

A set Y is said to be a subset of a set X if every element of Y is also an element of X, or equivalently, the set Y is a subset of the set X if and only if the statement "If $y \in Y$, then $y \in X$" is true. The fact that Y is a subset of X is often denoted by $Y \subset X$, which is sometimes also read as "Y is contained in X."

One easily verified consequence of this definition is that if X is any set, then the empty set $\emptyset$ is a subset of X. For the statement "If $x \in \emptyset$, then $x \in X$," is true for any set X because the statement "$x \in \emptyset$" is always false. Also associated with an element x of X is the subset $\{x\}$ of X consisting precisely of the element x of X. Further, the reader should have no difficulty verifying the following.

Basic Properties 1.1
Let X, Y, and Z be sets. Then:

(a) $X \subset X$.
(b) $X = Y$ if and only if $X \subset Y$ and $Y \subset X$.
(c) If $X \subset Y$ and $Y \subset Z$, then $X \subset Z$.

We are now in a position to define the power set 2^X of a set X. The set 2^X is the set whose elements are precisely the subsets of X. Stated symbolically, the power set 2^X of a set X is the set with the property that $Y \in 2^X$ if and only if $Y \subset X$.

It is worth noting that 2^X is never empty, even if X is empty. This is because the empty set $\emptyset$ is always contained in X and is thus an element of 2^X. Also, as we have already observed, there is associated with each element x of X the element $\{x\}$ of 2^X. Hence, 2^X consists of a single element if and only if X is empty.

We now recall the familiar notions of union and intersection of sets. Suppose X is a set and $\mathcal{S}$ a subset of 2^X. The **intersection** of the subsets of X in $\mathcal{S}$ is the

subset $\bigcap_{X' \in \mathcal{S}} X'$ of X consisting of all x in X such that the statement "If $X' \in \mathcal{S}$, then $x \in X'$" is true. It should be noted that if the subset $\mathcal{S}$ of 2^X is empty, then $\bigcap_{X' \in \mathcal{S}} X' = X$. For if $\mathcal{S}$ is empty, then the statement "If $X' \in \mathcal{S}$, then $x \in X'$" is true for all x in X since the statement "$X' \in \mathcal{S}$" is false.

The **union** of the subsets of X in $\mathcal{S}$ is the subset $\bigcup_{X' \in \mathcal{S}} X'$ of X consisting of all x in X with the property that the statement "There is an $X' \in \mathcal{S}$ such that $x \in X'$" is true. It should be noted that if $\mathcal{S}$ is empty, then $\bigcup_{x' \in \mathcal{S}} X' = \emptyset$. For if $\mathcal{S}$ is empty, then the statement "There is an $X' \in \mathcal{S}$ such that $x \in X'$" is false for all $x \in X$ since there are no X' in 2^X satisfying the condition that X' is in $\mathcal{S}$.

In practice, a particularly useful way of studying a set is to represent it as a union of some of its subsets. For this reason it is convenient to make the following definition.

Definition
Suppose X is a set. A subset $\mathcal{C}$ of 2^X is called a **covering of X** if $X = \bigcup_{X' \in \mathcal{C}} X'$.

Although coverings of various types play an important role in all of mathematics, we will be particularly concerned with the type of coverings called partitions.

Definition
A covering $\mathcal{C}$ of a set X is said to be a **partition of X** provided:

(a) If $X' \in \mathcal{C}$, then $X' \neq \emptyset$.
(b) If X' and X'' are distinct elements of $\mathcal{C}$, then $X' \cap X'' = \emptyset$.

The reader should convince himself that a set $\mathcal{C}$ of nonempty subsets of a set X is a partition of X if and only if each element in X is in one and only one subset of X in $\mathcal{C}$. For this reason, if $\mathcal{C}$ is a partition of a set X, it makes sense to talk about the element of $\mathcal{C}$ containing a particular element x of X. We will usually denote by $[x]_{\mathcal{C}}$ the unique element of the partition $\mathcal{C}$ of X containing the element x of X. When there is no danger of ambiguity concerning the particular partition $\mathcal{C}$ of a set X, we will write $[x]$ for $[x]_{\mathcal{C}}$.

Finally, we recall what is meant by the **product** $X \times Y$ of two sets X and Y. The set $X \times Y$ consists of all symbols (x, y) with x an element of X and y an element of Y. Hence, two elements (x, y) and (x', y') in $X \times Y$ are the same if and only if $x = x'$ and $y = y'$. Obviously, $X \times Y$ is empty if and only if either X or Y is empty.

2. MAPS

A **map of sets** consists of three things: a set X called the **domain** of the map, a set Y called the **range** of the map, and a subset f of $X \times Y$ having the property that if x is in X, then there is a unique y in Y such that the element (x, y) in $X \times Y$ is in f. These data X, Y, f will be denoted by $f: X \to Y$ which is to be read as "f is a map

from X to Y." If x is in X, then the unique element y in Y such that (x, y) is in f is called the **value of the map f at x** and is denoted by $f(x)$.

It is important to observe that according to this definition two maps cannot be the same unless they have the same domain and range. Also, two maps $f: X \to Y$ and $g: X \to Y$ with the same domains and ranges are the same if and only if their values are the same for each x in X, that is, if and only if $f(x) = g(x)$ for all x in X. Thus, once having specified the domain and range of a map, it only remains to describe its values for each x in X in order to completely determine the map. In the future, when defining particular maps from a set X to a set Y, we shall generally describe them by prescribing their values for each x in X rather than by writing down a subset of $X \times Y$. In following this procedure it is of course necessary to make sure that one and only one value in the range has been assigned to each element of the domain. As an illustration of this point suppose that $\mathscr{C}$ is a partition of a set X. Then we have already seen that for each x in X there is one and only one element $[x]_{\mathscr{C}}$ of $\mathscr{C}$ containing x. Thus, we obtain a map $k_{\mathscr{C}}: X \to \mathscr{C}$ by setting $k_{\mathscr{C}}(x) = [x]_{\mathscr{C}}$. Of course, we could have also defined the map $k_{\mathscr{C}}$ as the subset of $X \times \mathscr{C}$ consisting of all elements $(x, [x]_{\mathscr{C}})$ in $X \times \mathscr{C}$ with x in X.

We now describe some important maps of sets.

Example 2.1 Suppose $f: X \to Y$ is a map and X' is a subset of X. We define a map $f|X': X' \to Y$ called the **restriction of f to X'** by $(f|X')(x') = f(x')$ for all x' in X'.

Example 2.2 Associated with each subset X' of a set X is the **inclusion map** from X' to X which is denoted by $\text{inc}: X' \to X$ and is defined by $\text{inc}(x) = x$ if x is an element of X which is in X'.

Example 2.3 The inclusion map of a set X to itself is called the **identity map** and is usually denoted by id_X for each set X.

Example 2.4 Since the empty set $\emptyset$ is a subset of any set X, we always have the inclusion map $\text{inc}: \emptyset \to X$. Actually, this is the only map from $\emptyset$ to X and this unique map from $\emptyset$ to a set X is called the **empty map**. In this connection, the reader should convince himself that there are no maps from a nonempty set to the empty set.

Example 2.5 We have already seen that associated with a partition $\mathscr{C}$ of a set X is the map $k_{\mathscr{C}}: X \to \mathscr{C}$ given by $k_{\mathscr{C}}(x) = [x]$ for each x in X where $[x]$ is the unique subset of X in $\mathscr{C}$ containing the element x. This map $k_{\mathscr{C}}: X \to \mathscr{C}$ is called the **canonical** or **natural** map from the set X to the partition $\mathscr{C}$.

Suppose X and Y are sets. Then each map with domain X and range Y is completely determined by a subset of $X \times Y$ and hence by an element of $2^{X \times Y}$. Thus, the collection of all maps from X to Y which we denote by (X, Y) is a set which is a subset of $2^{X \times Y}$.

Of fundamental importance in constructing and analyzing maps is the notion of the **composition of maps**. Given two maps $f: X \to Y$ and $g: Y \to Z$ with the range of f the same as the domain of g, we define their composition gf to be the map $gf: X \to Z$ given by $gf(x) = g(f(x))$ for each x in X. It follows immediately from this definition that if we are given three maps $f: U \to X$, $g: X \to Y$, and $h: Y \to Z$, then the two maps $h(gf): U \to Z$ and $(hg)f: U \to Z$ are the same. This property of

the composition of maps is referred to as the **associativity of the composition of maps.**

As an example of the composition of maps we point out that if $f: X \to Y$ is a map of sets and X' is a subset of X, then $f|X': X' \to Y$, the restriction of f to X', is the composition $X' \xrightarrow{\text{inc}} X \xrightarrow{f} Y$ where $\text{inc}: X' \to X$ is the inclusion map.

3. ISOMORPHISMS OF SETS

One of the most important problems in mathematics is deciding when two mathematical objects have the same or similar mathematical properties and can therefore be considered essentially the same. Since all the mathematical objects we will be considering in this book consist of an underlying set together with some additional structure, it is reasonable to first consider how sets are compared and the circumstances under which they are considered essentially the same.

Because a map from a set X to a set Y associates with each element x in X an element y in Y, a map clearly can be viewed as a method for comparing the sets X and Y. If this is a reasonable idea, then we should be able to state in terms of maps what is probably the simplest comparison of sets we can make: the fact that a set is the same as itself. The reader should have no difficulty convincing himself that the identity map on a set does indeed express this fact. It is interesting to note that the identity map on a set can be completely described in terms of maps as is done in the following.

Basic Property 3.1
For a map $f: X \to X$, the following statements are equivalent:

(a) $f = \text{id}_X$.
(b) Given any map $g: X \to Y$, then $gf = g$.
(c) Given any map $h: Y \to X$, then $fh = h$.

Having decided that the identity map expresses the fact that a set is the same as itself, it is reasonable to ask what kind of maps between two sets X and Y must exist in order to conclude that X and Y resemble each other as much as possible. In view of our previous discussion, this amounts to asking when is a map $f: X \to Y$ close to being an identity map? A possible answer might be that there is a map $g: Y \to X$ such that the composition $gf: X \to X$ is the identity on X. But there is no reason to favor the set X over the set Y. Hence, we should also require that there be a map $h: Y \to X$ such that $fh = \text{id}_Y$. However, the associativity of the composition of maps implies that under these circumstances the two maps g and h are the same. Therefore, it seems reasonable to consider two sets X and Y as being essentially the same if there exists a pair of maps $f: X \to Y$ and $g: Y \to X$ such that $gf = \text{id}_X$ and $fg = \text{id}_Y$. In fact, this amounts to nothing more than the familiar notion of two sets being isomorphic, as we see in the following.

Definition
Let X and Y be sets. A map $f: X \to Y$ is said to be an **isomorphism** if and only if there is a map $g: Y \to X$ such that $gf = \text{id}_X$ and $fg = \text{id}_Y$. If $f: X \to Y$ is an isomorphism, then there is only one map $g: Y \to X$ with these properties, and this

uniquely determined map from Y to X, which we denote by f^{-1}, is called the **inverse of** f. Finally, the set X is said to be **isomorphic to** Y if there is a map $f: X \to Y$ which is an isomorphism.

We remind the reader of the following.

Basic Properties 3.2

(a) All identity maps are isomorphisms which are their own inverses. Hence, all sets are isomorphic to themselves.
(b) If $f: X \to Y$ is an isomorphism, then the inverse $f^{-1}: Y \to X$ is also an isomorphism whose inverse is f, that is, $(f^{-1})^{-1} = f$. Hence, if X is isomorphic to Y, then Y is isomorphic to X.
(c) The composition gf of two isomorphisms g and f is also an isomorphism with inverse $f^{-1}g^{-1}$. Thus, if X is isomorphic to Y and Y is isomorphic to Z, then X is isomorphic to Z.
(d) If gf is an isomorphism, then g is an isomorphism if and only if f is an isomorphism.

Experience has shown, roughly speaking, that a map $f: X \to Y$ is an isomorphism if and only if it gives a way of identifying the set X with the set Y. A precise formulation of this idea is given in the following familiar characterization of isomorphisms.

Basic Property 3.3
A map $f: X \to Y$ is an isomorphism if and only if it satisfies both of the following conditions:

(a) If $y \in Y$, then there is an x in X such that $f(x) = y$.
(b) If x_1 and x_2 are in X and $f(x_1) = f(x_2)$, then $x_1 = x_2$.

Although technically equivalent to the notion of an isomorphism, the conditions (a) and (b) of the above basic property are conceptually quite different from our original definition of an isomorphism since these conditions describe what the map does to the elements of the sets involved rather than how it is related to other maps. We will often refer to an isomorphism as a **bijective map** when we wish to emphasize this different approach to the concept of an isomorphism of sets.

4. EPIMORPHISMS AND MONOMORPHISMS

Yet another aspect of the notion of an isomorphism of sets is given in the following.

Basic Property 4.1
A map $f: X \to Y$ which is an isomorphism satisfies the following conditions:

(a) If $g_1, g_2: Y \to Z$ are two maps such that $g_1 f = g_2 f$, then $g_1 = g_2$.
(b) If $h_1, h_2: U \to X$ are two maps such that $fh_1 = fh_2$, then $h_1 = h_2$.

It turns out that there are many important maps which satisfy one but not necessarily both of the above conditions. For this reason we make the following definitions.

Definitions
Let $f: X \to Y$ be a map.

(a) f is called an **epimorphism** if given two maps $g_1, g_2: Y \to Z$, we have $g_1 = g_2$ whenever $g_1 f = g_2 f$.
(b) f is called a **monomorphism** if given two maps $h_1, h_2: U \to X$, we have $h_1 = h_2$ whenever $f h_1 = f h_2$.

Thus, if a map is an isomorphism, it is both an epimorphism and a monomorphism.

We now list some easily verified properties of epimorphisms and monomorphisms.

Basic Properties 4.2
Let $f: X \to Y$ and $g: Y \to Z$ be two maps.

(a) If f and g are both epimorphisms (monomorphisms), then the composition $gf: X \to Z$ is an epimorphism (monomorphism).
(b) If $gf: X \to Z$ is an epimorphism, then so is g.
(c) If $gf: X \to Z$ is a monomorphism, then so is f.

We have already seen how to describe in terms of what a map does to elements the fact that it is an isomorphism. The same can be done for the notions of epimorphisms and monomorphisms. In order to state this result, it is convenient to have the following.

Definitions
Let $f: X \to Y$ be a map.

(a) f is said to be a **surjection**, or a **surjective map**, if for each y in Y there is an x in X such that $f(x) = y$.
(b) f is said to be an **injection**, or an **injective map**, if given x_1 and x_2 in X with the property that $f(x_1) = f(x_2)$, then $x_1 = x_2$.

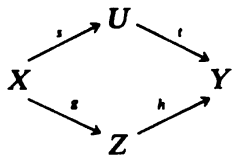
Basic Properties 4.3

(a) A map is an epimorphism if and only if it is a surjective map.
(b) A map is a monomorphism if and only if it is an injective map.
(c) A map is an isomorphism if and only if it is an epimorphism and a monomorphism.

As with isomorphisms, we will refer to an epimorphism (monomorphism) as a surjective map (injective map) whenever we wish to emphasize what the map does to the elements of the sets involved rather than its relation to other maps.

We conclude this section with the following useful property of surjective and injective maps.

Proposition 4.4
Suppose we are given a diagram of maps of sets



satisfying:

(a) $ts = hg$.
(b) s is a surjective map and h is an injective map.

Then there is one and only one map $j: U \to Z$ such that $js = g$ and $hj = t$.

5. THE IMAGE ANALYSIS OF A MAP

A map $f: X \to Y$ of sets not only serves as a way of comparing the sets X and Y, but also as a way of comparing subsets of X and subsets of Y. In the following definitions we point out some of these relationships. Others will be discussed later on.

Definitions
Suppose $f: X \to Y$ is a map of sets. If X' is a subset of X, then the subset of Y consisting of all elements $f(x)$ in Y with x in X' is called the **image of X' under f** and is denoted by $f(X')$. The subset $f(X)$ of Y is called the **image of the map f** and is usually denoted by $\operatorname{Im} f$.

Suppose we are given a map $f: X \to Y$. It is clear that f is a surjective map, or equivalently an epimorphism, if and only if $\operatorname{Im} f = Y$. However, regardless of whether the map f itself is surjective, the map $f_0: X \to \operatorname{Im} f$, defined by $f_0(x) = f(x)$ for all x in X is always surjective. Hence, associated with each map $f: X \to Y$ is the surjective map $f_0: X \to \operatorname{Im} f$.

The importance of the map f_0 lies in the fact that it completely determines the map f if we assume that we know the range of f. For it is easily checked that the map f is the composition

$$X \xrightarrow{f_0} \operatorname{Im} f \xrightarrow{\text{inc}} Y$$

where $\text{inc}: \operatorname{Im} f \to Y$ is the inclusion map of the subset $\operatorname{Im} f$ of Y. This representation of a map $f: X \to Y$ as the composition inc f_0 is called the **image analysis of the map f.**

Although we have already pointed out that the map f_0 is always a surjective map, it is equally important to observe that all inclusion maps are injective maps, or equivalently monomorphisms. The image analysis of a map therefore shows that every map can be written as the composition of a surjective map followed by an injective map. Because the representation of a map as the composition of a surjective map followed by an injective map plays a critical role in analyzing maps, we make the following definition.

Definition
Let $f: X \to Y$ be a map. By an **analysis of f** we mean a set A together with a surjective map $g: X \to A$ and an injective map $h: A \to Y$ such that $f = hg$.

We end this preliminary discussion of the analysis of a map by pointing out that all analyses of a map are essentially the same. Precisely, we have the following.

Basic Property 5.1
Suppose

$$X \xrightarrow{g} A \xrightarrow{h} Y$$

and

$$X \xrightarrow{g'} A' \xrightarrow{h'} Y$$

are both analyses of the map $f: X \to Y$. Then there is one and only one map $j: A \to A'$ such that $jg = g'$ and $h'j = h$, and this uniquely determined map $j: A \to A'$ is an isomorphism.

6. THE COIMAGE ANALYSIS OF A MAP

Another important standard analysis of a map is the coimage analysis. Before describing this analysis, it is convenient to have the following.

Definitions
Suppose $f: X \to Y$ is a map. If Y' is a subset of Y, then the set of all x in X with the property $f(x)$ is in Y' is called the **preimage of Y' under f** and is denoted by $f^{-1}(Y')$. If y is an element of Y, we write $f^{-1}(y)$ for $f^{-1}(\{y\})$.

Suppose we are given a map $f: X \to Y$. Then it is not difficult to show that the subset Coim f of 2^X consisting of all subsets of X of the form $f^{-1}(y)$ with y in Im f is a partition of X which we call the **coimage of f**. We have already seen that associated with any partition $\mathscr{C}$ of a set X is the canonical map $k_{\mathscr{C}}: X \to \mathscr{C}$ defined by $k_{\mathscr{C}}(x) = [x]_{\mathscr{C}}$ for all x in X, where $[x]_{\mathscr{C}}$ is the unique element of $\mathscr{C}$ containing x. Because each set in $\mathscr{C}$ is nonempty, it follows that the canonical map $k_{\mathscr{C}}: X \to \mathscr{C}$ is a surjective map. In particular, the map $k_{\text{Coim}\, f}: X \to \text{Coim}\, f$ is surjective.

The map $k_{\text{Coim}\, f}: X \to \text{Coim}\, f$ has another important property: There is a unique map $j_f: \text{Coim}\, f \to Y$ such that $f = j_f k_{\text{Coim}\, f}$. We first show that such a map exists. Suppose the subset X' of X is an element of Coim f. Then by definition there is a y in Im f such that $f^{-1}(y) = X'$. Hence, $f(X') = \{y\}$. Define the map $j_f: \text{Coim}\, f \to Y$ by $j_f(X')$ is the unique element y of Y such that $f(X') = \{y\}$. Then $j_f k_{\text{Coim}\, f}(x) = j_f([x]) = f(x)$ for all x in X which shows that the map $j_f: \text{Coim}\, f \to Y$ does indeed have the property $f = j_f k_{\text{Coim}\, f}$. That there is only one such map from Coim $f \to Y$ follows from the fact that $k_{\text{Coim}\, f}$ is surjective and hence an epimorphism. We leave it to the reader to verify that $j_f: \text{Coim}\, f \to Y$ is also an injective map with Im j_f = Im f and so $(j_f)_0: \text{Coim}\, f \to \text{Im}\, j$ is an isomorphism from Coim f to Im f. The map $j_f: \text{Coim}\, f \to Y$ is called the **map from Coim f to Y induced by the map $f: X \to Y$.**

It therefore follows that the composition $X \xrightarrow{k_{\text{Coim}\,f}} \text{Coim}\,f \xrightarrow{j_f} Y$ is an analysis of the map f. This analysis is called the **coimage analysis of the map f.** Further, because

$$X \xrightarrow{k_{\text{Coim}\,f}} \text{Coim}\,f \xrightarrow{j_f} Y$$

$$X \xrightarrow{f_0} \text{Im}\,f \xrightarrow{\text{inc}} Y$$

are both analyses of f, we know that there is a unique isomorphism $g: \text{Coim}\,f \to \text{Im}\,f$ such that $gk_{\text{Coim}\,f} = f_0$ and $\text{inc}\,g = j_f$. The map g is easily seen to be the isomorphism $(j_f)_0: \text{Coim}\,f \to \text{Im}\,f$. This isomorphism is called the **canonical isomorphism from Coim f to Im f.**

7. DESCRIPTION OF SURJECTIVE MAPS

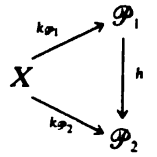
Suppose $f: X \to Y$ is a map. Then the coimage analysis $X \xrightarrow{k_{\text{Coim}\,f}} \text{Coim}\,f \xrightarrow{j_f} Y$ of f has the property that f is a surjective map if and only if $j_f: \text{Coim}\,f \to Y$ is a surjective map and hence an isomorphism. This certainly suggests that when $f: X \to Y$ is a surjective map, the maps $k_{\text{Coim}\,f}: X \to \text{Coim}\,f$ and $f: X \to Y$ are intimately connected. It is precisely this connection that we describe in this section. We begin with the following.

Definition

Let $\mathscr{P}_1$ and $\mathscr{P}_2$ be two partitions on a set X. We say that $\mathscr{P}_1$ is a **refinement** of $\mathscr{P}_2$ if given a subset X' of X in $\mathscr{P}_1$, there is a subset X'' of X in $\mathscr{P}_2$ such that $X'' \supset X'$. We shall denote the fact that $\mathscr{P}_1$ is a refinement of $\mathscr{P}_2$ by writing $\mathscr{P}_1 \geq \mathscr{P}_2$.

Now if the partition $\mathscr{P}_1$ is a refinement of the partition $\mathscr{P}_2$, then given any subset X' of X in $\mathscr{P}_1$, not only is there a subset X'' of X in $\mathscr{P}_2$ containing X', but there is only one such subset of X in $\mathscr{P}_2$. Hence, if $\mathscr{P}_1 \geq \mathscr{P}_2$, we can define the map $g_{\mathscr{P}_1, \mathscr{P}_2}: \mathscr{P}_1 \to \mathscr{P}_2$ by setting $g_{\mathscr{P}_1, \mathscr{P}_2}(X')$, for each element X' in $\mathscr{P}_1$, to be the unique element of $\mathscr{P}_2$ containing the element X' of $\mathscr{P}_1$. The map $g_{\mathscr{P}_1, \mathscr{P}_2}: \mathscr{P}_1 \to \mathscr{P}_2$ is called the **canonical map from $\mathscr{P}_1$ to $\mathscr{P}_2$.**

The map $g_{\mathscr{P}_1, \mathscr{P}_2}: \mathscr{P}_1 \to \mathscr{P}_2$ can be characterized as the unique map $h: \mathscr{P}_1 \to \mathscr{P}_2$ such that the diagram



commutes, that is, such that $hk_{\mathscr{P}_1} = k_{\mathscr{P}_2}$. This, of course, shows that $g_{\mathscr{P}_1, \mathscr{P}_2}$ is a surjective map. It also shows that $g_{\mathscr{P}_1, \mathscr{P}_2}$ is an injective map or, equivalently, is an isomorphism if and only if $\mathscr{P}_1 = \mathscr{P}_2$. Further, if $\mathscr{P}_1 = \mathscr{P}_2$, then $g_{\mathscr{P}_1, \mathscr{P}_1} = \text{id}_{\mathscr{P}_1}$. Finally, it is straightforward to show that if two partitions $\mathscr{P}_1$ and $\mathscr{P}_2$ of X have the property that there is a map $h: \mathscr{P}_1 \to \mathscr{P}_2$ such that $hk_{\mathscr{P}_1} = k_{\mathscr{P}_2}$, then $\mathscr{P}_1 \geq \mathscr{P}_2$ and $h = g_{\mathscr{P}_1, \mathscr{P}_2}$. In summary, we have the following.

Basic Properties 7.1
Let $\mathscr{P}_1$ and $\mathscr{P}_2$ be partitions of a set X. Then:

(a) There is a map $h: \mathscr{P}_1 \to \mathscr{P}_2$ such that $hk_{\mathscr{P}_1} = k_{\mathscr{P}_2}$ if and only if $\mathscr{P}_1 \geq \mathscr{P}_2$.
(b) If $\mathscr{P}_1 \geq \mathscr{P}_2$, there is only one map $h: \mathscr{P}_1 \to \mathscr{P}_2$ such that $hk_{\mathscr{P}_1} = k_{\mathscr{P}_2}$; namely, the canonical map $g_{\mathscr{P}_1, \mathscr{P}_2}$.
(c) If $\mathscr{P}_1 \geq \mathscr{P}_2$, then the canonical map $g_{\mathscr{P}_1, \mathscr{P}_2}: \mathscr{P}_1 \to \mathscr{P}_2$ is always a surjective map which is an isomorphism if and only if $\mathscr{P}_1 = \mathscr{P}_2$.
(d) $g_{\mathscr{P}_1, \mathscr{P}_1} = \mathrm{id}_{\mathscr{P}_1}$.

We can now state the main results concerning the connections between arbitrary surjective maps $f: X \to Y$ and their associated surjective maps $k_{\mathrm{Coim}\, f}: X \to \mathrm{Coim}\, f$.

Proposition 7.2
Let $f_1: X \to Y_1$ and $f_2: X \to Y_2$ be two surjective maps.

(a) The following statements are equivalent:
  (i) There is a map $h: Y_1 \to Y_2$ such that $hf_1 = f_2$.
  (ii) There is a map $g: \mathrm{Coim}\, f_1 \to \mathrm{Coim}\, f_2$ such that $gk_{\mathrm{Coim}\, f_1} = k_{\mathrm{Coim}\, f_2}$.
  (iii) $\mathrm{Coim}\, f_1 \geq \mathrm{Coim}\, f_2$.
(b) If there is a map $h: Y_1 \to Y_2$ such that $hf_1 = f_2$, then:
  (i) There is only one such map.
  (ii) There is only one map $g: \mathrm{Coim}\, f_1 \to \mathrm{Coim}\, f_2$ such that

$$gk_{\mathrm{Coim}\, f_1} = k_{\mathrm{Coim}\, f_2}, \quad \text{namely,} \quad g_{\mathrm{Coim}\, f_1, \mathrm{Coim}\, f_2}$$

(c) The following are equivalent:
  (i) There is an isomorphism $h: Y_1 \to Y_2$ such that $hf_1 = f_2$.
  (ii) $\mathrm{Coim}\, f_1 = \mathrm{Coim}\, f_2$.
(d) If Y is a partition of X and $f: X \to Y$ is the canonical map, then $\mathrm{Coim}\, f = Y$ and $f = k_{\mathrm{Coim}\, f}$.

Roughly speaking, this proposition says that all surjective maps $f: X \to Y$ with a fixed domain X are essentially given by the canonical maps $k_{\mathscr{C}}: X \to \mathscr{C}$ for all partitions $\mathscr{C}$ of X. Hence, it is of considerable importance to know how to create partitions of a set X. One of the most widely used devices for accomplishing this is known as an equivalence relation, a notion we discuss in the next section. Before doing this we point out the following generalization of some of our results to date concerning surjective maps.

Proposition 7.3
Let $f: X \to Y$ be a surjective map of sets. If $g: X \to Z$ is a map of sets, then there exists a map $h: Y \to Z$ such that $hf = g$ if and only if $\mathrm{Coim}\, f \geq \mathrm{Coim}\, g$. If $\mathrm{Coim}\, f \geq \mathrm{Coim}\, g$, then there is only one map $h: Y \to Z$ such that $hf = g$.

8. EQUIVALENCE RELATIONS

By definition, a **relation** R on a set X is simply any subset of $X \times X$. We usually denote the fact that an element (x_1, x_2) in $X \times X$ is in the relation R of X by writing

$x_1 \mathrm{R} x_2$. If R is a relation on a set X and X' is a subset of X, then we denote by $\mathrm{R}|X'$ the relation on X' given by $\mathrm{R} \cap (X' \times X')$. That is, if x_1' and x_2' are in X', then $(x_1', x_2') \in \mathrm{R}|X'$ if and only if $(x_1', x_2') \in \mathrm{R}$. The relation $\mathrm{R}|X'$ on X' is called the **relation on X' induced by R.**

Definition
A relation R on a set X is called an **equivalence relation** if it satisfies the following conditions:

(a) $x \mathrm{R} x$ holds for all x in X.
(b) If $x_1 \mathrm{R} x_2$ holds, then $x_2 \mathrm{R} x_1$ also holds.
(c) If $x_1 \mathrm{R} x_2$ and $x_2 \mathrm{R} x_3$ hold for x_1, x_2, and x_3 in X, then $x_1 \mathrm{R} x_3$ also holds.

We now describe how to associate with each equivalence relation R on a set X a partition X/R of X. For each element x in X, denote by $[x]_{\mathrm{R}}$ the subset of X consisting of all elements x' in X such that $x \mathrm{R} x'$ holds. Let X/R be the subset of 2^X consisting of the subsets $[x]_{\mathrm{R}}$ of X as x ranges through all elements of X. Then it is not difficult to show that X/R is a partition of X with the property $[x]_{X/\mathrm{R}} = [x]_{\mathrm{R}}$ for each x in X. Hence, one way to create a partition on a set X is to start out with an equivalence relation R on the set X and construct the partition X/R of X.

On the other hand, with each partition $\mathcal{P}$ on X, there is associated an equivalence relation $\mathrm{R}(\mathcal{P})$. Namely, for x_1 and x_2 in X define $x_1 \mathrm{R}(\mathcal{P}) x_2$ if and only if there is a subset X' of X in $\mathcal{P}$ such that x_1 and x_2 are both in X'. It is easy to check that the relation $\mathrm{R}(\mathcal{P})$ we just defined is actually an equivalence relation.

Moreover, it is equally easy to see that if R is an equivalence relation on a set X, then $\mathrm{R}(X/\mathrm{R}) = \mathrm{R}$. Similarly, if $\mathcal{P}$ is a partition on a set X, then $X/\mathrm{R}(\mathcal{P}) = \mathcal{P}$. This description of how to go back and forth between equivalence relations and partitions of a set shows that these are really interchangeable notions, a fact that we shall use freely from now on.

To illustrate this point, the reader should check the validity of the following proposition.

Proposition 8.1
Let R and R′ be two equivalence relations on a set X. Then $X/\mathrm{R} \geq X/\mathrm{R}'$ if and only if $x_1 \mathrm{R} x_2$ implies $x_1 \mathrm{R}' x_2$ for all x_1 and x_2 in X.

This suggests the following definition.

Definition
Let R and R′ be two relations on a set X. Then $\mathrm{R} \leq \mathrm{R}'$ if and only if $x_1 \mathrm{R} x_2$ implies $x_1 \mathrm{R}' x_2$; or, equivalently, $\mathrm{R} \subset \mathrm{R}'$.

It should be noted that if R and R′ are equivalence relations on a set X, then $\mathrm{R} \leq \mathrm{R}'$ if and only if $X/\mathrm{R} \geq X/\mathrm{R}'$.

As our final example of the correspondence between the partitions and equivalence relations on a set, we point out that if $f: X \to Y$ is a map, then the equivalence relation R on X corresponding to the partition Coim f of X is given by $x_1 \mathrm{R} x_2$ if and only if $f(x_1) = f(x_2)$. This equivalence relation is called the **equivalence relation associated with f** and is sometimes denoted by $\mathrm{R}(f)$.

9. CARDINALITY OF SETS

One of the earliest and most important mathematical processes one learns is that of counting; and one of the basic problems in counting is to determine when two sets of things have the same number of objects. This is usually done by showing that the objects in one collection can be matched up with the objects in the other collection. But the matching up of the objects in a set X with the objects in a set Y is nothing more than a bijective map from X to Y. This leads us to say that an arbitrary set X has the same number of elements as a set Y if and only if there is a bijective map $f: X \to Y$. Obviously, a set X has the same number of elements as a set Y if and only if the set Y has the same number of elements as X. The fact that two sets X and Y have the same number of elements is often denoted by $\text{card}(X) = \text{card}(Y)$ where $\text{card}(X)$ is read as the **cardinality of X**. Or stated slightly differently, two sets X and Y have the same cardinality (that is, the same number of elements) if and only if they are isomorphic sets.

In addition to knowing when two sets have the same number of elements, it is also important to know when one set Y has at least as many elements as another set X. A little thought should convince the reader that in usual practice this simply means that there is injective map $f: X \to Y$. This observation leads us to define the **cardinality of an arbitrary set X as being less than or equal to the cardinality of a set Y**; symbolically, $\text{card}(X) \leq \text{card}(Y)$ if and only if there is an injective map $f: X \to Y$.

If this definition of $\text{card}(X) \leq \text{card}(Y)$ really corresponds to the notion that the set X has at most as many elements as the set Y, then it should have the following properties:

(a) $\text{card}(X) \leq \text{card}(X)$ for all sets X.
(b) If $\text{card}(X) \leq \text{card}(Y)$ and $\text{card}(Y) \leq \text{card}(X)$, then $\text{card}(X) = \text{card}(Y)$.
(c) If $\text{card}(X) \leq \text{card}(Y)$ and $\text{card}(Y) \leq \text{card}(Z)$, then $\text{card}(X) \leq \text{card}(Z)$.

It is trivial to verify that (a) and (c) are true. The fact that (b) is true is less obvious and is equivalent to the following well-known Bernstein-Schroeder Theorem.

Theorem 9.1
Suppose X and Y are two sets and $f: X \to Y$ and $g: Y \to X$ are injective maps. Then X and Y are isomorphic sets.

An outline of a proof of this theorem is given in the exercises for the convenience of those readers not familiar with this result.

There are two more properties of the cardinality of sets that one might expect to be true judging from one's ordinary experience with counting. Namely, (1) given a set X there is a set Y such that Y has actually more elements than X, that is, $\text{card}(X) \leq \text{card}(Y)$ but $\text{card}(X) \neq \text{card}(Y)$; or, more simply, $\text{card}(X) < \text{card}(Y)$ and (2) if X and Y are sets, then either $\text{card}(X) \leq \text{card}(Y)$ or $\text{card}(Y) \leq \text{card}(X)$.

The fact that given any set X there is a set Y such that $\text{card}(X) < \text{card}(Y)$ follows from the following proposition, a proof of which is outlined in the exercises.

Proposition 9.2
Let X be any set. Then there is no surjective map from X to 2^X.

Hence, X and 2^X are never isomorphic sets, which means $\text{card}(X) \neq \text{card}(2^X)$. On the other hand, the map $f: X \to 2^X$ given by $f(x) = \{x\}$ for each x in X is clearly an injective map. Therefore, $\text{card}(X) \leq \text{card}(2^X)$ which implies that $\text{card}(X) < \text{card}(2^X)$ for each set X. Thus, given any set X there is a set Y such that $\text{card}(X) < \text{card}(Y)$, which settles the first question raised.

However, the second question, whether given two sets X and Y, either $\text{card}(X) \leq \text{card}(Y)$ or $\text{card}(Y) \leq \text{card}(X)$, is much more complicated. In fact, it cannot be settled except by the introduction of a notion of set theory which we have not discussed at all; namely, the axiom of choice. Therefore, we shall return to this second question in a later section after we have discussed this axiom of set theory.

10. ORDERED SETS

There are various equivalent forms of the axiom of choice. We shall be concerned with only three of them: the existence of choice functions, the well-ordering axiom, and Zorn's lemma. Because all but the first of these forms of the axiom of choice use the notion of an ordered set in their formulation, we shall begin this discussion with the notion of an ordered set.

Definition
A relation R on a set X is said to be an **order relation on X** or an **ordering of X**, if it satisfies:

(a) $x \,\mathrm{R}\, x$ for all x in X.
(b) If $x_1 \,\mathrm{R}\, x_2$ and $x_2 \,\mathrm{R}\, x_1$, then $x_1 = x_2$.
(c) If $x_1 \,\mathrm{R}\, x_2$ and $x_2 \,\mathrm{R}\, x_3$, then $x_1 \,\mathrm{R}\, x_3$.

An ordering R of X is called a **total ordering of X** if it also satisfies:

(d) If x_1 and x_2 are in X, then either $x_1 \,\mathrm{R}\, x_2$ or $x_2 \,\mathrm{R}\, x_1$.

Finally, a set X together with an ordering R (total ordering R) is called an **ordered set (totally ordered set)**.

The reader should observe that if the relation R on a set X is an order relation and X' is a subset of X, then the relation $\mathrm{R}|X'$ is an order relation on X' called the **induced ordering on X'**. Unless stated explicitly to the contrary, if X' is a subset of an ordered set X, we always consider X' an ordered set under the induced ordering. Obviously, if X is a totally ordered set so is X' for each subset X' of X.

When there is no danger of confusion concerning which ordering we mean, we shall follow the usual practice of writing $x_1 \leq x_2$ for $x_1 \,\mathrm{R}\, x_2$ when R is an order relation on the set X.

We now offer as examples certain ordered sets that will be occurring frequently in the rest of the book.

Example 10.1 Suppose X is a set. It is easy to check that the relation R on 2^X

given by $X' \mathrm{R} X''$ if and only if $X' \subset X''$ is an order relation. This is the only order relation we shall ever consider on the set 2^X. Hence, when we consider 2^X an ordered set it is always with respect to this ordering. The reader should observe that 2^X is a totally ordered set if and only if X has at most one element.

The next example, which is closely related to our first one, is extremely useful in constructing maps, as we shall see later on.

Example 10.2 Suppose we are given two sets X and Y. Let $\mathcal{T}(X, Y)$ consist of all triples (X', Y', f) where X' and Y' are subsets of X and Y, respectively, and f is a map from X' to Y'. Consider the relation R on $\mathcal{T}(X, Y)$ given by $(X', Y', f')\,\mathrm{R}\,(X'', Y'', f'')$ if and only if $X' \subset X''$, $Y' \subset Y''$, and $f''(x) = f'(x)$ for all x in X'. It is easily seen that R is an order relation on $\mathcal{T}(X, Y)$. Hence, when we refer to $\mathcal{T}(X, Y)$ as an ordered set, it is always with respect to this ordering.

Finally, we have the following familiar ordered sets.

Example 10.3 All of the following sets with their usual ordering are totally ordered sets:

(a) The set **N** of all nonnegative integers, that is, all integers $n \geq 0$.
(b) **Z**, the set of all integers.
(c) **Q**, the set of all rational numbers.
(d) **R**, the set of all real numbers.

Now that we have defined the notion of an ordered set we can start discussing the axiom of choice.

11. AXIOM OF CHOICE

That every set X has a choice function is perhaps the simplest and most appealing form of the axiom of choice. What this amounts to saying is that given any nonempty collection of nonempty subsets of a set X, it is possible to choose an element out of each one. Although this seems self-evident, it nonetheless cannot be proven on the basis of the types of manipulations of sets we have permitted ourselves until now. Formulated somewhat more precisely, this assertion becomes the following.

Axiom of Choice 1
Given any set X, there is a map $c : 2^X - \{\emptyset\} \to X$ (where $2^X - \{\emptyset\}$ is the set of all nonempty subsets of X) such that $c(X') \in X'$ for all nonempty subsets X' of X. Such a map c is called a **choice function on the set X**.

As an illustration of how this form of the axiom of choice is used, we prove the following proposition.

Proposition 11.1
Let $f : X \to Y$ be a surjective map of sets. Then there is a map $g : Y \to X$ such that $fg = \mathrm{id}_Y$.

PROOF: If $Y = \emptyset$, then X is empty and f is an isomorphism so there is nothing to prove.

Suppose now that $Y \neq \emptyset$ and $c: 2^X - \{\emptyset\} \to X$ is a choice function on X. Then define $g: Y \to X$ by $g(y) = c(f^{-1}(y))$ for each y in Y. Since $c(f^{-1}(y))$ is in $f^{-1}(y)$, it follows that $f(c(f^{-1}(y))) = y$ for all y in Y. Therefore, the map $g: Y \to X$ has the property $fg = \mathrm{id}_Y$.

In order to state the next form of the axiom of choice that interests us, it is necessary to recall the definition of a well-ordered set.

Definition
An ordered set X is said to be **well ordered** if

(a) X is totally ordered.
(b) If X' is a nonempty subset of X, then there is an element x_0 in X', called the **first element** of X', having the property $x_0 \leq x$ for all x in X'.

It is important to note that if X is a well-ordered set, then the first element of a nonempty subset X' of X is uniquely determined. For if x_0 and x_0' are both first elements in X', then $x_0 \leq x_0'$ and $x_0' \leq x_0$ which means that $x_0 = x_0'$.

Axiom of Choice 2
If X is a set, then there is an ordering on X which makes X a well-ordered set.

As stated earlier, these two forms of the axiom of choice that have been given are equivalent. Although it is certainly not trivial to show that the assumption that every set has a choice function implies that every set can be well ordered (see the exercises for a discussion of this point), the reverse implication is quite simple to establish as we now show.

Proposition 11.2
Let X be a well-ordered set. Then X has a choice function.

PROOF: Since the first element of any nonempty subset X' of X is a uniquely determined element of X', we obtain a choice function c on X by defining the map $c: 2^X - \{\emptyset\} \to X$ as follows: $c(X')$ is the first element of X' for each nonempty subset X' of X.

As a check on his understanding of well-ordered sets the reader should convince himself that while the set $\mathbf{N}$ of nonnegative integers is a well-ordered set, neither the integers, rational numbers, nor real numbers is a well-ordered set even though each of them is totally ordered.

We now turn our attention to the third and final form of the axiom of choice which is of concern to us, namely, Zorn's lemma. Although this form of the axiom of choice is much more technical and therefore has less intuitive appeal than the others, it has the advantage of being the easiest to apply in most situations of interest to us.

Before stating Zorn's lemma we review the notion of an inductive set.

Definition
An ordered set X is said to be an **inductive set** if every nonempty totally ordered subset X' of X has an upper bound in X. That is, for each nonempty totally ordered X' of X there is an element x in X such that $x \geq x'$ for all elements x' in X'.

To help clarify this definition we give some important examples of inductive sets.

Example 11.3 If X is a set, then the ordered set 2^X is an inductive set.

PROOF: Suppose $\mathcal{S}$ is any subset of 2^X and $Y = \bigcup_{X' \in \mathcal{S}} X'$. Then it is obvious that Y is an element of 2^X which is an upper bound of $\mathcal{S}$ in the sense that $Y \supset X'$ for all X' in $\mathcal{S}$. Thus, certainly every nonempty totally ordered subset $\mathcal{S}$ of 2^X has an upper bound in 2^X.

Example 11.4 Let X and Y be sets. Then the ordered set $\mathcal{T}(X, Y)$ is an inductive set.

PROOF: Suppose $\mathcal{S}$ is a nonempty totally ordered subset of $\mathcal{T}(X, Y)$, say $\mathcal{S} = \{(X'_\alpha, Y'_\alpha, f'_\alpha)\}$. Let $X_0 = \cup X'_\alpha$ and $Y_0 = \cup Y'_\alpha$.

We claim that because $\mathcal{S}$ is totally ordered there is a map $f_0 : X_0 \to Y_0$ such that (X_0, Y_0, f_0) is in $\mathcal{T}$. For suppose x_0 is in X_0, then by the definition of X_0 there is a triple $(X'_\alpha, Y'_\alpha, f'_\alpha)$ in $\mathcal{S}$ such that x_0 is in X'_α. Hence, it is tempting to define $f_0(x_0)$ to be $f'_\alpha(x_0)$. In order for this to be legitimate, we must show that if there is some other $(X'_\beta, Y'_\beta, f'_\beta)$ in $\mathcal{S}$ with x_0 in X'_β, then $f'_\alpha(x_0) = f'_\beta(x_0)$, for otherwise the value $f_0(x_0)$ would not be uniquely determined but would depend on the particular element of $\mathcal{S}$ used in its construction. But the fact that $f'_\alpha(x_0) = f'_\beta(x_0)$ follows from the fact that $\mathcal{S}$ is totally ordered. For we know that either $(X'_\alpha, Y'_\alpha, f'_\alpha) \leq (X'_\beta, Y'_\beta, f'_\beta)$ or $(X'_\beta, Y'_\beta, f'_\beta) \leq (X'_\alpha, Y'_\alpha, f'_\alpha)$. Now we suppose that $(X'_\alpha, Y'_\alpha, f'_\alpha) \leq (X'_\beta, Y'_\beta, f'_\beta)$. Then we know that $X'_\alpha \subseteq X'_\beta$ and $f'_\alpha(x) = f'_\beta(x)$ for all x in X'_α and hence, in particular, $f'_\alpha(x_0) = f'_\beta(x_0)$. A similar argument works in case $(X'_\beta, Y'_\beta, f'_\beta) \leq (X'_\alpha, Y'_\alpha, f'_\alpha)$.

Hence, we have shown that there is a map $f_0 : X_0 \to Y_0$. It is not hard to show now that (X_0, Y_0, f_0) is an upper bound for $\mathcal{S}$ since $(X_0, Y_0, f_0) \geq (X', Y', f')$ for all (X', Y', f') in $\mathcal{S}$.

The form of the axiom of choice known as Zorn's lemma is simply the following.

Axiom of Choice 3
If X is an inductive set, then there is an element x_0 in X such that if x is in X and $x \geq x_0$, then $x = x_0$. Such an element x of X is called a **maximal element of X.**

As an illustration of how Zorn's lemma is used, we finally give the much delayed proof that if X and Y are two sets, then either $\text{card}(X) \leq \text{card}(Y)$ or $\text{card}(Y) \leq \text{card}(X)$.

Proposition 11.5
If X and Y are sets, then there is either an injective map from X to Y or from Y to X.

PROOF: Let $\mathrm{Inj}(X, Y)$ be the ordered subset of $\mathcal{T}(X, Y)$ consisting of all triples (X', Y', f') with the property that $f' : X' \to Y'$ is an injective map. Using the same type of argument as in Example 11.4, it is not difficult to see that $\mathrm{Inj}(X, Y)$ is an inductive set.

Because $\mathrm{Inj}(X, Y)$ is an inductive set we know by Zorn's lemma that there is a maximal element (X', Y', f') in $\mathrm{Inj}(X, Y)$. This maximal element (X', Y', f') has the property that either $X' = X$ or $Y' = Y$. For suppose there is an x_0 in X but not in X' and an element y_0 in Y but not in Y'. Then the map $g : X' \cup \{x_0\} \to Y' \cup \{y_0\}$ defined by $g(x) = f'(x)$ for x in X' and $g(x_0) = y_0$ is injective. Therefore, $(X' \cup \{x_0\}, Y' \cup \{y_0\}, g)$ is an element of $\mathrm{Inj}(X, Y)$ with the property that $(X', Y', f') < (X' \cup \{x_0\}, Y' \cup \{y_0\}, g)$. This contradicts the fact that (X', Y', f') is a maximal element of $\mathrm{Inj}(X, Y)$. Therefore, our contention that either $X' = X$ or $Y' = Y$ has been established.

If $X' = X$, then the composition $X \xrightarrow{f'} Y' \xrightarrow{\mathrm{inc}} Y$ of injective maps is an injective map from X to Y.

On the other hand, if $Y' = Y$, we can define the map $g : Y \to X$ by letting $g(y)$ be the unique element in X such that $f'(g(y)) = y$. It is obvious that g is an injective map, and so in this case we obtain an injective map from Y to X. This completes the proof of the proposition.

We finish this discussion of the axiom of choice by pointing out that in the exercises there is an outline of a proof of our repeated assertions that the various conditions Axiom of Choice 1, 2, and 3 are equivalent. From now on we will make free use of these forms of the axiom of choice, especially Zorn's lemma.

12. PRODUCTS AND SUMS OF SETS

In discussing the product of sets, it is convenient to have the notion of an indexed family of subsets of a fixed set.

Definition
A **family of subsets of a set X indexed by a set I** is a map $\psi : I \to 2^X$. The set I is called the **indexing set** and the subset $\psi(i)$ of X is usually denoted by X_i. In practice one denotes the map ψ by $\{X_i\}_{i \in I}$.

In connection with this definition we observe that associated with an indexed family $\{X_i\}_{i \in I}$ of subsets of a set X is the subset $\mathcal{S}$ of 2^X consisting of all subsets X' of X such that $X' = X_i$ for some i in I. The reader should construct examples of different indexed families of subsets of a fixed set X which give rise to the same associated subset $\mathcal{S}$ of 2^X.

Definition
The **product of an indexed family $\{X_i\}_{i \in I}$ of subsets of a set X** is the set of all maps $f : I \to X$ such that $f(i) \in X_i$ for all i in I. This product is denoted by $\prod_{i \in I} X_i$.

Usually a different notation is used for the elements of a product $\prod_{i\in I} X_i$ of an indexed set $\{X_i\}_{i\in I}$ of subsets of X. If f is in $\prod_{i\in I} X_i$, then the element $f(i)$ in X_i is denoted by x_i and the element f of $\prod_{i\in I} X_i$ is denoted by $\{x_i\}_{i\in I}$ where $x_i = f(i)$ for all i in I.

If $\{X_i\}_{i\in I}$ is an indexed family of subsets of a fixed set X, then certain things about the product $\prod_{i\in I} X_i$ are obvious. First of all, if $I = \emptyset$, then $\prod_{i\in I} X_i$ consists of one element, namely, the empty map. So suppose from now on that $I \neq \emptyset$. Then it is clear that if $X_i = \emptyset$ for some i in I, then $\prod_{i\in I} X_i = \emptyset$. What might not be quite so obvious, since it depends on the existence of choice functions, is that $\prod_{i\in I} X_i \neq \emptyset$ if each $X_i \neq \emptyset$. For if $c: 2^X - \{\emptyset\} \to X$ is a choice function, then $\{c(X_i)\}_{i\in I}$ is in $\prod_{i\in I} X_i$ since each $c(X_i)$ is in X_i for all i in I. Hence, if $I \neq \emptyset$, then $\prod_{i\in I} X_i = \emptyset$ if and only if $X_i = \emptyset$ for some i in I.

In studying and using products of families of sets, the maps called projection maps play an important role.

Definition
Let $\{X_i\}_{i\in I}$ be an indexed family of subsets of a set X indexed by the nonempty set I. Then for each $k \in I$ we have the map $\text{proj}_k: \prod_{i\in I} X_i \to X_k$ given by $\text{proj}_k(\{x_i\}_{i\in I}) = x_k$. The map $\text{proj}_k: \prod_{i\in I} X_i \to X_k$ is called the ***k*th projection map**. Also, $\text{proj}_k(\{x_i\}_{i\in I}) = x_k$ is called the ***k*th coordinate of the element $\{x_i\}_{i\in I}$ in the product $\prod_{i\in I} X_i$.**

The following facts concerning projection maps are not difficult to establish.

Basic Properties 12.1
Let $\{X_i\}_{i\in I}$ be an indexed family of nonempty subsets of set X indexed by the nonempty set I. Then:

(a) For each k in I, the map $\text{proj}_k: \prod_{i\in I} X_i \to X_k$ is surjective.

(b) For each element $\{x_i\}_{i\in I}$ in $\prod_{i\in I} X_i$, we have $\{x_i\}_{i\in I} = \{\text{proj}_k(\{x_i\}_{i\in I})\}_{k\in I}$.

We end this discussion of products of sets with a description of the set of all maps $(X, \prod_{i\in I} Y_i)$ from a set X to the product $\prod_{i\in I} Y_i$ of the indexed family of sets $\{Y_i\}_{i\in I}$. To do this, we first observe that if f is a map from X to $\prod_{i\in I} Y_i$, then for each k in I, the composition $\text{proj}_k f$ is a map from X to Y_k. Hence, associated with an element f in $(X, \prod_{i\in I} Y_i)$ is the element $\{\text{proj}_k f\}_{k\in I}$ of $\prod_{i\in I}(X, Y_i)$. Thus, we obtain the map $\beta: (X, \prod_{i\in I} Y_i) \to \prod_{i\in I}(X, Y_i)$ given by $\beta(f) = \{\text{proj}_k f\}_{k\in I}$ for each f in $(X, \prod_{i\in I} Y_i)$.

Proposition 12.2
Let $\{Y_i\}_{i\in I}$ be an indexed family of subsets of a set Y indexed by the nonempty set I and X an arbitrary set. Then the map $\beta: (X, \prod_{i\in I} Y_i) \to \prod_{i\in I}(X, Y_i)$ given by $\beta(f) = \{\text{proj}_k f\}_{k\in I}$ is an isomorphism of sets.

Another useful construction associated with an indexed family $\{X_i\}_{i\in I}$ of subsets of a set X is the sum of the indexed family.

Definition
Let $\{X_i\}_{i\in I}$ be an indexed family of subsets of a set X. The **sum of this indexed family** is the subset $\coprod_{i\in I} X_i$ of $X\times I$ consisting of all elements (x, i) in $X\times I$ such that x is in X_i.

If $\{X_i\}_{i\in I}$ is an indexed family of subsets of a set X, the reader should observe the following facts about the sum $\coprod_{i\in I} X_i$. If $I=\emptyset$, then $X\times I=\emptyset$ and hence $\coprod_{i\in I} X_i=\emptyset$. Also regardless of whether I is empty or not, $\coprod_{i\in I} X_i=\emptyset$ if each $X_i=\emptyset$. Finally, $\coprod_{i\in I} X_i\neq\emptyset$ if $I\neq\emptyset$ and some $X_i\neq\emptyset$.

Analogous to the projection maps for the product of an indexed family of subsets of a set X is the injection maps for a sum of the indexed family of subsets of X.

Definition
Let $\coprod_{i\in I} X_i$ be the sum of the indexed family $\{X_i\}_{i\in I}$ of subsets of X indexed by the nonempty set I. For each k in I the map $\text{inj}_k: X_k \to \coprod_{i\in I} X_i$ defined by $\text{inj}_k(x_k)=(x_k, k)$ for each x_k in X_k is called the **kth injection map**.

We leave it to the reader to verify the following.

Basic Properties 12.3
Let $\coprod_{i\in I} X_i$ be the sum of the indexed family $\{X_i\}_{i\in I}$ of subsets of a set X indexed by the nonempty set I. The injection maps $\text{inj}_k: X_k \to \coprod_{i\in I} X_i$ have the following properties:

(a) For each k in I the map $\text{inj}_k: X_k \to \coprod_{i\in I} X_i$ is injective.

(b) Letting $\text{Im inj}_k = Y_k$ we have:
  (i) $Y_k \cap Y_{k'} = \emptyset$ if $k\neq k'$.
  (ii) $\coprod_{i\in I} X_i = \bigcup_{k\in I} Y_k$.

We end this section by establishing the analog of Proposition 12.2 for sums. To this end, we observe that if $f: \coprod_{i\in I} X_i \to Y$ is a map from the sum of the indexed family $\{X_i\}_{i\in I}$ to the set Y, then for each $k \in I$ the composition $f \text{ inj}_k$ is a map from X_k to Y. Hence, associated with an element f in $(\coprod_{i\in I} X_i, Y)$ is the element $\{f \text{ inj}_k\}_{k\in I}$ of $\prod_{i\in I} (X_i, Y)$. Thus, we obtain the map $\gamma: (\coprod_{i\in I} X_i, Y) \to \prod_{i\in I} (X_i, Y)$ given by $\gamma(f) = \{f \text{ inj}_k\}_{k\in I}$ for each f in $(\coprod_{i\in I} X_i, Y)$.

Proposition 12.4
Let $\{X_i\}_{i\in I}$ be an indexed family of subsets of X indexed by the nonempty set I and let Y be an arbitrary set. Then the map $\gamma: (\coprod_{i\in I} X_i, Y) \to \prod_{i\in I} (X_i, Y)$ given by $\gamma(f) = \{f \text{ inj}_k\}_{k\in I}$ for each f in $(\coprod_{i\in I} X_i, Y)$ is an isomorphism of sets.

EXERCISES

(1) Throughout this exercise $\mathbf{R}$ denotes the set of real numbers.

(a) Let $f:\mathbf{R}\to\mathbf{R}\times\mathbf{R}$ be the map defined by $f(x)=(\cos x, \sin x)$ for all x in R. Describe:
  (i) Im f.
  (ii) $f^{-1}(y)$ for each y in Im f.

(b) Let $\mathbf{R}^+$ be the set of nonnegative real numbers. Show that the logarithmic map $\log_e:\mathbf{R}^+\to\mathbf{R}$ is an isomorphism of sets whose inverse is the exponential map from $\mathbf{R}$ to $\mathbf{R}^+$.

(c) Show that any two circles, regardless of size, are isomorphic sets.

(d) Show that the subset of $\mathbf{R}\times\mathbf{R}$ consisting of all ordered pairs (x, y) satisfying $0\le y<\pi$ and x arbitrary is isomorphic to the set $\mathbf{R}\times\mathbf{R}$. [*Hint*: Use polar coordinates.]

(e) Let $C^\infty(\mathbf{R})$ denote the set of all maps from $\mathbf{R}$ to $\mathbf{R}$ which have all derivatives. Consider the maps $d/dx: C^\infty(\mathbf{R})\to C^\infty(\mathbf{R})$ and $\int_0^x: C^\infty(\mathbf{R})\to C^\infty(\mathbf{R})$. Describe:
  (i) the images of these maps;
  (ii) the preimages of elements in their images;
  (iii) the compositions of these two maps.

(2) Let X be a set consisting of a single point which we denote by x. For each set Y we define the map $\alpha:(X, Y)\to Y$ by $\alpha(f)=f(x)$ for each f in (X, Y). Prove that the map α is an isomorphism for each set Y.

(3) Prove that a map $f:X\to Y$ is a monomorphism if and only if it is an injective map. [*Hint*: To prove that the map f being a monomorphism implies that f is an injective map, use the description of the maps of a single point to the set X given in Exercise 2.]

(4) Prove that a map $f:X\to Y$ is an epimorphism if and only if it is a surjective map. [*Hint*: To prove that the map f being an epimorphism implies that it is a surjective map, consider the maps of Y to a set consisting of two distinct points.]

(5) Suppose X is a set and $\mathscr{C}\subset 2^X$ is a set of subsets of X. Show that the following statements are equivalent:

(a) $\mathscr{C}$ is a covering of X.

(b) If Y is an arbitrary set, then two maps $f:X\to Y$ and $g:X\to Y$ are the same if the restriction maps $f|X'$ and $g|X'$ are the same for each subset X' of X in $\mathscr{C}$.

(6) Suppose that $\mathscr{C}\subset 2^X$ is a covering for the set X and Y is an arbitrary set. Show that the following statements are equivalent for a family of maps $\{f_{X'}:X'\to Y\}_{X'\in\mathscr{C}}$:

(a) There is a map $f:X\to Y$ such that $f|X'=f_{X'}$ for each X' in $\mathscr{C}$.

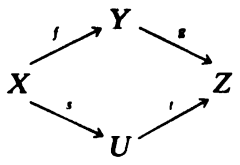
(b) For each pair of elements X' and X'' of $\mathscr{C}$ we have that $f_{X'}|X'\cap X''=f_{X''}|X'\cap X''$.

(7) Show that a subset $\mathscr{C}$ of 2^X is a partition of the set X if and only if $\mathscr{C}$ satisfies:

(a) Each subset of X in $\mathscr{C}$ is not empty.

(b) Given any set Y and family of maps $\{f_{X'}:X'\to Y\}_{X'\in\mathscr{C}}$ there is a unique map $f:X\to Y$ such that $f|X'=f_{X'}$ for all subsets of X in $\mathscr{C}$.

(8) Consider the following diagram of sets and maps:



Suppose that the diagram is commutative, that is, $gf = ts$.

(a) Prove that if f is surjective and t is injective, then there is one and only one map $h: Y \to U$ such that $hf = s$ and $th = g$. [*Hint*: For each $y \in Y$, choose $x \in X$ such that $f(x) = y$. Show that the element $s(x) \in U$ is independent of the choice of x and define $h(y)$ to be $s(x)$.]

(b) Prove that if f and s are surjective and g and t are injective, there are unique isomorphisms $h: Y \to U$ and $h': U \to Y$ such that $hf = s$, $h's = f$, and $h' = h^{-1}$.

(9) Prove that for any set X, there is no surjective map from X to 2^X. [*Hint*: If $f: X \to 2^X$ is any map and $X' = \{x \in X | x \notin f(x)\}$, show that X' is not in Im f.]

(10) The set of subsets of a set X has been denoted by 2^X. Show that if Y is a set consisting precisely of two distinct elements which we denote by 0 and 1, then the map $\beta: (X, Y) \to 2^X$ defined by $\beta(f) = f^{-1}(0)$ for all f in (X, Y) is an isomorphism of sets.

(11) Show that if Y is a set with at least two distinct elements, then $\text{card}((X, Y)) > \text{card}(X)$ for all sets X.

(12) Let X be the subset of the set of real numbers satisfying the condition x is in X if and only if $0 \leq x < 1$.

(a) Show that each real number in X can be written in one and only one way as an infinite decimal $.a_1a_2a_3 \ldots a_n \ldots$ having the property that given any integer n there is an integer $m > n$ such that $a_m \neq 9$.

(b) Show that given any infinite decimal $.a_1a_2a_3 \ldots a_n \ldots$ there is a unique element x in X such that $x = .a_1a_2a_3 \ldots a_n \ldots$.

(c) Suppose $f: \mathbf{N} \to X$ is a map of the set of nonnegative integers $\mathbf{N}$ to X. For each k in $\mathbf{N}$, let $.a_{k1}a_{k2} \ldots a_{kn} \ldots$ be the unique infinite decimal expansion of $f(k)$ satisfying the condition specified in part (a). For each k in $\mathbf{N}$ let b_k be different from 9 and a_{kk}. Show that the number b in X whose decimal expansion is $.b_1b_2 \ldots b_n \ldots$ is not in the image of f.

(d) Show that $\text{card}(X) > \text{card}(\mathbf{N})$ and hence $\text{card}(\mathbf{R}) > \text{card}(\mathbf{N})$.

(13) Let X_1, X_2 be subsets of a set X and Y_1, Y_2 subsets of a set Y. Suppose that $X = X_1 \cup X_2$, $Y = Y_1 \cup Y_2$, and that $X_1 \cap X_2 = \emptyset = Y_1 \cap Y_2$. Show that if $\text{card}(X_i) = \text{card}(Y_i)$ for $i = 1, 2$, then $\text{card}(X) = \text{card}(Y)$.

(14) (a) Let X_1, X_2, X be sets with $X_1 \subset X_2 \subset X$ and let $f: X \to X_1$ be a map. Let $\mathcal{V}$ be the set of all subsets V of X such that $f(V) \cup (X_2 - X_1) \subset V$, where $X_2 - X_1 = \{x \in X_2 | x \notin X_1\}$. Prove that if $U = \bigcap_{V \in \mathcal{V}} V$, then $U \subset X_2$ and $f(U) \cup (X_2 - X_1) = U$. [*Hint*: First prove that $U \in \mathcal{V}$. Then show that $f(U) \cup (X_2 - X_1) \in \mathcal{V}$.]

(b) From (a), prove that $X_2 = (X_1 - f(U)) \cup U$. Then prove that $\text{card}(X_1) = \text{card}(X_2)$ if f is **an injective map**. [*Hint*: Use the fact that $X_1 = (X_1 - F(U)) \cup f(U)$ and Exercise 13 above. Note that $(X_1 - f(U)) \cap U = \emptyset$.]

(c) Show, finally, that if f is bijective, then $\text{card}(X_2) = \text{card}(X)$.

(15) Let X and Y be sets such that $\text{card}(X) \leq \text{card}(Y)$ and $\text{card}(Y) \leq \text{card}(X)$. Prove that $\text{card}(X) = \text{card}(Y)$. [*Hint*: let $g: X \to Y$ and $h: Y \to X$ be injective maps. Let $X_2 = \text{Im } h$ and $X_1 = \text{Im } hg$ and let $f: X \to X_1$ be the bijective map $(hg)_0$. Then use the preceding exercise. This is the Bernstein-Schroeder theorem.]

(16) Let X be the subset of the real numbers consisting of all reals of the form $n + \sum_{j=1}^{t-1} (\frac{1}{2})^j$ where $t > 0$ and n is an integer. [We use the convention that $\sum_{j=1}^{0} (\frac{1}{2})^j = 0$.] Is X a well-ordered set?

(17) Let X be a well-ordered set. Show that if $x \in X$ and x is not an upper bound for X (that is, there exist $y \in X$ such that $y > x$), then there is one and only one element $s(x) \in X$ having the following two properties:

(i) $x < s(x)$;

(ii) if $x \leq y \leq s(x)$, then either $y = x$ or $y = s(x)$.

(18) Let X be a nonempty set.

(a) Prove that there are subsets X' of X which have a well ordering, that is, there exists an ordering on X' which makes X' a well-ordered set. [*Hint*: Consider subsets of X which consist of only one point.]

(b) If $X_1' \subset X_2' \subset X$ and R is an order relation on X_2', prove that $R \cap (X_1' \times X_1')$ is an order relation on X_1'. As usual denote this relation on X_1' by $R|X_1'$.

(c) Let $\mathcal{W}$ be the set of all pairs (X', R) where X' is a subset of X and R is a well ordering of X'. Define an order relation on $\mathcal{W}$ by setting $(X_1', R_1) < (X_2', R_2)$ if:

(i) $X_1' \subset X_2'$.

(ii) $R_2|X_1' = R_1$.

(iii) If $x_1 \in X_1'$ and $x_2 \in X_2' - X_1'$, then $x_1 < x_2$. Prove that $\mathcal{W}$ is an inductive set.

(d) Using Zorn's lemma, prove that $\mathcal{W}$ has a maximal element (X_0, R). Show that $X_0 = X$. This proves that Zorn's lemma implies well ordering.

The next two exercises constitute a proof that the axiom of choice implies Zorn's lemma.

(19) Let X be an inductive set having the additional property that every totally ordered subset X' of X has a least upper bound in X, that is, among the upper bounds of X' there is a first element. Let $f: X \to X$ be a map satisfying the following two conditions:

(i) $x \leq f(x)$ for all $x \in X$.

(ii) $x \leq y \leq f(x)$ implies $y = x$ or $y = f(x)$.

We now outline a proof that these hypotheses imply the existence of an element x in X such that $f(x_0) = x_0$. If a is an element of X, define a subset C of X to be an a-chain if C satisfies the following conditions:

(i) $a \in C$.

(ii) If $a \in C$, then $f(a) \in C$.

(iii) Every totally ordered subset of C has a least upper bound in C.

(a) Prove that if $C = \{x \in X | a \leq x\}$, then C is an a-chain.

(b) If K_a is the intersection of all a-chains, prove that K_a is an a-chain.

(c) Prove that if $x \in K_a$ and x is comparable with every element of K_a, that is, for all $y \in K_a$ either $x \leq y$ or $y \leq x$, then $f(x)$ is comparable with every element of K_a. [*Hint*: Let $C = \{y \in K_a | f(x)$ is comparable with $y\}$ and prove that C is an a-chain.]

(d) Prove that if $x \in K_a$, then x is comparable with every element of K_a, so that K_a is totally ordered. [*Hint*: Let C be the set of those elements x in K_a which are comparable with every element of K_a. Show that C is an a-chain.]

(e) Prove that K_a is well ordered.

(f) Because K_a is totally ordered and is an a-chain, we may conclude that K_a has a least upper bound x_0 in K_a. Prove that $f(x_0) = x_0$.

(20) Let X be an ordered set and let $\mathcal{T}$ be the set of all totally ordered subsets T of X. $\mathcal{T}$ is an ordered set under the usual order relation of inclusion.

(a) Prove that $\mathcal{T}$ is an inductive set satisfying the additional condition that every totally ordered subset of $\mathcal{T}$ has a least upper bound in $\mathcal{T}$.
(b) For each $T \in \mathcal{T}$, let $B_T = \{x \in X | x \geq t \text{ for all } t \in T\}$. Assume that φ is a choice function for X, and define $f: \mathcal{T} \to \mathcal{T}$ by

$$f(T) = \begin{cases} T \cup \varphi(B_T), & \text{if } B_T \neq \emptyset \\ T & \text{if } B_T = \emptyset \end{cases}$$

Prove that for all $T \in \mathcal{T}$, $T \leq f(T)$ and that if $T \leq T' \leq f(T)$, then $T = T'$ or $T' = f(T)$.
(c) Using Exercise 19 and still assuming the existence of φ, prove that $\mathcal{T}$ has a maximal element. This establishes the fact that if the axiom of choice holds, then every ordered set contains a maximal totally ordered subset.
(d) From (c) conclude that the axiom of choice implies Zorn's lemma.
(21) Let C, Y, and Z be sets. Define a map $\varphi: (X \times Y, Z) \to (X, (Y, Z))$ by setting, for $f: X \times Y \to Z$, $\varphi(f)(x)(y) = f((x, y))$.
(a) Prove that φ is always an isomorphism.
(b) If $g: Z \to Z'$ is a map, then for all sets W there is a map $g_W: (W, Z) \to (W, Z')$ defined by $g_W(f) = gf$. Prove that the following square is commutative:

$$\begin{array}{ccc} (X \times Y, Z) & \xrightarrow{\varphi} & (X, (Y, Z)) \\ \downarrow{\scriptstyle g_{X \times Y}} & & \downarrow{\scriptstyle (g_Y)_X} \\ (X \times Y, Z') & \xrightarrow{\varphi'} & (X, (Y, Z')) \end{array}$$

where $\varphi': (X \times Y, Z') \to (X, (Y, Z'))$ is defined as in (a).

(22) Show that if X and Y are arbitrary sets, then the sets $2^{X \times Y}$ and $(Y, 2^X)$ are isomorphic.
(23) Suppose that R is an order relation on a set X which is not a total ordering. Suppose, in particular, that x and y are elements of X such that neither x R y nor y R x holds. Show that there is an order relation R′ on X such that x R′ y and x_1 R′ x_2 holds if x_1 R x_2 holds. [*Hint*: Show that the relation R′ on X given by x_1 R′ x_2 if either x_1 R x_2 or both x_1 R x and y R x_2 hold is an order relation on X with the desired properties.]
(24) Let R be an order relation on a set X. Show that there is a total ordering R′ of X such that x_1 R x_2 implies x_1 R′ x_2.

Chapter 2 MONOIDS AND GROUPS

For the convenience of handy reference as well as to fix notation and definitions, we presented in the previous chapter a rapid survey of those aspects of set theory that we shall need in this book. In a similar spirit, some fundamental facts and notions concerning monoids and groups are presented in this chapter. Because, as in the case of set theory, we are assuming the reader is familiar with most of this material, few proofs are given in the text. Illustrative material as well as outlines of the more difficult proofs are included in the exercises to aid the reader in gaining familiarity with the few notions or results he encounters here.

1. MONOIDS

Certainly everyone will agree that adding, multiplying, or somehow combining two quantities to obtain a third is central to all of our experiences with algebra. These are all instances of what is called a binary law of composition, a notion we now define.

Definition

A binary law of composition on a set X is simply a map $m: X \times X \to X$. Two properties of a law of composition $m: X \times X \to X$ of particular interest are

(1) **associativity**, which means that $m(x_1, m(x_2, x_3)) = m(m(x_1, x_2), x_3)$ for all x_1, x_2, and x_3 in X;
(2) **commutativity**, which means that $m(x_1, x_2) = m(x_2, x_1)$ for all x_1 and x_2 in X.

It is not difficult to create laws of composition which are neither associative nor commutative. Although noncommutative laws of composition are of interest to us in this book, nonassociative ones are not. Therefore, from now on when we speak of laws of composition we shall always mean associative, but not necessarily commutative, ones.

We shall say that a law of composition $m: X \times X \to X$ on a set X has an identity if there is an element e in X, such that $m(e, x) = m(x, e) = x$ for all x in X. It is important to note that if the law of composition m has an identity, then it has only one. For if e_1, e_2 in X are both identity elements for m, then $e_1 = m(e_1, e_2) = e_2$. This uniquely determined element e in X is called **the identity of the law of composition** m.

We can now define the most general type of algebraic object of interest to us.

Definition

By a **monoid structure** on a set X we mean an associative law of composition $m: X \times X \to X$ with identity. A **monoid** is a pair (X, m) where X is a set and m is a monoid structure on X. If (X, m) is a monoid, then X is called the **underlying set** of the monoid (X, m). Finally, a monoid (X, m) is called a **commutative** or **abelian monoid** if m is a commutative law of composition on X.

We have already seen that the subsets of a set play an important role in set theory. The analogous notion for monoids is that of a submonoid which we now define.

Definition

A monoid (X', m') is said to be a **submonoid** of a monoid (X, m) if:

(a) X' is a subset of X containing the identity e of m.
(b) $m(x_1, x_2) = m'(x_1, x_2)$ for all x_1 and x_2 in X'. We shall denote the fact that (X', m') is a submonoid of the monoid (X, m) by writing $(X', m') \subset (X, m)$.

It is worth noting that if (X', m') is a submonoid of a monoid (X, m), then X' is a subset of X containing the identity e of m such that $m(x_1, x_2)$ is in X' if x_1 and x_2 are in X'. On the other hand, if X' is a subset of X containing the identity e of m such that $m(x_1, x_2)$ is in X' whenever x_1 and x_2 are in X', then (X', m') is a submonoid of (X, m) where $m': X' \times X' \to X'$ is defined by $m'(x_1, x_2) = m(x_1, x_2)$ for all x_1 and x_2 in X'. In other words, the submonoids of a monoid (X, m) are completely determined by the subsets X' of X containing e which have the property that $m(x_1, x_2)$ is in X' whenever x_1 and x_2 are in X'. For this reason, it is legitimate to refer to a submonoid (X', m') of a monoid (X, m) simply as the submonoid X' of (X, m).

If $m: X \times X \to X$ is a law of composition, then it is common practice to use either the multiplicative notation $x_1 x_2$ or additive notation $x_1 + x_2$ to denote $m(x_1, x_2)$. When there is no danger of ambiguity concerning which particular law of composition we mean, we shall tend to use the multiplicative notation for laws of composition. The additive notation shall be used only for commutative laws of composition. Of course, this does not preclude using the multiplicative notation for a commutative composition. Whenever we use the multiplicative (or additive)

notation for the law of composition of a monoid, we shall denote the identity of the monoid by 1 (or 0). Finally, we shall refer to a monoid (X, m) as simply the monoid X whenever there is no possible doubt as to which law of composition m on X we have in mind.

Before giving some important examples of monoids and submonoids, we state the following easily verified properties.

Basic Properties 1.1
Let $\mathcal{S}$ be a set of submonoids of the monoid X. Then:

(a) $\bigcap_{X' \in \mathcal{S}} X'$ is a submonoid of X.

(b) If $\mathcal{S}$ is a totally ordered subset of 2^X, then $\bigcup_{X' \in \mathcal{S}} X'$ is a submonoid of X.

Example 1.2 Let X be a set. Then the set of all maps (X, X) together with the law of composition $m : (X, X) \times (X, X) \to (X, X)$ given by $m(f_1, f_2) = f_1 f_2$, where $f_1 f_2$ is the composition $X \xrightarrow{f_2} X \xrightarrow{f_1} X$ of maps, is a monoid with identity element id_X. Since the maps from a set X to itself are called **endomorphisms** of X, this monoid is called the **monoid of endomorphisms** of X and is denoted by **End**(X). It is easily seen that the following are submonoids of $\mathrm{End}(X)$.

(a) $\mathrm{Inj}(X)$, the set of all injective endomorphisms of X.
(b) $\mathrm{Sur}(X)$, the set of all surjective endomorphisms of X.
(c) $\mathrm{Aut}(X)$, the set of all **automorphisms** of X, that is, isomorphisms from X to X.

The reader should also observe that $\mathrm{Aut}(X) = \mathrm{Inj}(X) \cap \mathrm{Sur}(X)$.

Example 1.3 A little thought should suffice to convince the reader that the set **N** of nonnegative integers is a commutative monoid under the usual addition of integers whose identity is 0 and satisfying:

(a) If $x + y = z + y$, then $x = z$ for all integers x, y, z in **N**.
(b) There is an element t different from 0 in N with the property that a subset N' of **N** is all of **N** if 0 is in N' and $x + t$ is in N' whenever x is in N'.
(c) Further, the element t has the property that $x + t \neq 0$ for all x in **N**.

Condition (a) is obvious and (b) is the familiar induction principle which we see by letting $t = 1$. What is perhaps not so obvious is that conditions (a), (b), and (c) completely determine the monoid of positive integers under addition, a fact we shall establish in the next section. In the meantime we will accept conditions (a), (b), and (c) as axioms for the monoid **N** of nonnegative integers under addition.

We now describe the ordinary order relation on **N** in terms of these axioms. If x and y are in N, then define $x \leq y$ if there is a z in N such that $x + z = y$. It is not difficult to show just using axioms (a), (b), and (c) that this defines an ordering on **N** satisfying:

(i) If $s \leq t$ and $u \leq v$, then $s + u \leq t + v$.
(ii) If $0 \leq x \leq t$, then $x = 0$ or t.
(iii) **N** is a totally ordered set under this ordering.
(iv) **N** is a well-ordered set with 0 the first element of N.

Since property (ii) implies there is only one element t in **N** satisfying axiom

(b), we follow the usual practice of denoting this element by 1 which we call the element one of **N**.

Example 1.4 Since the ordinary product of nonnegative integers is associative and the product of nonnegative integers is again a nonnegative integer, the ordinary multiplication of nonnegative integers is also a law of composition on **N**. In fact, it is well known that **N** together with this law of composition is a commutative monoid with identity 1 having the property $(x + y)z = xz + yz$ for all x, y, and z in **N**. In this connection, the reader should not have too much difficulty showing that if **N** is a commutative monoid satisfying axioms (a), (b), and (c) above, then there is one and only one law of composition, which we write multiplicatively, satisfying:

(c) $1x = x = x1$ for all x in **N**,
(d) $(x + y)z = xz + yz$,

and that this uniquely determined law of composition is commutative. Thus, the axioms we gave for the additive monoid **N** of nonnegative integers also enables us to construct the law of composition on **N** corresponding to the ordinary multiplication of positive integers. Finally, one should observe that $0x = 0$ for all x in **N**. For we have $0 = 0 + 0$ and hence $0x = (0+0)x = 0x + 0x$. Therefore, $0x + 0 = 0x + 0x$, which by axiom (a), implies $0x = 0$.

2. MORPHISMS OF MONOIDS

In the previous chapter, we saw that maps between sets give a useful way of comparing sets. Since monoids are sets together with additional structure, namely, laws of composition, it seems reasonable that maps between the underlying sets of two monoids which are somehow compatible with their laws of composition should give useful ways of comparing the monoids. This approach leads to the notion of a morphism from one monoid to another which we now define precisely.

Definition

Suppose (X', m') and (X, m) are two monoids with identity elements e' and e, respectively. A **morphism** from (X', m') to (X, m) is a map $f: X' \to X$ satisfying:

(a) $f(e') = e$.
(b) $f(m'(x_1, x_2)) = m(f(x_1), f(x_2))$ for all x_1 and x_2 in X'.

The set of all morphisms from (X', m') to (X, m) will be denoted by $((X', m'), (X, m))$.

The reader should note that if we are given two monoids X' and X whose laws of composition are written multiplicatively, then a morphism from X' to X is a map of sets $f: X' \to X$ satisfying:

(a) $f(1) = 1$.
(b) $f(x_1x_2) = f(x_1)f(x_2)$ for all x_1 and x_2 in X'.

Similarly, if we write the laws of composition additively, then a morphism from X' to X is a map of sets $f: X' \to X$ satisfying:

(a) $f(0) = 0$.
(b) $f(x_1 + x_2) = f(x_1) + f(x_2)$ for all x_1 and x_2 in X'.

When we denote the two monoids (X', m') and (X, m) by their underlying sets X' and X, then we denote the set of morphisms from X' to X by $\mathrm{Morph}(X', X)$ instead of $((X', m'), (X, m))$.

We now illustrate this notion with some important examples.

Example 2.1 Let (X', m') be a submonoid of (X, m). Then the inclusion map $\mathrm{inc}: X' \to X$ is a morphism from (X', m') to (X, m) called the **inclusion morphism** and is denoted by $\mathrm{inc}: (X', m') \to (X, m)$. In particular, the identity map $\mathrm{id}_X: X \to X$ is a morphism from (X, m) to (X, m) which is called the **identity morphism** and is denoted by $\mathrm{id}_{(X,m)}$ or more simply, id_X.

Example 2.2 Suppose X is a multiplicative monoid, that is, its law of composition is written multiplicatively. Associated with each element x in X are the maps $l_x: X \to X$, called left multiplication by x and defined by $l_x(y) = xy$ for all y in X, and $\rho_x: X \to X$, called right multiplication by x and defined by $\rho_x(y) = yx$ for all y in X. Thus, associated with a monoid X are the two maps $L: X \to \mathrm{End}(X)$ defined by $L(x) = l_x$ for all x in X and $R: X \to \mathrm{End}(X)$ defined by $R(x) = \rho_x$ for all x in X where $\mathrm{End}(X)$ is the monoid consisting of all endomorphisms of the set X. It is easily checked that $L: X \to \mathrm{End}(X)$ is a morphism from the monoid X to the monoid $\mathrm{End}(X)$. The map $R: X \to \mathrm{End}(X)$ is not a morphism but satisfies the condition $R(x_1x_2) = R(x_2)R(x_1)$. Obviously, though, $L = R$ if and only if X is a commutative monoid.

Example 2.3 Let $\mathbf{N}$ be the monoid of nonnegative integers under addition and X an arbitrary monoid. Then it is not very hard to show that two morphisms $f_1, f_2: \mathbf{N} \to X$ are the same if and only if $f_1(1) = f_2(1)$. What is more difficult to show (see the exercises) is that given any element x in X, there is a morphism $f: \mathbf{N} \to X$ such that $f(1) = x$. Suppose for each x in X we denote the unique morphism $f: \mathbf{N} \to X$ such that $f(1) = x$ by f_x. Then it is fairly obvious that the two maps of sets $X \to \mathrm{Morph}(\mathbf{N}, X)$ and $\mathrm{Morph}(\mathbf{N}, X) \to X$ given respectively by $x \to f_x$ and $f \to f(1)$ are isomorphisms of sets which are inverses of each other.

An interesting aspect of Example 2.3 is that it is really a reformulation of the familiar notion of raising a number to an integral exponent. For suppose x is an element of a multiplicatively written monoid X and $f_x: \mathbf{N} \to X$ is the unique morphism such that $f_x(1) = x$. If we denote $f_x(n)$ by x^n, then we obtain the usual properties of exponentiation: $x^0 = 1$, $x^1 = x$, and $x^{(n_1+n_2)} = x^{n_1}x^{n_2}$. For this reason we will usually use the notation x^n to denote $f_x(n)$ for all n in $\mathbf{N}$ and x in X.

Of course, in order to completely justify this definition of exponentiation we should show that $(x^{n_1})^{n_2} = x^{n_1 n_2}$ for all n_1, n_2 in $\mathbf{N}$ and x in X. To do this it suffices to show that for a fixed x in X and n_1 in $\mathbf{N}$ we have $(x^{n_1})^n = x^{n_1 n}$ for all n in $\mathbf{N}$. This can be easily carried out by induction on n. If $n = 0$, we have $(x^{n_1})^0 = 1$ while $x^{(n_1 0)} = x^0 = 1$. Suppose for $n \geq 0$ we have that $(x^{n_1})^n = x^{(n_1 n)}$. Then $(x^{n_1})^{(n+1)} = (x^{n_1})^n \cdot x^{n_1} =$

$x^{n_1 n} \cdot x^{n_1} = x^{(n_1 n + n_1)} = x^{n_1(n+1)}$. Hence, we have shown that $(x^{n_1})^n = x^{(n_1 n)}$ for all n in $\mathbf{N}$, which is our desired result.

Suppose X is a commutative multiplicative monoid. If x_1 and x_2 are elements of X, then it is not difficult to check that the map $g: \mathbf{N} \to X$ given by $g(n) = f_{x_1}(n) f_{x_2}(n)$ is a morphism of monoids. Because $g(1) = f_{x_1}(1) f_{x_2}(1) = x_1 x_2$, it follows that $f_{x_1 x_2}(n) = f_{x_1}(n) f_{x_2}(n)$ for all n in $\mathbf{N}$. Thus, we obtain in this case the usual formula $(x_1 x_2)^n = x_1^n x_2^n$ for all n in $\mathbf{N}$.

Suppose now that X is a commutative, additive monoid. For each element x in X, let $f_x: \mathbf{N} \to X$ be the unique morphism of monoids such that $f_x(1) = x$. Then we will usually denote $f_x(n)$ by nx for all n in $\mathbf{N}$. We have the usual rules $0x = 0$, $1x = x$, $(n_1 + n_2)x = n_1 x + n_2 x$, and $(n_1 n_2)x = n_1(n_2 x)$ for all n_1 and n_2 in $\mathbf{N}$. Further, as above for x_1 and x_2 in X, we obtain $n(x_1 + x_2) = nx_1 + nx_2$ for all n in $\mathbf{N}$.

As an application of these ideas we cite the following.

Example 2.4 Let X be a set and $\mathrm{End}(X)$ the monoid of endomorphisms of X. Then for each f in $\mathrm{End}(X)$ and each nonnegative integer n, the endomorphism f^n is called the **nth iterate** of f. The nth iterate of f is the formal way of expressing the endomorphism of X which is the composition of f with itself n times. Notice that $f^0 = \mathrm{id}_X$.

We now list some easily verified properties.

Basic Properties 2.5
Let $f: X \to Y$ be a morphism of monoids.

(a) If X' is a submonoid of X, then $f(X')$ is a submonoid of Y. If X' is commutative monoid, then so is $f(X')$.
(b) If Y' is a submonoid of Y, then $f^{-1}(Y')$ is a submonoid of X.
(c) Suppose that $g: Y \to Z$ is a morphism of monoids, then the composition of maps of sets $gf: X \to Z$ is a morphism of monoids.

This last basic property suggests the following.

Definition
Let $f: X \to Y$ and $g: Y \to Z$ be morphisms of monoids. The **composition** of f followed by g is defined to be the morphism $gf: X \to Z$ given by the composition of the maps of sets f followed by g.

It is obvious that the associativity of the composition of maps of sets implies the associativity of the composition of morphisms of monoids.

3. SPECIAL TYPES OF MORPHISMS

In this section we develop for monoids the analog of the notions of isomorphic, surjective, and injective maps already given for maps of sets.

We begin by pointing out the following characterization of the identity morphisms of monoids which is the exact analog of the characterization given for identity maps of sets.

Basic Property 3.1
For a morphism $f: X \to X$ of monoids, the following are equivalent:

(a) f is the identity morphism on X.
(b) If $g: X \to Y$ is an arbitrary morphism of monoids, then $gf = f$.
(c) If $h: U \to X$ is an arbitrary morphism of monoids, then $fh = h$.

Again in analogy with the situation for sets, we define an isomorphism of monoids as follows.

Definition
A morphism $f: X \to Y$ of monoids is an **isomorphism** if there is a morphism of monoids $g: Y \to X$ such that $gf = \mathrm{id}_X$ and $fg = \mathrm{id}_Y$. The monoid X is said to be **isomorphic** to the monoid Y if there is an isomorphism $f: X \to Y$.

Basic Properties 3.2

(a) For each monoid X the identity morphism is an isomorphism.
(b) If $f: X \to Y$ is an isomorphism of monoids, then there is only one morphism $g: Y \to X$ such that $gf = \mathrm{id}_X$ and $fg = \mathrm{id}_Y$. This uniquely determined morphism $g: Y \to X$ is also an isomorphism which is called the **inverse** of f and is denoted by f^{-1}. Clearly, $(f^{-1})^{-1} = f$.
(c) If $f: X \to Y$ and $g: Y \to Z$ are isomorphisms of monoids, then $gf: X \to Z$ is also an isomorphism of monoids with $(gf)^{-1} = f^{-1}g^{-1}$.
(d) If $f: X \to Y$ and $g: Y \to Z$ are morphisms such that $gf: X \to Z$ is an isomorphism, then f is an isomorphism if and only if g is an isomorphism.
(e) Suppose X, Y, and Z are monoids. Then:
  (i) X is isomorphic to Y if and only if Y is isomorphic to X.
  (ii) If X is isomorphic to Y and Y is isomorphic to Z, then X is isomorphic to Z.

An obvious question to ask at this point is how the fact that a morphism of monoids $f: X \to Y$, which is an isomorphism of monoids, is related to its being an isomorphism when viewed solely as a map of the underlying sets of X and Y. This question is answered in the following easily verified proposition.

Proposition 3.3
Let X and Y be monoids.

(a) Suppose $f: X \to Y$ is a bijective map of the underlying sets of X and Y. Then f is a morphism of monoids if and only if the inverse map $f^{-1}: Y \to X$ of the underlying sets of Y and X is a morphism of monoids. Consequently:
(b) A morphism of monoids $f: X \to Y$ is an isomorphism of monoids if and only if it is an isomorphism when viewed solely as a map of the underlying sets of X and Y. Hence:
(c) A morphism $f: X \to Y$ of monoids is an isomorphism of monoids if and only if:
  (i) Given y in Y, there is an x in X such that $f(x) = y$.
  (ii) If x_1 and x_2 are in X and $f(x_1) = f(x_2)$, then $x_1 = x_2$.

Finally:

(d) If the morphism of monoids $f: X \to Y$ is an isomorphism, its inverse as a morphism of monoids coincides with its inverse as a map of sets.

To illustrate some of the material developed until now, we substantiate our earlier claim that the monoid $(\mathbf{N}, +)$ of nonnegative integers under ordinary addition is completely described by the axioms we gave for $(\mathbf{N}, +)$.

Theorem 3.4

Suppose N and N' are commutative monoids (which we write additively) which satisfy the axioms for the nonnegative integers under addition. Then there is one and only one isomorphism N to N'.

PROOF: We first show that N and N' are isomorphic monoids. We have already stated in Section 2 that if M is any commutative monoid satisfying the axioms for the additive monoid of nonnegative integers and X is an arbitrary monoid, then given x in X there is precisely one morphism $f_x: M \to X$ such that $f(1) = x$ where 1 is the one in M. Hence, in particular, there are unique morphisms $f: N \to N'$ and $g: N' \to N$ such that $f(1) = 1'$ and $g(1') = 1$ where 1 is the one in N and $1'$ is the one in N'. Therefore, the compositions gf and fg have the properties that they are endomorphisms respectively of N and N' such that $gf(1) = 1$ and $fg(1') = 1'$. Since the endomorphisms of N and N' are completely determined by their values on 1 and $1'$, respectively, the fact that the endomorphisms gf and fg have the property $gf(1) = 1 = \mathrm{id}_N(1)$ and $fg(1') = 1' = \mathrm{id}_{N'}(1')$, it follows that $gf = \mathrm{id}_N$ and $fg = \mathrm{id}_{N'}$. Hence, $f: N \to N'$ is an isomorphism with inverse g.

We now show that if $h: N \to N'$ is an isomorphism of monoids, then $h(1) = 1'$. Because $f: N \to N'$ also has the property $f(1) = 1'$, it follows that $h = f$, which establishes that there is only one isomorphism from N to N', namely, the isomorphism f.

Suppose $h: N \to N'$ is an isomorphism. In order to show that $h(1) = 1'$ it suffices to prove that the element $h(1)$ of N' satisfies axiom (b) for the monoid of nonnegative integers. Namely, we must show that if X is a subset of N' which contains 0 and which contains $x + h(1)$ whenever it contains x, then $X = N'$, for we have already seen that the axioms for the monoid of nonnegative integers under addition implies that $1'$, the one of N', is the only element of N' with this property. Thus, if $h(1)$ does indeed have this property our claim that $h(1) = 1'$ will be verified.

Suppose X is a subset of N' containing 0 and such that $x + h(1)$ is in X whenever x is in X. Because $h(0) = 0$, we have that $0 \in h^{-1}(X)$. We also claim that if n is in $h^{-1}(X)$, then $n + 1$ is in $h^{-1}(X)$. For if n is in $h^{-1}(X)$, then $h(n+1) = h(n) + h(1)$ is also in X since $h(n)$ is in X and X has the property that $x + h(1)$ is in X whenever x is in X. Thus, $h^{-1}(X) = N$ since $h^{-1}(X)$ is a subset of N containing 0 which also contains $n + 1$ whenever it contains n. Because $h(h^{-1}(X)) \subset X$, it follows that $h(N) \subset X$. But $h(N) = N'$ because all isomorphisms are surjective maps. Therefore, we have our desired result that $X = N$. This completes the proof of the theorem.

Of course, the whole proof of this theorem depends on establishing the fact that if M is a monoid satisfying the axioms for the nonnegative integers and x is an

element of a monoid X, then there is a unique morphism of monoids $f_x: M \to X$ such that $f_x(1) = X$. The reader is reminded that a detailed outline of the proof of this fact is given in the exercises.

Returning to the general discussion of morphisms of monoids, we now define surjective and injective morphisms of monoids which are the exact analogs of the corresponding notions for maps of sets.

Definitions
Let $f: X \to Y$ be a morphism of monoids.

(a) f is said to be a **surjective** morphism if as a map of the underlying sets of X and Y it is surjective. In other words, f is a surjective morphism if and only if given any y in Y, there is an x in X such that $f(x) = y$.
(b) f is said to be an **injective** morphism if as a map of the underlying sets of X and Y it is injective. In other words, f is an injective morphism if and only if $f(x_1) = f(x_2)$ implies $x_1 = x_2$ for all x_1 and x_2 in X.

We also have the following analogs of the formal properties of surjective and injective maps of sets.

Basic Properties 3.5
Suppose $f: X \to Y$ and $g: Y \to Z$ are morphisms of monoids. Then:

(a) If f and g are surjective (injective) morphisms, then the morphism $gf: X \to Z$ is also a surjective (injective) morphism.
(b) If the composition $gf: X \to Z$ is a surjective morphism, then $g: Y \to Z$ is also a surjective morphism.
(c) If $gf: X \to Z$ is an injective morphism, then $f: X \to Y$ is also an injective morphism.

In dealing with sets, we have already seen that the notions of monomorphism and injective map coincide as do also the notions of epimorphism and surjective map. This is not quite the case for monoids. The reader will see in Section 9 that epimorphisms of monoids need not be surjective. However, the following connections between these concepts are valid.

Proposition 3.6
Let $f: X \to Y$ be a morphism of monoids.

(a) f is a monomorphism if and only if f is an injective morphism.
(b) If f is a surjective morphism, then f is an epimorphism.

PROOF: We only prove that if f is a monomorphism, then f is injective. The rest of the proposition is left as an exercise.

Suppose $f: X \to Y$ is a monomorphism and x_1 and x_2 are elements in X such that $f(x_1) = f(x_2)$. Let $h_1, h_2: \mathbf{N} \to X$ be the morphisms given by $h_1(n) = x_1^n$ and $h_2(n) = x_2^n$ for all n in $\mathbf{N}$. Then the compositions $fh_1, fh_2: \mathbf{N} \to Y$ have the property that $fh_1(1) = fh_2(1)$. Hence, $fh_1 = fh_2$ since a morphism from $\mathbf{N}$ to an arbitrary monoid is completely determined by its value on 1. Since $f: X \to Y$ is a monomorphism, this implies that $h_1 = h_2$. Therefore, $x_1 = h_1(1) = h_2(1) = x_2$. Consequently, f is injective.

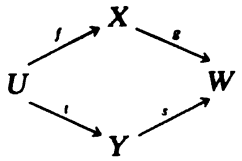
We now point out some easily demonstrable properties of surjective and injective morphisms of monoids that are analogs of properties of maps of sets. These are important because they are often useful in showing that analogs of results already obtained for maps of sets also hold for morphisms of monoids.

Proposition 3.7
Suppose X, Y, and Z are monoids and $f: X \to Y$ and $g: Y \to Z$ are maps of the underlying sets of the monoids involved such that the map $gf: X \to Z$ is a morphism of monoids.

(a) If $f: X \to Y$ is a surjective morphism of monoids, then the map $g: Y \to Z$ is also a morphism of monoids.
(b) If $g: Y \to Z$ is an injective morphism of monoids, then the map $f: X \to Y$ is also a morphism of monoids.

As an example of how these observations can be used, we establish for monoids the analog of the following result we have already obtained for sets.

Corollary 3.8
Suppose we are given a diagram of morphisms of monoids

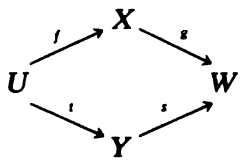


satisfying:

(a) $gf = st$.
(b) f is a surjective morphism and s is an injective morphism.

Then there is one and only one morphism of monoids $h: X \to Y$ which makes the diagram commutative, that is, such that $hf = t$ and $sh = g$.

PROOF: Viewing the diagram



simply as a diagram of maps of sets, the hypothesis that f is a surjective map and s is an injective map such that $gf = st$ implies that there is a unique map of sets $h: X \to Y$ such that $sh = g$ and $hf = t$. Since $shf = st$, it follows that shf is a morphism. Because s is an injective morphism and the composition $s(hf)$ is also a morphism, it follows from our previous proposition that the map hf is a morphism of monoids. Hence, by the same proposition, the fact that f is a surjective morphism and hf is a morphism of monoids implies that the map h is actually a morphism of monoids. Now it is not difficult to check that the morphism $h: X \to Y$ has our desired properties and is the only morphism from X to Y having these properties.

4. ANALYSES OF MORPHISMS

We recall that a map of sets always has an analysis, that is, a factorization into a surjective map followed by an injective map, and that any two analyses are uniquely "isomorphic." Our purpose in this section is to establish analogous results for morphisms of monoids.

Suppose $f: X \to Y$ is a morphism of monoids. We have already seen that the subset $f(X)$ of Y is a submonoid of Y. This submonoid of Y is called the image of f and is often denoted by $\operatorname{Im} f$. Further, the map $f_0: X \to \operatorname{Im} f$ defined by $f_0(x) = f(x)$ for all x in X is obviously a surjective morphism. Because the inclusion morphism $\operatorname{inc}: \operatorname{Im} f \to Y$ is an injective map, the fact that $f = \operatorname{inc} f_0$ shows that every morphism of monoids can be written as the composition of a surjective morphism followed by an injective morphism. We summarize this discussion in the following.

Definitions

Let $f: X \to Y$ be a morphism of monoids. Then $\operatorname{Im} f$ is a submonoid of Y called the **image of** f. Further, the map $f_0: X \to \operatorname{Im} f$ is a surjective morphism while the inclusion morphism $\operatorname{inc}: \operatorname{Im} f \to Y$ is an injective morphism. Finally, the representation of f as the composition of morphisms

$$X \xrightarrow{f_0} \operatorname{Im} f \xrightarrow{\operatorname{inc}} Y$$

is called the **image analysis of** f. More generally, any representation of f as the composition of morphisms of monoids hg with g a surjective morphism and h an injective morphism is called an **analysis of** f.

Using Corollary 3.8 it is easy to show that analyses of morphisms of monoids are unique in exactly the same sense that analyses of maps are unique. Specifically we have the following.

Basic Property 4.1

Let

$$X \xrightarrow{g} U \xrightarrow{h} Y$$

$$X \xrightarrow{g'} U' \xrightarrow{h'} Y$$

be two analyses of the same morphism of monoids $f: X \to Y$. Then there exists one and only one morphism of monoids $j: U \to U'$ such that $jg = g'$ and $h'j = h$ and this uniquely determined morphism j is an isomorphism.

Having developed the notion of the image analysis of a morphism of monoids, we now discuss the coimage analysis of a morphism of monoids.

Suppose $f: X \to Y$ is a morphism of monoids. Viewing f as a map of the underlying sets of X and Y, we know that there is associated with the map f the partition $\operatorname{Coim} f$ of X whose elements are the subsets of X of the form $f^{-1}(y)$ for all y in $\operatorname{Im} f$. Suppose $f^{-1}(y_1)$ and $f^{-1}(y_2)$ are two elements of $\operatorname{Coim} f$. The fact that f is a morphism of monoids implies that if x_1 is in $f^{-1}(y_1)$ and x_2 is in $f^{-1}(y_2)$, then x_1x_2 is in $f^{-1}(y_1y_2)$. This is equivalent to saying that if we denote by $f^{-1}(y_1)f^{-1}(y_2)$ the set of all elements in X of the form x_1x_2 with x_1 and x_2 in $f^{-1}(y_2)$, then $f^{-1}(y_1)f^{-1}(y_2) \subset$

$f^{-1}(y_1y_2)$. This condition can be restated as follows: If the subsets X_1 and X_2 of X are elements of Coim f, then there is one and only one element X_3 in Coim f containing X_1X_2 where X_1X_2 is the set of all elements of X of the form x_1x_2 with x_1 in X_1 and x_2 in X_2.

This property of the partition Coim f of X suggests considering the map $m:\text{Coim } f\times\text{Coim } f\to\text{Coim } f$ where $m(X_1, X_2)$ is the unique element of Coim f containing the subset X_1X_2 of X. Now it is not difficult to check that the map $m:\text{Coim } f\times\text{Coim} f\to \text{Coim } f$ has the following interesting property. Suppose $k:X\to\text{Coim } f$ is the canonical map from X to the partition Coim f of X. Then $k(x_1x_2)=m(k(x_1), k(x_2))$ for all x_1 and x_2 in X. But this implies, as we shall see presently, that the map $m:\text{Coim } f\times\text{Coim } f\to\text{Coim } f$ is not only associative and hence a law of composition but also that $k(1)=[1]$ is the identity element for this law of composition. In other words, (Coim f, m) is a monoid with [1] as identity and the canonical map $k:X\to\text{Coim } f$ is a surjective morphism. The validity of these observations is an easy consequence of the following general proposition.

Proposition 4.2
Let X be a monoid, Y a set, and $f:X\to Y$ a surjective map.

(a) There is at most one monoid structure m on Y such that the map $f:X\to Y$ is a morphism from the monoid X to the monoid (Y, m).
(b) This unique monoid structure exists on Y if and only if there is a map $m:Y\times Y\to Y$ such that $f(x_1x_2)=m(f(x_1), f(x_2))$ for all x_1 and x_2 in X. Such a map m, when it exists, is the unique monoid structure on Y making $f:X\to Y$ a morphism of monoids.

Thus, we see that the map $m:\text{Coim } f\times\text{Coim } f\to\text{Coim } f$ defined above is the unique monoid structure on Coim f such that the canonical surjective map $k_{\text{Coim } f}:X\to\text{Coim } f$ is a morphism of monoids. It is not difficult to check directly that the injective map $j_f:\text{Coim } f\to Y$ is a morphism of monoids. However, it is worthwhile noting that this also follows from the fact that $f=j_fk_{\text{Coim } f}$. For we have already seen that the composition of maps $f=j_fk_{\text{Coim } f}$ being a morphism of monoids together with the fact that $k_{\text{Coim } f}$ is a surjective morphism, implies that the map j_f is a morphism of monoids (see Proposition 3.7). Finally, from this it follows that the bijective map $(j_f)_0:\text{Coim } f\to\text{Im } f$ is an isomorphism of monoids.

We now summarize this discussion in the following.

Definitions
Let $f:X\to Y$ be a morphism of monoids. Then the set Coim f together with the unique monoid structure on Coim f which makes the canonical map $k_{\text{Coim } f}:X\to$ Coim f a morphism, is called the **coimage of the morphism** f and is denoted by Coim f.

The monoid Coim f has the further property that the unique map $j_f:\text{Coim } f\to Y$ which gives the coimage analysis of the map f is also a morphism. The morphism $j_f:\text{Coim } f\to Y$ is called the morphism from Coim f to Y **induced by the morphism** $f:X\to Y$. Thus, the composition

$$X\xrightarrow{k_{\text{Coim } f}}\text{Coim } f\xrightarrow{j_f}Y$$

is an analysis of f called the **coimage analysis of** f.

Finally, the unique isomorphism $g: \operatorname{Coim} f \to \operatorname{Im} f$ such that $gk_{\operatorname{Coim} f} = f_0$ and $\operatorname{inc} g = j_f$ is simply the isomorphism $(j_f)_0: \operatorname{Coim} f \to \operatorname{Im} f$ which is called the **canonical isomorphism from Coim f to Im f.**

5. DESCRIPTION OF SURJECTIVE MORPHISMS

As in the case of maps of sets, a morphism $f: X \to Y$ of monoids is surjective if and only if the injective morphism $j_f: \operatorname{Coim} f \to Y$ in the coimage analysis of f is surjective or what is the same thing, an isomorphism. This suggests that connections analogous to those already obtained for surjective maps of sets should exist between arbitrary surjective morphisms of monoids $f: X \to Y$ and their associated surjective morphisms of monoids $k_{\operatorname{Coim} f}: X \to \operatorname{Coim} f$. In fact, as we show in this section, the results along these lines for monoids are identical with those obtained for sets once one decides which partitions of the underlying set of a monoid are sufficiently compatible with the monoid structure to be properly considered partitions of the monoid itself and not just of its underlying set.

We have already seen that if $f: X \to Y$ is a morphism, then the partition Coim f is not arbitrary but satisfies the condition that if X_1 and X_2 are subsets of X in Coim f, then there is one and consequently only one subset X_3 of X in Coim f such that $X_1X_2 \subset X_3$. This suggests that in dealing with a monoid X, one should consider only those partitions $\mathcal{P}$ of X which satisfy the condition that if X_1 and X_2 are subsets of X in $\mathcal{P}$, then there is one (and consequently only one) subset X_3 of X in $\mathcal{P}$ such that $X_1X_2 \subset X_3$. The appropriateness of this remark is reinforced by the following.

Proposition 5.1
The following conditions are equivalent for a partition $\mathcal{P}$ of the underlying set of a monoid X:

(a) If X_1 and X_2 are elements of $\mathcal{P}$, then there is one (and consequently only one) element X_3 in $\mathcal{P}$ containing X_1X_2.
(b) There exists one (and consequently only one) map $m: \mathcal{P} \times \mathcal{P} \to \mathcal{P}$ such that the canonical surjective map $k_{\mathcal{P}}: X \to \mathcal{P}$ has the property $k_{\mathcal{P}}(x_1x_2) = m(k_{\mathcal{P}}(x_1), k_{\mathcal{P}}(x_2))$ for all x_1 and x_2 in X.
(c) There exists one (and consequently only one) monoid structure on $\mathcal{P}$ such that the canonical surjective map $k_{\mathcal{P}}: X \to \mathcal{P}$ is a morphism of monoids.

In the light of this discussion it is reasonable to make the following.

Definitions
A **partition of a monoid** X is a partition $\mathcal{P}$ of the underlying set of X which has the following property: If X_1 and X_2 are elements of $\mathcal{P}$, there is one (and consequently only one) element X_3 in $\mathcal{P}$ containing X_1X_2.

If $\mathcal{P}$ is a partition of a monoid X, define the map $m: \mathcal{P} \times \mathcal{P} \to \mathcal{P}$ by letting $m(X_1, X_2)$ be the unique element of $\mathcal{P}$ containing the product X_1X_2. This map is called the **canonical monoid structure** on $\mathcal{P}$ since it is the unique monoid structure on P which makes the canonical surjective map $k_{\mathcal{P}}: X \to \mathcal{P}$ a morphism of monoids.

If $\mathcal{P}$ is a partition of a monoid X, then we shall denote the monoid $(\mathcal{P}, m)$ consisting of the set $\mathcal{P}$ together with the canonical monoid structure m simply by $\mathcal{P}$.

It should be noted that if $\mathcal{P}$ is a partition of the monoid X, then the canonical monoid structure on $\mathcal{P}$ is completely described by the appealing formula $[x_1][x_2] = [x_1x_2]$ for all x_1 and x_2 in X. In fact, this formulation has so much appeal that it is the way we will describe the canonical monoid structure on $\mathcal{P}$ from now on.

In view of the correspondence between partitions and equivalence relations on a set, it is reasonable to ask which equivalence relations on the underlying set of a monoid correspond to the partitions of the monoid. This is answered in the following.

Proposition 5.2
Let R be an equivalence relation on the underlying set of a monoid X. Then the following are equivalent:

(a) The partition X/R of the underlying set of X is a partition of the monoid X.
(b) If $x_1 \mathrm{R}\, x_2$ holds, then $xx_1 \mathrm{R}\, xx_2$ and $x_1x \mathrm{R}\, x_2x$ both hold for x_1, x_2, and x in X.
(c) If $x_1 \mathrm{R}\, x_2$ and $x_1' \mathrm{R}\, x_2'$ are true, then $x_1x_1' \mathrm{R}\, x_2x_2'$ is true for all x_1, x_2, and x_1', x_2' in X.

This leads to the following.

Definition
Let X be a monoid. An equivalence relation R on the underlying set of X is an **equivalence relation on the monoid** X if it satisfies the following condition:

If $x_1 \mathrm{R}\, x_2$ and $x_1' \mathrm{R}\, x_2'$ hold, then $x_1x_1' \mathrm{R}\, x_2x_2'$ also holds for x_1, x_2, x_1', x_2' in X.

If R is an equivalence relation on a monoid X, we shall denote simply by X/R the monoid $(X/\mathrm{R}, m)$ consisting of the partition X/R of the monoid X together with the canonical monoid structure m.

Having established what we mean by partitions and equivalence relations on monoids, our earlier results describing the surjective maps for sets can now be transcribed verbatim for monoids. In order to give a little variety we shall state the results for monoids in terms of equivalence relations rather than partitions.

Suppose X is a monoid. If R_1 and R_2 are equivalence relations on the monoid X with $\mathrm{R}_1 \leq \mathrm{R}_2$, then $X/\mathrm{R}_1 \geq X/\mathrm{R}_2$ and the canonical map $g_{X/\mathrm{R}_1,\, X/\mathrm{R}_2} : X/\mathrm{R}_1 \to X/\mathrm{R}_2$ is a surjective morphism of monoids. This surjective morphism $g_{X/\mathrm{R}_1,\, X/\mathrm{R}_2}$ is called the **canonical morphism** from X/R_1 to X/R_2. We now list some of the basic properties of the canonical morphisms $g_{X/\mathrm{R}_1,\, X/\mathrm{R}_2}$.

Basic Properties 5.3
Suppose R_1 and R_2 are equivalence relations on the monoid X.

(a) There is a morphism $h : X/\mathrm{R}_1 \to X/\mathrm{R}_2$ of monoids such that $hk_{X/\mathrm{R}_1} = k_{X/\mathrm{R}_2}$ if and only if $\mathrm{R}_1 \leq \mathrm{R}_2$.
(b) If $\mathrm{R}_1 \leq \mathrm{R}_2$, there is only one morphism of monoids $h : X/\mathrm{R}_1 \to X/\mathrm{R}_2$ such that $hk_{X/\mathrm{R}_1} = k_{X/\mathrm{R}_2}$, namely, the canonical morphism $g_{X/\mathrm{R}_1,\, X/\mathrm{R}_2}$.

(c) If $R_1 \leq R_2$, then $g_{X/R_1, X/R_2}: X/R_1 \to X/R_2$ is always a surjective morphism which is an isomorphism if and only if $R_1 = R_2$.
(d) $g_{X/R_1, X/R_1} = \mathrm{id}_{X/R_1}$.

We now state the main results concerning the connections between arbitrary surjective morphisms of monoids $f: X \to Y$ and their associated surjective morphisms $k_{\mathrm{Coim}\, f}: X \to \mathrm{Coim}\, f$.

Proposition 5.4
Let $f_1: X \to Y_1$ and $f_2: X \to Y_2$ be two surjective morphisms of monoids.

(a) The following statements are equivalent:
  (i) There is a morphism $h: Y_1 \to Y_2$ of monoids such that $hf_1 = f_2$.
  (ii) There is a morphism $g: \mathrm{Coim}\, f_1 \to \mathrm{Coim}\, f_2$ of monoids such that $gk_{\mathrm{Coim}\, f_1} = k_{\mathrm{Coim}\, f_2}$.
  (iii) $\mathrm{Coim}\, f_1 \geq \mathrm{Coim}\, f_2$.
(b) If there is a morphism $h: Y_1 \to Y_2$ such that $hf_1 = f_2$, then:
  (i) There is only one such morphism.
  (ii) There is only one morphism $g: \mathrm{Coim}\, f_1 \to \mathrm{Coim}\, f_2$ such that $gk_{\mathrm{Coim}\, f_1} = k_{\mathrm{Coim}\, f_2}$, namely, $g_{\mathrm{Coim}\, f_1, \mathrm{Coim}\, f_2}$.
(c) The following are equivalent:
  (i) There is an isomorphism of monoids $h: Y_1 \to Y_2$ such that $hf_1 = f_2$.
  (ii) $\mathrm{Coim}\, f_1 = \mathrm{Coim}\, f_2$.
(d) If Y is a partition of the monoid X and $f: X \to Y$ is the canonical morphism, then $\mathrm{Coim}\, f = Y$ and $f = k_{\mathrm{Coim}\, f}$.

As in the case of sets, the main content of this proposition is that all surjective morphisms of monoids $f: X \to Y$ with a fixed domain X are essentially given by the canonical morphisms $k_{\mathcal{P}}: X \to \mathcal{P}$ for all partitions $\mathcal{P}$ of the monoid X.

We end this section with the following generalization of some of the properties of surjective morphisms of monoids we have discussed so far. The reader should have no difficulty recognizing this as the exact analog of a result already established for surjective maps of sets (Chapter 1, Proposition 7.3).

Proposition 5.5
Let $f: X \to Y$ be a surjective morphism of monoids. If $g: X \to Z$ is an arbitrary morphism of monoids, then there is a morphism of monoids $h: Y \to Z$ such that $hf = g$ if and only if $R(f) \leq R(g)$. If $R(f) \leq R(g)$, then there is only one morphism $h: Y \to Z$ of monoids satisfying $hf = g$.

6. GROUPS AND MORPHISMS OF GROUPS

Although the general notion of a monoid is important in a good deal of mathematics, we shall be concerned in this book primarily with the special type of monoids called groups. Because groups are monoids, everything we have shown about monoids generally also holds for groups. However, there are special features of

the theory of groups which do not hold for arbitrary monoids. This section is devoted to outlining some of these special features.

Before defining groups, we discuss the notion of invertible elements of a monoid. An element x in a monoid (X, m) with identity element e is said to be invertible if there is an element y in X such that $m(x, y) = m(y, x) = e$. It is not difficult to show that if x in the monoid (X, m) is invertible there is only one element y in X with the property $m(x, y) = m(y, x) = e$.

Definitions
Let (X, m) be a monoid with identity e. An element x in X is said to be **invertible** if and only if there is an element y in X such that $m(x, y) = m(y, x) = e$. If x in X is invertible, then the unique element y in X such that $m(x, y) = m(y, x) = e$ is called the **inverse of** x. If we write the law of composition in the monoid (X, m) multiplicatively or additively, then the inverse of an invertible element x will be denoted by x^{-1} or $-x$, respectively.

As an immediate consequence of these definitions we have the following.

Basic Properties 6.1
Let X be a multiplicative monoid.

(a) The identity 1 of X is invertible with $1^{-1} = 1$.
(b) If x in X is invertible, then its inverse x^{-1} is also invertible and $(x^{-1})^{-1} = x$.
(c) If x and y are elements of X and xy is invertible, then x and y are both invertible in X.
(d) The set of invertible elements of X is a submonoid of X which we denote by $\mathrm{Inv}(X)$.
(e) If $f: X \to Y$ is a morphism of monoids and x is an invertible element of X, then $f(x)$ is invertible in Y and $f(x^{-1}) = f(x)^{-1}$. Hence:
(f) $f(\mathrm{Inv}(X)) \subset \mathrm{Inv}(Y)$.

By way of illustrating some of these points, we give the following examples.

Example 6.2 Let X be a set and $\mathrm{End}(X)$ the monoid of endomorphisms of X. Then an element f in $\mathrm{End}(X)$ is invertible if and only if f is an automorphism of X. Hence, $\mathrm{Inv}(\mathrm{End}(X)) = \mathrm{Aut}(X)$.

Example 6.3 Let X be a monoid and $L: X \to \mathrm{End}(X)$, the injective morphism given by $L(x) = l_x$, left multiplication by x, for all x in X. Then $L(\mathrm{Inv}(X)) \subset \mathrm{Aut}(X)$.

Example 6.4 Let $\mathbf{N}$ be the nonnegative integers. Then $\mathrm{Inv}((\mathbf{N}, +)) = \{0\}$ while $\mathrm{Inv}((\mathbf{N}, \times)) = \{1\}$.

We now give some definitions concerning groups.

Definitions

(a) A **group** is a monoid X with the property that every element of X is invertible, or, what is the same thing, $\mathrm{Inv}(X) = X$.

(b) A **subgroup** X' of a group X is a submonoid X' of X which is also a group.
(c) Suppose X and Y are groups. A **morphism** $g: X \to Y$ **of groups** is simply the same thing as a morphism from X to Y when X and Y are viewed as monoids.
(d) By a **partition (equivalence relation) of a group** X we mean a partition (equivalence relation) of X when viewed as a monoid.

The reader should have no difficulty checking the following.

Basic Properties 6.5
Let X be a group.

(a) A subset X' of X is a subgroup of X if and only if:
  (i) $1 \in X'$.
  (ii) If x is in X', then x^{-1} is also in X'.
  (iii) If x and y are in X', then xy is in X'.
(b) Suppose X and Y are groups. A map $f: X \to Y$ of the underlying sets of X and Y is a morphism of groups if and only if f satisfies $f(x_1x_2) = f(x_1)f(x_2)$ for all x_1, x_2 in X. Further, if $f: X \to Y$ is a morphism of groups, then $f(x^{-1}) = f(x)^{-1}$ for all x in X.
(c) Suppose $f: X \to Y$ is a surjective morphism of monoids with X a group. Then Y is a group.

Combining what has been shown about monoids with our discussion of groups, we obtain the following important facts concerning groups.

Proposition 6.6
Suppose X is a group.

(a) If $\mathcal{P}$ is a partition of the group X, then $\mathcal{P}$ with its canonical monoid structure is a group.
(b) Suppose Y is a monoid and $f: X \to Y$ a morphism of monoids.
  (i) If $X \xrightarrow{g} Z \xrightarrow{h} Y$ is an analysis of f, then Z is a group, g is a surjective morphism of groups, and h is an injective morphism of monoids.
  (ii) In particular, Im f and Coim f are groups, the morphisms $f_0: X \to \text{Im } f$ and $k_{\text{Coim } f}: X \to \text{Coim } f$ are surjective morphisms of groups. The inclusion morphism Im $f \to Y$ and the canonical morphism Coim $f \xrightarrow{i_f} Y$ are injective morphisms of monoids.
(c) If $X \xrightarrow{g} Z \xrightarrow{h} Y$ and $X \xrightarrow{g'} Z' \xrightarrow{h'} Y$ are analyses of the morphism f, then there is a unique morphism of groups $t: Z \to Z'$ such that $tg = g'$ and $h = h't$.

7. KERNELS OF MORPHISMS OF GROUPS

One of the basic differences between the theory of groups and the theory of arbitrary monoids is that it is much easier to describe the partitions of a group than it is to describe the partitions of an arbitrary monoid. In general, in order to describe a partition $\mathcal{P}$ on a set or monoid X it is necessary to describe each of the subsets of X individually. However, if X is a group, a partition $\mathcal{P}$ of the group X

can be completely described in terms of the subset $[1]_{\mathcal{P}}$ of X containing the identity 1 of X. Exactly how this is accomplished is made clear in the following.

Proposition 7.1
Let $f: X \to Y$ be a morphism of groups and let $K = f^{-1}(1)$. Then K has the following properties:

(a) K is a subgroup of X satisfying $xK = Kx$ for all x in X.
(b) If x_1 and x_2 are in X, then $f(x_1) = f(x_2)$ if and only if there is a $k \in K$ such that $x_1 k = x_2$. Hence, if x is in X, then $f^{-1}(f(x)) = xK$.
(c) Thus, the elements of the partition Coim f of X are precisely the subsets of X of the form xK for all x in X. Hence, the partition Coim f of X is completely determined by $K = [1]_{\text{Coim } f}$.

This result clearly indicates that subgroups K of a group X with the property $xK = Kx$ for all x in X play a fundamental role in studying groups. For this reason they are given a special name.

Definitions
A subgroup K of a group X is called a **normal** or **invariant subgroup** of X if $xK = Kx$ for all x in X. If $f: X \to Y$ is a morphism of groups, the normal subgroup $f^{-1}(1)$ of X is called the **kernel of** f and is denoted by Ker f.

Suppose now we are given a partition $\mathcal{P}$ of a group X. As we have already seen (Proposition 6.6), $\mathcal{P}$ is a group and the canonical map $k_{\mathcal{P}}: X \to \mathcal{P}$ is a surjective morphism of groups. Hence, our previous proposition shows:

(a) $K = [1]_{\mathcal{P}}$ is a normal subgroup of X.
(b) If x is in X, then $[x]_{\mathcal{P}} = xK$.
(c) Denoting $[x]_{\mathcal{P}}$ by xK, the law of composition in $\mathcal{P}$ takes on the form $(x_1K)(x_2K) = x_1x_2K$.
(d) For each x in X, we have $(xK)^{-1} = x^{-1}K$.

Hence, associated with each partition $\mathcal{P}$ of a group is the normal subgroup $[1]_{\mathcal{P}}$ of X which completely determines the partition $\mathcal{P}$. It is natural to ask in this connection if for each normal subgroup K of a group X is there a partition $\mathcal{P}$ of the group X such that $[1]_{\mathcal{P}} = K$? This question is answered in the affirmative in the following.

Proposition 7.2
Let K be a normal subgroup of the group X. Then:

(a) The set X/K of all subsets of X of the form xK with x in X is a partition of the group X.
(b) The canonical law of composition for the partition X/K of X is given by $(x_1K)(x_2K) = x_1x_2K$.
(c) The monoid X/K is a group with $(xK)^{-1} = x^{-1}K$ for all x in X.
(d) The canonical morphism of groups $k_{X/K}: X \to X/K$ given by $k_{X/K}(x) = xK$ for all x in X has the property Ker $k_{X/K} = K$.
(e) If $\mathcal{P}$ is a partition of the group X with $[1]_{\mathcal{P}} = K$, then $\mathcal{P} = X/K$.

To emphasize the point that partitions of a group X are completely determined by the normal subgroups of X, we introduce a new name and notation for the partitions of a group.

Definition
Suppose K is a normal subgroup of a group X. The partition of the group consisting of the subsets of X of the form xK for all x in X together with the law of composition $(x_1K)(x_2K) = x_1x_2K$ is a group which we denote by X/K. The group X/K is called either the **factor group of X by K** or the **residue class group of X by K**. The canonical surjective morphism of groups $k_{X/K}: X \to X/K$ is called the **canonical morphism** from the group X to the factor group X/K.

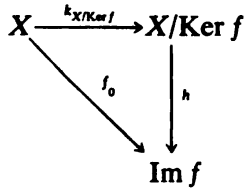
We now state in terms of this new terminology some of our previous results.

Proposition 7.3
Let X be a group.

(a) The partitions of the group X are precisely the factor groups X/K for all normal subgroups K of X.
(b) Suppose $f: X \to Y$ is a morphism of groups. Then $\operatorname{Coim} f = X/\operatorname{Ker} f$ and the coimage analysis of f

$$X \xrightarrow{k_{X/\operatorname{Ker} f}} X/\operatorname{Ker} f \xrightarrow{j_f} Y$$

can be described by $k_{X/\operatorname{Ker} f}(x) = xK$; and $j_f(xK) = f(x)$ for all x in X.
(c) Since $j_f: X/\operatorname{Ker} f \to Y$ is injective and $\operatorname{Im} j_f = \operatorname{Im} f$, we have that the induced morphism $(j_f)_0: X/\operatorname{Ker} f \to \operatorname{Im} f$ is an isomorphism of groups. The isomorphism $(j_f)_0$ can also be characterized as the unique morphism $h: X/\operatorname{Ker} f \to \operatorname{Im} f$ such that the diagram



commutes, that is, $hk_{X/\operatorname{Ker} f} = f_0$. Hence:
(d) The morphism of groups $f: X \to Y$ is a surjective morphism if and only if the morphism $j_f: X/\operatorname{Ker} f \to Y$ is surjective and therefore an isomorphism.
(e) The morphism of groups $f: X \to Y$ is injective if and only if $\operatorname{Ker} f = \{1\}$.
(f) If K is a normal subgroup of X and $k_{X/K}: X \to X/K$ the canonical morphism of groups, then $\operatorname{Ker}(k_{X/K}) = K$.

Now it is not difficult to see that if K_1 and K_2 are normal subgroups of a group X, then $K_1 \subset K_2$ if and only if X/K_1 is a refinement of X/K_2. Moreover, if $K_1 \subset K_2$, then the canonical morphism $g_{X/K_1, X/K_2}: X/K_1 \to X/K_2$ can be described by $g_{X/K_1, X/K_2}(xK_1) = xK_2$ for all x in X. Clearly, if $K_1 = K_2$, then $g_{X/K_1, X/K_2} = \mathrm{id}_{X/K_1}$.

Further, we have the following analogs of results already obtained earlier for sets and monoids.

Proposition 7.4
Suppose $f_1: X \to Y_1$ and $f_2: X \to Y_2$ are surjective morphisms of groups. Then:

(a) The following statements are equivalent:
  (i) There is a morphism of groups $h: Y_1 \to Y_2$ such that $hf_1 = f_2$.
  (ii) $\operatorname{Ker} f_1 \subset \operatorname{Ker} f_2$.
  (iii) There is a morphism $h': X/K_1 \to X/K_2$ such that $h'k_{X/K_1} = k_{X/K_2}$. The morphism h' is nothing more than the canonical morphism $g_{X/K_1, X/K_2}$.

(b) If there is a morphism $h: Y_1 \to Y_2$ such that $hf_1 = f_2$, then:
  (i) h is the unique morphism of groups with this property.
  (ii) h is a surjective morphism.
  (iii) h is an isomorphism if and only if $\operatorname{Ker} f_1 = \operatorname{Ker} f_2$.

What this proposition says in essence is that the surjective morphisms $f: X \to Y$ of groups with a fixed domain X are, roughly speaking, completely determined by the normal subgroups of X.

We devote the rest of this section to pointing out various important applications of the notion of the kernel of a morphism of groups.

We begin with the following generalization of our previous proposition.

Proposition 7.5

(a) If $X \xrightarrow{f} Y \xrightarrow{g} Z$ are morphisms of groups, then $\operatorname{Ker}(gf) = f^{-1}(\operatorname{Ker} g) \supset \operatorname{Ker} f$. Hence, if g is injective, then $\operatorname{Ker}(gf) = \operatorname{Ker} f$.

(b) If we are given a diagram of morphisms of groups

$$\begin{array}{ccc} X_1 & \xrightarrow{f} & Y \\ \downarrow{\scriptstyle \kappa} & & \\ X_2 & \xrightarrow{h} & Z \end{array}$$

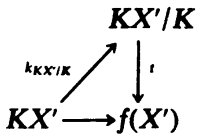
with f a surjective morphism, then there exists a morphism $i: Y \to Z$ which makes the above diagram commute, that is, $hg = if$, if and only if $g(\operatorname{Ker} f) \subset \operatorname{Ker} h$. Moreover, there is at most one such morphism from Y to Z.

In the next proposition we investigate the connection a morphism of groups $f: X \to Y$ establishes between the subgroups of Y and those of X.

Proposition 7.6
Suppose $f: X \to Y$ is a morphism of groups and $K = \operatorname{Ker} f$. Then:

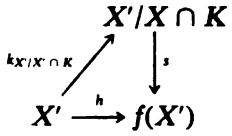
(a) If Y' is a subgroup of Y, then $f^{-1}(Y')$ is a subgroup of X containing K. Moreover, $f^{-1}(Y)$ is a normal subgroup of X if Y' is a normal subgroup of Y.

(b) Suppose X' is a a subgroup of X.
  (i) $f^{-1}(f(X'))$ is a subgroup KX' of X.
  (ii) Moreover, the morphism of groups $g: KX' \to f(X')$ given by $g(y) = f(y)$ for all y in KX' is a surjective morphism with kernel K. Thus, there is a unique isomorphism of groups $t: KX'/K \to f(X')$ which makes the diagram



commute.

(iii) The morphism of groups $h: X' \to f(X')$ given by $h(y) = f(y)$ for all y in X' is a surjective morphism with kernel $X' \cap K$. Hence, there is a unique isomorphism of groups $s: X'/X' \cap K \to f(X')$ which makes the diagram



commute.

(iv) Because $s: X'/X' \cap K \to f(X')$ and $t: X'K/K \to f(X')$ are isomorphisms of groups, we have that $t^{-1}s: X'/X' \cap K \to X'K/K$ is also an isomorphism of groups.

It is worth noting that although the morphism $f: X \to Y$ was used to obtain the isomorphism $t^{-1}s: X'/X' \cap K \to X'K/K$, this isomorphism in fact depends only on the subgroups X' and K of X and not on the morphism $f: X \to Y$. More precisely, it is nothing more than the canonical isomorphism we describe below which clearly has nothing to do with the morphism $f: X \to Y$.

Proposition 7.7

Suppose X' and K are subgroups of a group X with K a normal subgroup of X.

(a) $KX' = X'K$ is a subgroup of X. Clearly, K and X' are subgroups of KX' with K a normal subgroup.

(b) The morphism of groups $g: X' \to X'K/K$ given by $g(x) = xK$ for each x in X' is a surjective morphism with kernel $X' \cap K$. Hence:

(c) The morphism $j_g: X'/X' \cap K \to X'K/K$ induced by the surjective morphism $g: X' \to X'K/K$ is an isomorphism.

Because the isomorphism j_g we just described is used a great deal it is convenient to make the following definition.

Definition

Suppose X' and K are subgroups of a group X with K a normal subgroup of X. Then the isomorphism $j_g: X'/X' \cap K \to X'K/K$ induced by the surjective morphism $g: X' \to X'K/K$ given by $g(x) = xK$ for all x in X' is called the **canonical isomorphism from** $X'/X' \cap K$ to $X'K/K$. Unless stated otherwise, this is the only isomorphism we consider between $X'/X' \cap K$ and $X'K/K$.

As might be expected, if we assume that we are given a surjective morphism $f: X \to Y$, then much more can be said about the relationships between the subgroups of X and Y than we were able to say when $f: X \to Y$ was an arbitrary morphism of groups.

Proposition 7.8
Let $f: X \to Y$ be a surjective morphism of groups with kernel K.

(a) Suppose X' is a normal subgroup of X. Then:
  (i) $f(X')$ is a normal subgroup of Y.
  (ii) The morphism $g: X \to Y/f(X')$ of groups given by $g(x) = f(x)f(X')$ is a surjection with kernel $f^{-1}(f(X')) = X'K$. Hence:
  (iii) The morphism $j_g: X/X'K \to Y/f(X')$ induced by $g: X \to Y/f(X')$ is an isomorphism.
  (iv) The isomorphism $j_g: X/X'K \to Y/f(X')$ is the unique morphism from $X/X'K \to Y/f(X')$ which makes the diagram

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle k_{X/X'K}}\downarrow & & \downarrow{\scriptstyle k_{Y/f(X')}} \\ X/X'K & \longrightarrow & Y/f(X') \end{array}$$

  commute.

(b) Let $\mathcal{S}$ be the set of subgroups of X containing K and $\mathcal{T}$ the set of all subgroups of Y. Then the maps $\alpha: \mathcal{S} \to \mathcal{T}$ and $\beta: \mathcal{T} \to \mathcal{S}$ defined by $\alpha(X') = f(X')$ and $\beta(Y') = f^{-1}(Y')$ for all X' in $\mathcal{S}$ and Y' in $\mathcal{T}$ have the following properties:
  (i) α and β are bijective maps which are inverses of each other.
  (ii) X' in $\mathcal{S}$ is a normal subgroup of X if and only if $\alpha(X') = f(X')$ is a normal subgroup of Y. Similarly, Y' is a normal subgroup of Y if and only if $\beta(Y') = f^{-1}(Y')$ is a normal subgroup of X.
  (iii) If X' in $\mathcal{S}$ is normal in X, then the morphism of groups $g: X \to Y/f(X')$ given by $g(x) = f(x)f(X')$ for all x in X is surjective with kernel X'. Hence, the induced morphism $j_g: X/X' \to Y/f(X')$ is an isomorphism.

For ease of reference we make the following definition.

Definition
Let $f: X \to Y$ be a surjective morphism of groups with kernel K and suppose X' is a normal subgroup of X. Then the isomorphism $X/X'K \to Y/f(X')$, which is the unique morphism from $X/X'K \to Y/f(X')$ which makes the diagram

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle k_{X/X'K}}\downarrow & & \downarrow{\scriptstyle k_{Y/f(X')}} \\ X/X'K & \longrightarrow & Y/f(X') \end{array}$$

commute, is called the isomorphism from $X/X'K \to Y/f(X')$ **induced by** f.

A particularly important special case of the above are the surjective morphisms of the form $k_{X/K}: X \to X/K$ where K is a normal subgroup of X. Suppose K is a normal subgroup of X and X' is a subgroup of X. Then it is easily seen that $X'K$ is a subgroup of X which is a normal subgroup of X if X' is a normal

subgroup of X. Also the group $X'K/K$ is a subgroup of X/K which is a normal subgroup of X/K if and only if X', and hence $X'K$, is a normal subgroup of X. Thus, if X' is a normal subgroup of X, we obtain that $X/X'K$ and $(X/K)/(X'K/K)$ are isomorphic groups by means of the isomorphism $X/X'K \to (X/K)/(X'K/K)$ induced by the canonical surjective morphism $k_{X/K}: X \to X/K$. This isomorphism $X/X'K \to (X/K)/(X'K/K)$ is called the **canonical isomorphism** and is the only isomorphism between $X/X'K$ and $(X/K)/(X'K/K)$ we will consider. It should be noted that if $X' \supset K$, then $X'K = X'$ and we have the notationally appealing result that X/X' is isomorphic to $(X/K)/(X'/K)$ under the canonical isomorphism.

One last word concerning the canonical isomorphisms $X'/X' \cap K \to X'K/K$ and $X/X'K \to (X/K)/(X'K/K)$. We usually consider these isomorphisms as identifications. Therefore, we shall write $X'/X' \cap K = X'K/K$ and $X/X'K = (X/K)/(X'K)/K$ meaning that they are being identified by means of the canonical isomorphism or their inverses.

8. GROUPS OF FRACTIONS

We have already seen that associated with each monoid X is the group $\mathrm{Inv}(X)$ of invertible elements of X. It is not difficult to show that the inclusion morphism $\mathrm{inc}: \mathrm{Inv}(X) \to X$ has the property that given any group G and morphism of monoids $f: G \to X$, there is one and only one morphism of groups $g: G \to \mathrm{Inv}(X)$ such that $f = \mathrm{inc}\, g$. Moreover, this observation completely characterizes the group $\mathrm{Inv}(X)$ as we see in the following.

Proposition 8.1
Let $h: U \to X$ be a morphism of monoids with U a group. Suppose $h: U \to X$ has the property that given any group G and any morphism of monoids $f: G \to X$ there is one and only one morphism of groups $g: G \to U$ such that $f = hg$. Then $h: U \to X$ is an injective morphism with $\mathrm{Im}\, h = \mathrm{Inv}(X)$. Hence, $h_0: U \to \mathrm{Inv}(X)$ is an isomorphism of groups.

PROOF: Since U is a group, there is a unique morphism $g: U \to \mathrm{Inv}(X)$ such that $\mathrm{inc}\, g = h$ where $\mathrm{inc}: \mathrm{Inv}(X) \to X$ is the inclusion morphism. On the other hand, the hypothesis in $h: U \to X$ implies that there exists a unique morphism $g': \mathrm{Inv}(X) \to U$ such that $hg' = \mathrm{inc}$. From these relationships it follows that $hg'g = h$ and $\mathrm{inc}\, gg' = \mathrm{inc}$. But the morphism $h: U \to X$ has the property that there is only one morphism $t: U \to U$ such that $ht = h$. Since $g'g$ and id_U both have this property, it follows that $g'g = \mathrm{id}_U$. One can also show that $gg': \mathrm{Inv}(X) \to \mathrm{Inv}(X)$ is the $\mathrm{id}_{\mathrm{Inv}(X)}$ in a similar fashion or by using the fact that $\mathrm{inc}: \mathrm{Inv}(X) \to X$ is a monomorphism. Therefore, the morphism $g: U \to \mathrm{Inv}(X)$ is an isomorphism such that $\mathrm{inc}\, g = h$. From this it follows that h is a monomorphism and $\mathrm{Im}\, h = \mathrm{Inv}(X)$.

This characterization of the morphism $\mathrm{inc}: \mathrm{Inv}(X) \to X$ shows that the morphisms of a group G to X are uniquely determined by the morphisms of G to the group $\mathrm{Inv}(X)$. It is natural to ask if something similar can be done for the morphisms of monoids from X to a group G. Specifically, is there a group H

associated with the monoid X such that for each group G the morphisms of monoids from X to G are uniquely determined by the morphisms of groups from H to G? We now describe in what sense such a group H can be found.

Definition
Suppose X is a monoid. A group H together with a morphism $h: X \to H$ of monoids is called a **group of fractions** for X if for each group G and morphism of monoids $f: X \to G$ there is one and only one morphism of groups $g: H \to G$ such that $f = gh$.

The reader should have no difficulty in convincing himself that if a monoid X has a group of fractions $h: X \to H$, then the group H has the property that the morphisms from X to a group G are completely determined by the morphisms from the group H to G. For it follows from the definition of a group of fractions that the morphisms from X to G can be written uniquely as the compositions $X \xrightarrow{h} H \xrightarrow{g} G$ as g runs through group morphism from H to G. Hence, our original problem will be solved for a monoid X if we can show it has a group of fractions. The main result in this connection is the following.

Theorem 8.2
Let X be a monoid. Then:

(a) X has a group of fractions.
(b) If $h_1: X \to H_1$ and $h_2: X \to H_2$ are two groups of fractions for X, then there are unique morphisms of groups $t_1: H_1 \to H_2$ and $t_2: H_2 \to H_1$ such that $t_1 h_1 = h_2$ and $t_2 h_2 = h_1$. Further, these uniquely determined morphisms t_1 and t_2 are isomorphisms which are inverses of each other.

In other words, each monoid has an essentially unique group of fractions. Although this theorem is of interest for arbitrary monoids, we will prove it just for commutative monoids because this is the only case we will have need of in this book. Also, the proof in the commutative case is somewhat simpler. However, before proving the existence part of the theorem for commutative monoids, we prove part (b), the uniqueness statement, for arbitrary, not just commutative, monoids.

PROOF: (b) Suppose $h_1: X \to H_1$ and $h_2: X \to H_2$ are two groups of fractions for the monoid X. Because $t_1: X \to H_1$ is a morphism from X to a group we know by the definition of a group of fractions that there is a unique morphism $t_1: H_1 \to H_2$ of groups such that $h_2 = t_1 h_1$. Similarly, there is a unique morphism $t_2: H_2 \to H_1$ of groups such that $h_1 = t_2 h_2$.

We show that these unique morphisms are isomorphisms by showing that $t_2 t_1: H_1 \to H_1$ is the identity on H_1 and $t_1 t_2: H_2 \to H_2$ is the identity on H_2. First we observe that $(t_2 t_1) h_1 = h_1$, because $(t_2 t_1) h_1 = t_2 (t_1 h_1) = t_2 h_2 = h_1$. But this implies that $t_2 t_1 = \mathrm{id}_{H_1}$ because, by the definition of a group of fractions, there is only one morphism $f: H_1 \to H_1$ such that $f_1 h_1 = h_1$ and obviously $\mathrm{id}_{H_1} h_1 = h_1$. By symmetry it follows that $t_1 t_2: H_2 \to H_2$ is the identity on H_2. Hence, t_1 and t_2 are isomorphisms which are inverses of each other.

In showing that commutative monoids have groups of fractions, the notion of the product of two monoids comes up.

Definitions
Let X and Y be two monoids. The **product** $X \times Y$ is the monoid whose underlying set is the Cartesian product $X \times Y$ and whose law of composition is given by $(x_1, y_1)(x_2, y_2) = (x_1x_2, y_1y_2)$ for all x_1, x_2 in X and y_1, y_2 in Y. Clearly, $(1, 1)$ is the identity of $X \times Y$.

The maps $i_X: X \to X \times Y$ and $i_Y: Y \to X \times Y$ given by $i_X(x) = (x, 1)$ for all x in X and $i_Y(y) = (1, y)$ for all y in Y are injective morphisms of monoids called the **injections** of X and Y into $X \times Y$.

The projection maps $p_X: X \times Y \to X$ given by $p_X(x, y) = x$ and $p_Y: X \times Y \to Y$ given by $p_Y(x, y) = y$ are surjective morphisms called the projections of $X \times Y$ onto X and Y, respectively.

Basic Properties 8.3
Let X and Y be monoids.

(a) The injection morphisms $i_X: X \to X \times Y$ and $i_Y: Y \to X \times Y$ have the following properties:
  (i) $i_X(x)i_Y(y) = i_Y(y)i_X(x)$ for all x in X and $y \in Y$.
  (ii) If z is in $X \times Y$, then there are unique elements x in X and y in Y such that $i_X(x)i_Y(y) = z$.

(b) $X \times Y$ is commutative if and only if both X and Y are commutative.

In order to show that commutative monoids have groups of fractions, we need the following description of the morphisms from the monoid $X \times Y$ to an arbitrary monoid Z in terms of the morphisms from X and Y to Z.

Proposition 8.4
Let X, Y, and Z be monoids. Associated with each morphism $f: X \times Y \to Z$ are the morphisms $fi_X: X \to Z$ and $fi_Y: Y \to Z$.

(a) For each morphism $f: X \times Y \to Z$ the morphisms fi_X and fi_Y have the property that $z_1z_2 = z_2z_1$ for each z_1 in $\mathrm{Im}(fi_X)$ and z_2 in $\mathrm{Im}(fi_Y)$.

(b) Two morphisms $f_1, f_2: X \times Y \to Z$ are the same if and only if $f_1i_X = f_2i_X$ and $f_1i_Y = f_2i_Y$.

(c) If $g: X \to Z$ and $h: Y \to Z$ are two morphisms of monoids such that $z_1z_2 = z_2z_1$ for all z_1 in $\mathrm{Im}\, g$ and z_2 in $\mathrm{Im}\, h$, then there is one and only one morphism $f: X \times Y \to Z$ such that $fi_X = g$ and $fi_Y = h$.

(d) If Z is a commutative monoid, then the map $(X \times Y, Z) \to (X, Z) \times (Y, Z)$ given by $f \to (fi_X, fi_Y)$ is an isomorphism from the set $(X \times Y, Z)$ of all morphisms from $X \times Y$ to Z to the product $(X, Z) \times (Y, Z)$ of the sets of all morphisms (X, Z) and (Y, Z) from X and Y to Z, respectively.

With these preliminary results concerning products of monoids out of the way, we return to our problem of constructing a group of fractions for a commutative monoid X. In order to motivate the actual construction of this group

of fractions, we first show that with each morphism from a commutative monoid X to a group G, there is associated a morphism from $X \times X$ to G. The definition of this associated morphism is based on the following.

Lemma 8.5
Let Y be a commutative submonoid of the group G and let Y^{-1} be the subset of G consisting of all y^{-1} with y in Y. Then:

(a) Y^{-1} is a commutative submonoid of G.
(b) The map $Y \to Y^{-1}$ given by $y \to y^{-1}$ is an isomorphism of monoids.
(c) The elements of Y and Y^{-1} commute, that is, if g_1 is in Y and g_2 is in Y^{-1}, then $g_1 g_2 = g_2 g_1$.
(d) The subset YY^{-1} consisting of all elements in G of the form $g_1 g_2$ with g_1 in Y and g_2 in Y^{-1} is a commutative subgroup of G.

Suppose X is a commutative monoid and $f: X \to G$ is a morphism of monoids with G a group. Then $Y = \operatorname{Im} f$ is a commutative submonoid of G because X is a commutative monoid. From Lemma 8.5 it follows that Y^{-1} is a commutative submonoid of G and the morphism $t: Y \to Y^{-1}$ given by $t(y) = y^{-1}$ for all y in Y is an isomorphism of monoids. Hence, associated with the morphism $f: X \to G$ is the composite morphism

$$X \xrightarrow{h} Y \xrightarrow{t} Y^{-1} \xrightarrow{\text{inc}} G$$

which can be described more directly as the morphism $g: X \to G$ given by $g(x) = f(x)^{-1}$ for all x in X. Because $\operatorname{Im} g = Y^{-1}$ and $\operatorname{Im} f = Y$, it follows from Lemma 8.5 that the elements of $\operatorname{Im} g$ and $\operatorname{Im} f$ commute. Hence, we know that there is one and only one morphism $\tilde{f}: X \times X \to G$ such that $\tilde{f}(x, 1) = f(x)$ and $\tilde{f}(1, x) = g(x)$ for all x in X. Tracing through the various definitions we see that $\tilde{f}: X \times X \to G$ can be described more simply by $\tilde{f}(x_1, x_2) = f(x_1)f(x_2)^{-1}$ for all x_1 and x_2 in X. Hence, by Lemma 8.5(d), $\operatorname{Im} \tilde{f}$ is the commutative subgroup $(\operatorname{Im} f)(\operatorname{Im} f)^{-1}$ of G.

Now it is easily seen that if (x_1, x_2) and (x_1', x_2') are in $X \times X$, and if there is an $x \in X$ such that $xx_1x_2' = xx_1'x_2$, then $\tilde{f}(x_1, x_2) = \tilde{f}(x_1', x_2')$. For if $xx_1x_2' = xx_1'x_2$, then $f(x)f(x_1)f(x_2') = f(x)f(x_1')f(x_2)$ which, after multiplying both sides by $f(x)^{-1}$, gives $f(x_1)f(x_2') = f(x_1')f(x_2)$. But this implies $f(x_1)f(x_2)^{-1} = f(x_1')f(x_2')^{-1}$ since $(\operatorname{Im} f)(\operatorname{Im} f)^{-1}$ is a commutative subgroup of G. Hence, we have our desired result.

This observation suggests considering the relation R on $X \times X$ given by $(x_1, x_2)\,\mathrm{R}\,(x_1', x_2')$ if and only if there is an x in X such that $xx_1x_2' = xx_1'x_2$. It is not difficult to check that R is an equivalence relation on the commutative monoid $X \times X$ because X is a commutative monoid. Obviously, the equivalence relation R on the monoid $X \times X$ depends only on the commutative monoid X and is independent of the morphism $f: X \to G$ we started with. On the other hand, we also know as a consequence of our previous discussion that given any morphism $f: X \to G$ with G a group, the partition $(X \times X)/\mathrm{R}$ of the monoid $X \times X$ is a refinement of the partition $\operatorname{Coim} \tilde{f}$ of the monoid $X \times X$. Hence, given any morphism $f: X \to G$ with G a group, we know that there is a unique morphism of monoids $g_f: (X \times X)/\mathrm{R} \to G$ such that $\tilde{f} = g_f k_{(X \times X)/\mathrm{R}}$. It is also not difficult to show that if we define the morphism of monoids $h: X \to (X \times X)/\mathrm{R}$ by $h(x) = [(x, 1)]$, where $[(x, 1)]$ is the equivalence

class of $(x, 1)$ under R, then the diagram

$$\begin{array}{ccc} X & \xrightarrow{h} & (X \times X)/\mathrm{R} \\ & \searrow^{f} & \downarrow g_f \\ & & G \end{array}$$

commutes. Hence, if we can show that the monoid $(X \times X)/\mathrm{R}$ is a group, then the morphism $h: X \to (X \times X)/\mathrm{R}$ stands a reasonably good chance of being a group of fractions for the commutative monoid X since $h: X \to (X \times X)/\mathrm{R}$ has the property that given any morphism of monoids $f: X \to G$ with G a group, there is the morphism $g_f: (X \times X)/\mathrm{R} \to G$ such that $f = g_f h$.

We first observe that $(X \times X)/\mathrm{R}$ is a commutative monoid with identity $[(1,1)]$ since $X \times X$ is a commutative monoid with identity $(1, 1)$. Hence, to show that $(X \times X)/\mathrm{R}$ is a group, and thus a commutative group, it suffices to show that every element $[(x_1, x_2)]$ in $(X \times X)/\mathrm{R}$ has an inverse. But $[(x_2, x_1)][(x_1, x_2)] = [(x_2x_1, x_1x_2)] = [(x_2x_1, x_2x_1)] = [(1, 1)]$. Since $(X \times X)/\mathrm{R}$ is commutative, $[(x_2, x_1)]$ is the inverse of $[(x_1, x_2)]$ for all elements $[(x_1, x_2)]$ in $(X \times X)/\mathrm{R}$ which shows that $(X \times X)/\mathrm{R}$ is a commutative group.

Summarizing our results, we know that the commutative group $(X \times X)/\mathrm{R}$ has the property that given any morphism of monoids $f: X \to G$, with G a group, there is the morphism $g_f: (X \times X)/\mathrm{R} \to G$ of groups such that $g_f h = f$ where $h: X \to (X \times X)/\mathrm{R}$ is the morphism given by $h(x) = [(x, 1)]$ for all x in X. Therefore, we will have shown that the morphism of monoids $h: X \to (X \times X)/\mathrm{R}$ is a group of fractions for X if we show that $h: X \to (X \times X)/\mathrm{R}$ has the property that two morphisms of groups $g_1, g_2: (X \times X)/\mathrm{R} \to G$ are the same if $g_1 h = g_2 h$.

To this end we observe that each element $[(x_1, x_2)]$ in $(X \times X)/\mathrm{R}$ can be written as $[(x_1, 1)][(1, x_2)] = [(x_1, 1)][(x_2, 1)]^{-1} = h(x_1)h(x_2)^{-1}$. Suppose $g_1, g_2: (X \times X)/\mathrm{R} \to G$ are morphisms of groups such that $g_1 h = g_2 h$. By our previous remark, $g_1 = g_2$ if and only if $g_1(h(x_1)h(x_2)^{-1}) = g_2(h(x_1)h(x_2)^{-1})$ for all x_1 and x_2 in X. But $g_1(h(x_1)h(x_2)^{-1}) = g_1(h(x_1))g_1(h(x_2)^{-1}) = g_1h(x_1)(g_1(h(x_2))^{-1} = g_2h(x_1)(g_2h(x_2))^{-1} = g_2(h(x_1))g_2(h(x_2)^{-1}) = g_2(h(x_1)h(x_2)^{-1})$. Thus, $g_1 = g_2$ if $g_1 h = g_2 h$, which finishes the proof that $h: X \to X \times X/\mathrm{R}$ is a group of fractions for X.

Before showing how this notion of the group of fractions of a commutative monoid can be applied to defining and constructing the additive group $\mathbf{Z}$ of all integers from the monoid of nonnegative integers $\mathbf{N}$ under addition, we introduce certain notational conventions and point out certain basic properties of groups of fractions.

If we are writing our commutative monoid X multiplicatively, as we have been doing until now, then we will write the commutative law of composition in the group of fractions $(X \times X)/\mathrm{R}$ also multiplicatively. Further, we shall denote the element $[(x_1, x_2)]$ in $(X \times X)/\mathrm{R}$ by the fraction x_1/x_2. In this notation the law of composition in $(X \times X)/\mathrm{R}$ takes on the familiar form $(x_1/x_2)(x_1'/x_2') = x_1x_1'/x_2x_2'$. Also, we have the familiar formula $(x_1/x_2)^{-1} = x_2/x_1$.

However, the question of when two fractions x_1/x_2 and x_1'/x_2' are equal has a slightly different answer than the usual one; namely, $x_1/x_2 = x_1'/x_2'$ if and only if there is an x in X such that $xx_1x_2' = xx_2x_1'$. Why we need this rather than the

familiar criterion $x_1x_2' = x_2x_1'$ for determining when two fractions x_1/x_2 and x_1'/x_2' are equal is made clear in the exercises. However, for the moment the reader should observe that in the cases he is used to, each element x in X has the property that $xy_1 = xy_2$ implies $y_1 = y_2$ for all y_1 and y_2. Thus, under these circumstances it is easily seen that our criterion for when two fractions x_1/x_2 and x_1'/x_2' are equal coincides with the usual one that $x_1x_2' = x_1'x_2$.

One reason for introducing this notation is that it has the advantage of familiarity which makes calculation easier. Another is that many of the previously introduced morphisms have a more appealing description in this notation. For example, the morphism $h: X \to (X \times X)/\mathrm{R}$ has the form $h(x) = x/1$ for all x in X. Also, given a morphism of monoids $f: X \to G$ with G a group, then the unique morphism $g_f: (X \times X)/\mathrm{R} \to G$ such that $f = g_f h$ has the appealing description $g_f(x_1/x_2) = f(x_1)(f(x_2))^{-1}$ for all x_1 and x_2 in X.

Because this construction of a group of fractions for a commutative monoid will occur often in this book we make the following definition.

Definition
Let X be a commutative monoid. We denote by $G(X)$ the commutative group $(X \times X)/\mathrm{R}$ where R is the equivalence relation on the monoid $X \times X$ given by $(x_1, x_2)\,\mathrm{R}\,(x_1', x_2')$ if and only if there is an x in X such that $xx_1x_2' = xx_2x_1'$. Then the morphism $h: X \to G(X)$ given by $h(x) = x/1$ for all x in X is a group of fractions for X which is called the **standard group of fractions** for X.

If we are writing the law of composition in X additively, then we will write the law of composition in $G(X)$ additively also. In this case we will denote the element $[(x_1, x_2)]$ in $(X \times X)/\mathrm{R}$ by $x_1 - x_2$. Then the law of composition in $G(X)$ becomes $(x_1 - x_2) + (x_1' - x_2') = x_1 + x_1' - (x_2 + x_2')$ while the inverse $-(x_1 - x_2)$ of $x_1 - x_2$ is $(x_2 - x_1)$. Also, we have that $x_1 - x_2 = x_1' - x_2'$ if and only if there is an x in X such that $x + x_1 + x_2' = x + x_1' + x_2$. The morphism $h: X \to G(X)$ is given in this notation by $h(x) = x - 0$ for all x in X. Finally, if $f: X \to G$ is a morphism of monoids with G a group, then the unique morphism of groups $g_f: G(X) \to G$ such that $f = g_f h$ can be described by $g_f(x_1 - x_2) = f(x_1)f(x_2)^{-1}$ for all x_1 and x_2 in X.

Basic Properties 8.6
Let $h: X \to H$ be a group of fractions for the commutative monoid X. Then:

(a) h is an epimorphism of monoids.
(b) h is an isomorphism if and only if X is a group.
(c) h is an injective morphism if and only if $xy = xz$ implies $y = z$ for all x, y, z in X.

When the morphism $h: X \to G(X)$ is an injective morphism we view X as a submonoid of $G(X)$ by identifying the element $x/1$ in $G(X)$ with the element x in X for each x in X. In this case the morphism $h: X \to G(X)$ then is the inclusion of the submonoid X of $G(X)$.

Using these ideas, we now construct the group of integers from the monoid of nonnegative integers.

9. THE INTEGERS

Definition
We define the group of **integers**, which we denote by **Z**, to be $G(\mathbf{N})$ where **N** is the additive monoid of nonnegative integers. The law of composition in **Z** is of course written additively.

Basic Properties 9.1
Let $h: \mathbf{N} \to \mathbf{Z}$ be the standard group of fractions for **N**.

(a) h is an injective morphism and so, according to our convention, **N** is a submonoid of **Z** (that is, we denote the element $n-0$ in **Z** by n for each n in **N**).
(b) If g is an element of a group G, then there is one and only one morphism $f: \mathbf{Z} \to G$ such that $f(1) = g$.
(c) The inclusion morphism $\mathbf{N} \to \mathbf{Z}$ of monoids is an epimorphism of monoids which is not a surjective morphism.
(d) We obtain an order relation on **Z** by defining $z_1 \leq z_2$ for z_1 and z_2 in **Z** if and only if there is an element n in **N** such that $z_1 + n = z_2$. This order relation has the following standard properties:
  (i) **Z** is totally ordered.
  (ii) If $z_1 \leq z_2$ and $z_1' \leq z_2'$, then $z_1 + z_1' \leq z_2 + z_2'$.
  (iii) z in **Z** is in **N** if and only if $0 \leq z$.
  (iv) For each element $z \in \mathbf{Z}$, let S_z be the set of all x in **Z** satisfying $x \geq z$. Then each set S_z is well ordered even though **Z** is not well ordered.
  (v) If $x > z$, then $x \geq z + 1$ for each z in **Z**.
(e) If z is in **Z** but not in **N**, then $-z$ is in **N**.

We recall that this last property is the basis of the notion of absolute value $|z|$ of an integer z. The **absolute value** is the map $|\ |: \mathbf{Z} \to \mathbf{N}$ given by $|z| = z$ if z is in **N** and $|z| = -z$, if z is not in **N**.

We now introduce the monoid structure on **Z** given by multiplication of integers.

We have already seen that there is a unique multiplicative monoid structure on **N** with the properties:

(a) $1n = n$ for all n in **N**
(b) $(n_1 + n_2)n = n_1 n + n_2 n$

for all n, n_1, and n_2 in **N**. We also know that this monoid structure makes N a commutative monoid. Using this multiplicative monoid structure on **N** we obtain a commutative multiplicative monoid structure on $\mathbf{N} \times \mathbf{N}$ given by $(n_1, n_2)(n_1', n_2') = (n_1 n_1' + n_2 n_2', n_1 n_2' + n_1' n_2)$ for all n_1, n_1', n_2, n_2' in **N**. Thus, the set $\mathbf{N} \times \mathbf{N}$ has commutative additive and multiplicative monoid structures. Now we obtained the group **Z** of integers from the additive monoid $\mathbf{N} \times \mathbf{N}$ by introducing the relation R on $\mathbf{N} \times \mathbf{N}$ given by $(n_1, n_2)\ \mathrm{R}\ (n_1', n_2')$ if and only if $n_1 + n_2' = n_1' + n_2$. (Why don't we need that there exists an n in N such that $n + n_1 + n_2' = n + n_1' + n_2$?) Not only is this equivalence relation on the set $\mathbf{N} \times \mathbf{N}$, an equivalence relation on the additive

monoid $\mathbf{N} \times \mathbf{N}$, it is also an equivalence relation on the multiplicative monoid $\mathbf{N} \times \mathbf{N}$. Thus, $\mathbf{Z} = (\mathbf{N} \times \mathbf{N})/R$ is a commutative multiplicative as well as additive monoid. We now list some of the familiar properties of this multiplication, all of which can be derived from this description of the multiplication of integers without much difficulty.

Basic Properties 9.2
The addition and multiplication defined on $\mathbf{Z}$ have the following properties:

(a) For all z, z_1, and z_2 in $\mathbf{Z}$ we have:
  (i) $z_1 z_2 = z_2 z_1$.
  (ii) $z(z_1 + z_2) = zz_1 + zz_2$.
  (iii) $z1 = z$.
  (iv) $0z = 0$.
  (v) $(-z_1)z_2 = -(z_1 z_2)$.
  (vi) If $z_1 z_2 = 0$, then either $z_1 = 0$ or $z_2 = 0$.
(b) The injective map $h: \mathbf{N} \to \mathbf{Z}$ is a morphism of monoids when both $\mathbf{N}$ and $\mathbf{Z}$ are considered as multiplicative monoids. Thus, $\mathbf{N}$ can be considered a submonoid of $\mathbf{Z}$ by means of the injective map $h: \mathbf{N} \to \mathbf{Z}$, both as additive and multiplicative monoids.
(c) If $z_1 \geq z_2$ and n is in $\mathbf{N}$, then $nz_1 \geq nz_2$.
(d) The absolute value $|\,|: \mathbf{Z} \to \mathbf{N}$ is a morphism of the multiplicative monoids of $\mathbf{Z}$ and $\mathbf{N}$ (that is, $|z_1 z_2| = |z_1||z_2|$ for all z_1, z_2 in $\mathbf{Z}$).
(e) 1 and -1 are the only invertible elements in the multiplicative monoid of $\mathbf{Z}$.

The reader should observe that the first few properties cited above are nothing more than the assertion that $\mathbf{Z}$ is a commutative ring under addition and multiplication. For the convenience of the reader we recall the following.

Definition
A set X together with two monoid structures $+$ and $\times$ is called a **ring** if:

(a) Under addition X is an abelian group whose identity we denote by 0.
(b) $x(x_1 + x_2) = xx_1 + xx_2$, $(x_1 + x_2)x = x_1 x + x_2 x$, for all x, x_1, and x_2 in X.

A ring X is said to be a **commutative ring** if it is a commutative monoid under multiplication.

The reader should show that for any ring X we have $0x = x0 = 0$ and $(-x_1)x_2 = -(x_1 x_2)$ for all x, x_1, and x_2 in X.

We will return to the general subject of rings later on. In the meantime we point out certain interesting interpretations of some of our results to date concerning the ring $\mathbf{Z}$.

We have already seen that if x is an element of a monoid X, the unique morphism $f: \mathbf{N} \to X$ from the additive monoid $\mathbf{N}$ to X such that $f(1) = x$, if written in the notation $f(n) = x^n$, gives us the notion of raising x to nonnegative integral powers. An analogous interpretation exists for the morphism from the additive group of $\mathbf{Z}$ to a group G. For we know that given any element g in the multiplicative group G, there is one and only one morphism $f_g: \mathbf{Z} \to G$ such that $f_g(1) = g$. If we denote $f_g(z)$ by g^z for all z in Z, then we have the familiar rules of exponentiation:

$g^0 = 1$, $g^1 = g$, $g^{z_1}g^{z_2} = g^{(z_1+z_2)}$ for all z_1 and z_2 in Z. Also, we have g^{-1} is the inverse of g and for each n in $\mathbf{N}$ we have $(g^{-1})^n = g^{-n}$. Hence, the unique morphism $f_g: \mathbf{Z} \to G$ such that $f_g(1) = g$ gives us the familiar notion of raising an element of a group to positive as well as negative powers. For this reason we will usually use the notation g^z to denote $f_g(z)$ for all z in $\mathbf{Z}$. That $(g^{z_1})^{z_2} = g^{z_1 z_2}$ for all z_1, z_2 in $\mathbf{Z}$ can be deduced from the fact that $(g^{n_1})^{n_2} = g^{(n_1 n_2)}$ for all n_1 and n_2 in $\mathbf{N}$.

Suppose G is a commutative multiplicative group. If g_1 and g_2 are two elements of G, then it is easy to check that the map $b: \mathbf{Z} \to G$ given by $b(z) = f_{g_1}(z)f_{g_2}(z)$ is a morphism of monoids. Since $b(1) = f_{g_1}(1)f_{g_2}(1) = g_1 g_2$, it follows that $f_{g_1 g_2}(z) = f_{g_1}(z)f_{g_2}(z)$ for all z in $\mathbf{Z}$. Thus, we obtain in this case the usual formula $(g_1 g_2)^z = g_1^z g_2^z$ for all z in $\mathbf{Z}$.

Suppose now that G is a commutative, additive group. For each g in G, let $f_g: \mathbf{Z} \to G$ be the unique morphism such that $f_g(1) = g$. Then we will usually denote $f_g(z)$ by zg for z in $\mathbf{Z}$. What we have already established shows that we have the usual rules: $0g = 0$, $1g = g$, $(z_1 + z_2)g = z_1 g + z_2 g$, $(z_1 z_2)g = z_1(z_2 g)$ for all g in G, and z_1, and z_2 in $\mathbf{Z}$. Further, if g_1 and g_2 are in G, then $z(g_1 + g_2) = zg_1 + zg_2$ for all z in $\mathbf{Z}$.

Summarizing, we see that associated with each abelian group G is the map $\mathbf{Z} \times G \to G$ given by $(z, g) \to zg$ which has the following properties:

(a) $(z_1 + z_2)g = z_1 g + z_2 g$.
(b) $z(g_1 + g_2) = zg_1 + zg_2$.
(c) $1g = g$.
(d) $(z_1 z_2)g = z_1(z_2 g)$.

It should be observed that this operation of $\mathbf{Z}$ on abelian groups G also has the property that if $f: G_1 \to G_2$ is a morphism of abelian groups, then $f(zg) = zf(g)$ for all z in Z and g in G_1.

We end this section by giving a generalization to arbitrary rings R of this operation of the ring $\mathbf{Z}$ on abelian groups.

Definition
Let R be a ring. By an ***R*-module structure** on an abelian group M we mean a map $R \times M \to M$ which we denote by $(r, m) \to rm$ satisfying:

(a) $(r_1 + r_2)m = r_1 m + r_2 m$.
(b) $r(m_1 + m_2) = rm_1 + rm_2$.
(c) $(r_1 r_2)(m) = r_1(r_2 m)$.
(d) $1m = m$.

An abelian group together with an R-module structure is called an ***R*-module**.

We shall return later on to this general notion of a module. In fact, most of this book will be devoted to a detailed study of rings and modules.

10. FINITE AND INFINITE SETS

In this section we point out some fundamental facts concerning finite and infinite sets and give some applications to monoids and groups.

Our development of the theory of finite sets is founded on certain basic facts concerning the cardinality of subsets of **N**, the set of nonnegative integers. In order to carry out this program it is convenient to have the familiar definitions.

Definitions
Let X be an ordered set.

(a) A subset X' of X is said to be an **interval** of X if whenever we have $x_1 \leq x \leq x_2$ with x_1 and x_2 in X', then x is also in X'.
(b) Associated with any two elements x_1 and x_2 in X are the intervals:
  (i) $[x_1, x_2]$ consisting of all x in X satisfying $x_1 \leq x \leq x_2$.
  (ii) $[x_1, x_2)$ consisting of all x in X satisfying $x_1 \leq x < x_2$.
  (iii) $(x_1, x_2]$ consisting of all x in X satisfying $x_1 < x \leq x_2$.
  (iv) (x_1, x_2) consisting of all x in X satisfying $x_1 < x < x_2$.

We now list most of the facts concerning the cardinality of subsets of **N** which are of immediate concern to us. An outline of the proofs of the following assertions is given in the exercises.

Proposition 10.1
Let **N** be the set of nonnegative integers.

(a) $\text{card}([0, m)) \leq \text{card}([0, n))$ if and only if $m \leq n$. Hence:
(b) $\text{card}([0, m)) = \text{card}([0, n))$ if and only if $m = n$.
(c) If $n_1 + x = n_2$ and $n_1' + x' = n_2'$, then $\text{card}([n_1, n_2) \cup [n_1', n_2')) \leq \text{card}([0, x + x'))$. Further, $\text{card}([n_1, n_2) \cup [n_1', n_2')) = \text{card}([0, x + x'))$ if and only if $[n_1, n_2) \cap [n_1', n_2') = \emptyset$.
(d) $\text{card}([0, m) \times [0, n)) = \text{card}([0, mn))$.
(e) If $N' \subset [n_1, n_2)$ for some n_1 and n_2 in **N**, then there is a unique n in **N** such that $\text{card}(N') = \text{card}([0, n))$.
(f) A subset N' of **N** has $\text{card}(N') = \text{card}(\mathbf{N})$ if and only if N' has no upper bound in **N**, that is, N' is not contained in $[0, n)$ for any n in N.

The reader should have no difficulty convincing himself that the following definition of finite and infinite sets agrees with his intuitive understanding of these notions.

Definitions
A set X is said to be a **finite set** if there is an injective map $f: X \to [0, n)$ for some n in **N**. The set X is said to be **infinite** if no such injective map exists.

As a consequence of these definitions we have the following.

Basic Properties 10.2
Let X and Y be sets, with X a finite set.

(a) If $f: Y \to X$ is injective, then Y is a finite set.
(b) If $g: X \to Y$ is a surjective map, then Y is finite.
(c) The set Y is infinite if and only if there is an injective map $\mathbf{N} \to Y$, or equivalently $\text{card}(Y) \geq \text{card}(\mathbf{N})$.

We now turn our attention to studying finite sets. However, in order to do this efficiently we must develop some new notions for products and sums in a monoid. Until now we have only had to multiply or add at most three or four elements in a monoid at a time. This required no special notation. However, to talk about multiplying or adding an unspecified finite number of things together, as we will often have to do from now on, does require some special notational devices. Therefore, we interrupt our discussion of finite sets to develop these devices.

Let X be a multiplicative monoid. Then given a finite sequence $x_1, \ldots, x_i, \ldots, x_k$ of elements of X with i in the interval $[1, k]$ in $\mathbf{N}$, we want to define the product of the sequence which we denote by $\prod_{i=1}^{k} x_i$. We do this by induction on k. If $k = 0$, or what is the same thing, the sequence is empty, we define $\prod_{i=1}^{k} x_i = 1$. Assuming we know what $\prod_{i=1}^{k} x_i$ is for $k \geq 0$, we define $\prod_{i=1}^{k+1} x_i$ to be $(\prod_{i=1}^{k} x_i)x_{k+1}$. Hence the product of any finite sequence $x_1, \ldots, x_n$ of elements in X is defined. Although it is not trivial to do so, it can be shown that the associativity of the product of a sequence of three elements [that is, $(x_1x_2)x_3 = x_1(x_2x_3)$ for all x_1, x_2, x_3 in X] implies the associativity of the product of any finite sequence of elements in X. Even to explain precisely what this means is a bit complicated. We shall merely give some examples to illustrate the point.

Example 10.3 Suppose $x_1, \ldots, x_6$ is a sequence of elements in the monoid X. Then $\prod_{i=1}^{6} x_i = x_1[(((x_2x_3)x_4)x_5)x_6] = [[(x_1x_2)x_3](x_4x_5)]x_6 = \prod_{i=1}^{4} y_i$ where $y_1 = (x_1x_2)$, $y_2 = (x_3)$, $y_3 = (x_4x_5)$, $y_4 = x_6$.

Example 10.4 Let $x_1, \ldots, x_7$ be a sequence of elements in the monoid X.

Suppose $y_1 = x_1 = x_2 = x_3$, $y_2 = x_4 = x_5$, $y_3 = x_6 = x_7$. Then $\prod_{i=1}^{7} x_i = y_1^3(y_2^2y_3^2) = \prod_{i=1}^{3} z_i$ where $z_1 = y_1^3$, $z_2 = y_2^2$, $z_3 = y_3^2$.

Suppose now that $x_1, \ldots, x_k$ is a sequence in a monoid X and $f : [1, \ldots, k] \to [1, \ldots, k]$ is an isomorphism of sets. Then $x_{f(1)}, \ldots, x_{f(k)}$ is another sequence of elements in X and in general $\prod_{i=1}^{k} x_i \neq \prod_{i=1}^{k} x_{f(i)}$. However, it is not difficult to show that $\prod_{i=1}^{k} x_i = \prod_{j=1}^{k} x_{f(j)}$ if any pair of elements in the sequence $x_1, \ldots, x_k$ commute. Thus, in particular, if X is commutative, then $\prod_{i=1}^{k} x_i$ depends only on the elements which appear in the sequence and not on the order in which they appear. Therefore, if X is a commutative monoid and $\{x_i\}_{i \in I}$ is any finite family of elements in X (that is, I is a finite set), it makes sense to speak of the product $\prod_{i \in I} x_i$ of the elements in the finite family $\{x_i\}_{i \in I}$.

Similarly, if X is a commutative monoid which is written additively and $\{x_i\}_{i \in I}$ is any finite family of elements in X, we can speak of their sum $\Sigma_{i \in I} x_i$. If $I = \emptyset$, then $\Sigma_{i \in I} x_i = 0$.

We now return to our discussion of finite sets.

Suppose X is a finite set. Then we know that for some integer m in $\mathbf{N}$ there is an injective map $f: X \to [0, m)$. Hence, $\text{card}(X) = \text{card}(\text{Im } f)$. Since $\text{Im } f \subset [0, m)$, we know from our discussion of the cardinality of subsets of $\mathbf{N}$ that there is a unique integer n in $\mathbf{N}$ such that $\text{card}(\text{Im } f) = \text{card}([0, n))$. Thus, if X is a finite set, there is a unique integer n in $\mathbf{N}$ such that $\text{card}(X) = \text{card}([0, n))$. This observation leads to the following.

Definition
Let X be a finite set. Then the unique integer n in $\mathbf{N}$ such that $\text{card}(X) = \text{card}([0, n))$ is called the **number of elements** in X. We will often denote the fact that n is the number of elements in X by writing $\text{card}(X) = n$.

As an immediate consequence of this definition we have the following.

Basic Properties 10.5
Let X and Y be finite sets. Then the following are equivalent:

(a) $\text{Card}(X) \geq \text{card}(Y)$.
(b) There is an injective morphism $f: Y \to X$. If $Y \neq \emptyset$, then (a) and (b) are equivalent to the following.
(c) There is a surjective map $f: X \to Y$.

The following results concerning finite sets can be deduced from our earlier proposition dealing with the cardinality of subsets of $\mathbf{N}$.

Proposition 10.6
Let X be a set, I a finite set, and $\{X_i\}_{i \in I}$ a family of finite subsets of X with $\text{card}(X_i) = n_i$ for each $i \in I$. Then:

(a) $\bigcup_{i \in I} X_i$ is finite set with $\text{card}(\bigcup_{i \in I} X_i) \leq \Sigma_{i \in I} n_i$. Further, $\text{card}(\bigcup_{i \in I} X_i) = \Sigma_{i \in I} n_i$ if and only if $X_i \cap X_j = \emptyset$ whenever $i \neq j$ in I.
(b) $\prod_{i \in I} X_i$ is a finite set with $\text{card}(\prod_{i \in I} X_i) = \prod_{i \in I} n_i$.
(c) If X_1 and X_2 are finite sets, with $\text{card}(X_i) = n_i$ for $i = 1, 2$, then the set (X_1, X_2) of all maps from X_1 to X_2 is a finite set with $\text{card}((X_1, X_2)) = n_2^{n_1}$.
(d) In particular, since the set 2^X of all subsets of a set X is isomorphic to (X, Y) where $Y = [0, 2)$, it follows that if X is a finite set with n elements, then 2^X is a finite set with $\text{card}(2^X) = 2^n$.

We end our discussion of finite sets with this useful characterization of such sets.

Proposition 10.7
Let X be a set. Then the following statements are equivalent:

(a) X is finite.
(b) If $f: X \to X$ is injective, then f is an isomorphism.
(c) If $g: X \to X$ is surjective, then g is an isomorphism.

Before discussing infinite sets, we give some illustrations of how some of these facts concerning finite sets come up in the theory of monoids and groups.

Proposition 10.8
Let X be a finite monoid (that is, X is finite as a set). If $xy_1 = xy_2$ implies $y_1 = y_2$ for all x, y_1, y_2 in X, then X is a group.

PROOF: For each x in X, let $l_x : X \to X$ be the map of sets given by $l_x(y) = xy$ for all y in X. Since the fact that $xy_1 = xy_2$ implies $y_1 = y_2$, we see that each map $l_x : X \to X$ is injective. Because X is a finite set, we know that each l_x must be an isomorphism of sets since it is injective. Thus, for each x in X, there is an x' in X such that $l_x(x') = 1 = xx'$.

Therefore, in order to show that X is a group, we have to show that $x'x$ also is 1. We know there is an x'' in X such that $x'x'' = 1$. Hence, we have $x'' = (xx')x'' = x(x'x'') = x$. Thus, $1 = x'x'' = x'x$, and we see that X is a group.

The reader should give an example of a monoid X which is not a group but which nonetheless satisfies the condition that $xy_1 = xy_2$ implies $y_1 = y_2$ for all x, y_1, y_2 in X.

Our next illustration of how the notion of finite sets comes up in dealing with monoids and groups is based on the notion of a product of monoids which we now define.

Definition
Let $\{X_i\}_{i \in I}$ be a nonempty family of monoids. The law of composition on the product of sets $\prod_{i \in I} X_i$, given by $(x_i)_{i \in I}(x'_i)_{i \in I} = (y_i)_{i \in I}$ where $y_i = x_i x'_i$ for all $i \in I$, makes $\prod_{i \in I} X_i$ a monoid which is called the **product of the family** $\{X_i\}_{i \in I}$ and is denoted by $\prod_{i \in I} X_i$. Each of the projection maps $\text{proj}_k : \prod_{i \in I} X_i \to X_k$ given by $\text{proj}_k((x_i)_{i \in I}) = x_k$ is a surjective morphism of monoids called the ***k*-projection morphism.**

We now list some easily verified properties of the product of monoids.

Basic Properties 10.9
Let $\{X_i\}_{i \in I}$ be a nonempty family of monoids.

(a) The identity of $\prod_{i \in I} X_i$ is $\{1_i\}_{i \in I}$ where 1_i is the identity of X_i.

(b) $(x_i)_{i \in I}$ in $\prod_{i \in I} X_i$ is invertible if and only if x_i is invertible in X_i for each $i \in I$. If $(x_i)_{i \in I}$ is invertible, then $(x_i)^{-1}_{i \in I} = (x_i^{-1})_{i \in I}$. Thus:

(c) $\prod_{i \in I} X_i$ is a group if and only if each monoid X_i is a group.

(d) $\prod_{i \in I} X_i$ is commutative if and only if each X_i is commutative.

(e) If Y is a monoid, then a map $f : Y \to \prod_{i \in I} X_i$ is a morphism of monoids if and only if each of the maps $\text{proj}_i f : Y \to X_i$ is a morphism of monoids. This implies:

(f) If Y is a monoid, then the map of sets

$$\text{Morph}\left(Y, \prod_{i \in I} X_i\right) \to \prod_{i \in I} \text{Morph}(Y, X_i) \qquad \text{given by} \qquad f \to (\text{proj}_i f)_{i \in I}$$

is an isomorphism of sets.

In addition to the product, there is another monoid associated with a family of monoids $\{X_i\}_{i\in I}$ which plays an extremely important role throughout this book and mathematics generally.

In order to describe this we need the following.

Definition

Let $\{X_i\}_{i\in I}$ be a nonempty family of monoids. The **support of an element** $\{x_i\}_{i\in I}$ in $\prod_{i\in I} X_i$ is the subset of J of I consisting of all i in I such that $x_i \neq 1_i$. An element $\{x_i\}$ in $\prod_{i\in I} X_i$ is said to have **finite support** if its support is a finite set.

It is not difficult to check that the subset of $\prod_{i\in I} X_i$ consisting of those elements with finite support is a submonoid of $\prod_{i\in I} X_i$ which we denote by $\Sigma_{i\in I} X_i$. For each k in I we shall denote the composition $\sum_{i\in I} X_i \xrightarrow{\text{inc}} \prod_{i\in I} X_i \xrightarrow{\text{proj}_k} X_k$ by $\text{proj}_k : \sum_{i\in I} X_i \to X_k$ which we call the ***k*th projection morphism**. Also for each k in I, the map $\text{inj}_k : X_k \to \sum_{i\in I} X_i$ given by $\text{inj}_k(x) = \{x_i\}_{i\in I}$ where $x_k = x$ and $x_i = 1_i$ for all $i \neq k$, is a morphism of monoids which we call the ***k*th injection morphism**. Since for each k in I we have that the composition $X_k \xrightarrow{\text{inj}_k} \sum_{i\in I} X_i \xrightarrow{\text{proj}_k} X_k$ is id_{X_k}, it follows that each inj_k is an injective morphism and each proj_k is a surjective morphism. We now list some easily verified properties.

Basic Properties 10.10

Let $\{X_i\}_{i\in I}$ be a nonempty family of monoids.

(a) $\{x_i\}_{i\in I} X_i$ is invertible if and only if each x_i is invertible in X_i. If $\{x_i\}_{i\in I}$ is invertible in $\Sigma_{i\in I} X_i$, then $\{x_i\}_{i\in I}^{-1} = \{x_i^{-1}\}_{i\in I}$.

(b) $\Sigma_{i\in I} X_i$ is a group if and only if each X_i is a group.

(c) $\Sigma_{i\in I} X_i$ is commutative if and only if each X_i is commutative.

(d) If I is finite, then $\Sigma_{i\in I} X_i = \prod_{i\in I} X_i$.

(e) If k and k' are distinct elements of I, then $\text{inj}_k(x)\text{inj}_{k'}(y) = \text{inj}_{k'}(y)\text{inj}_k(x)$ for all x in X_k and $y \in X_{k'}$.

We now wish to describe for each monoid Y the morphisms from $\Sigma_{i\in I} X_i$ to Y. Associated with a morphism of monoids $f : \Sigma_{i\in I} X_i \to Y$ are the morphisms $f \text{inj}_k : X_k \to Y$ for each k in I. Now it is not hard to check that two morphisms $f_1, f_2 : \Sigma_{i\in I} X_i \to Y$ are the same if and only if $f_1 \text{inj}_k = f_2 \text{inj}_k$ for all k in I. Thus, the map $\text{Morph}(\Sigma_{i\in I} X_i, Y) \to \prod_{i\in I} \text{Morph}(X_i, Y)$ of sets given by $f \to (f \text{inj}_k)_{k\in I}$ is injective. This naturally raises the question: Is the map $\text{Morph}(\Sigma_{i\in I} X_i, Y) \to \prod_{i\in I} \text{Morph}(X_i, Y)$ surjective and thus an isomorphism?

In general, it is not surjective. For if $f : \Sigma_{i\in I} X_i \to Y$ is a morphism of monoids and k and k' are distinct elements of I, the fact that $\text{inj}_k(x_k)\,\text{inj}_{k'}(x_{k'}) = \text{inj}_{k'}(x_{k'})\text{inj}_k(x_k)$ tells us that the morphism $f \text{inj}_k : X_k \to Y$ and $f \text{inj}_{k'} : X_{k'} \to Y$ have the property $f \text{inj}_k(x_k) f \text{inj}_{k'}(x_{k'}) = f \text{inj}_{k'}(x_{k'}) f \text{inj}_k(x_k)$ for all x_k in X_k and $x_{k'}$ in $X_{k'}$. Thus, if $f_i : X_i \to Y$ is a family of morphisms, then in order for there to be a morphism $f : \Sigma X_i \to Y$ such that $f_k = f \text{inj}_k$ for all k in I, the family of morphisms

$f_k : X_k \to Y$ must have the property $f_k(x_k)f_{k'}(x_{k'}) = f_{k'}(x_{k'})f_k(x_k)$ for all x_k in X_k and $x_{k'}$ in $X_{k'}$ whenever k and k' are distinct elements of I.

On the other hand, suppose we are given a family $f_k : X_k \to Y$ of morphisms such that $f_k(x_k)f_{k'}(x_{k'}) = f_{k'}(x_{k'})f_k(x_k)$ for all x_k in X_k and $x_{k'}$ in $X_{k'}$ whenever k and k' are distinct elements of I. Let $\{x_i\}_{i\in I}$ be an element of $\Sigma_{i\in I} X_i$. Because only a finite number of the $x_i \neq 1_i$, it follows that only a finite number of the elements in $f_k(x_k)$ in Y are different from 1. Also, since the elements $f_k(x_k)$ commute with each other we can talk about the product of the elements $f_k(x_k) \neq 1$ which we can denote by $\prod_{k\in I} f_k(x_k)$ without any confusion. Hence, we obtain a map $f : \Sigma_{i\in I} X_i \to Y$ by setting $f(\{x_i\}_{i\in I}) = \prod_{k\in I} f_k(x_k)$. It is not difficult to check that $f : \Sigma_{i\in I} X_i \to Y$ is a morphism which also has the property that $f_k = f \operatorname{inj}_k$ for all k in I. Thus, we have shown the following.

Proposition 10.11
Let $\{X_i\}_{i\in I}$ be a nonempty family of monoids. For each monoid Y, the map of sets $\operatorname{Morph}(\Sigma_{i\in I} X_i, Y) \to \prod_{i\in I} \operatorname{Morph}(X_i, Y)$ given by $f \to (f \operatorname{inj}_k)_{k\in I}$ for all morphisms $f : \Sigma_{i\in I} X_i \to Y$ is an injective map. Further, an element $\{f_k\}_{k\in I}$ in $\prod_{i\in I} \operatorname{Morph}(X_i, Y)$ is in the image of $\operatorname{Morph}(\Sigma_{i\in I} X_i, Y) \to \prod_{i\in I} \operatorname{Morph}(X_i, Y)$ if and only if $f_k(x_k)f_{k'}(x_{k'}) = f_{k'}(x_{k'})f_k(x_k)$ for all x_k in X_k and $x_{k'}$ in $X_{k'}$ whenever k and k' are distinct elements of I. Consequently, the map $\operatorname{Morph}(\Sigma_{i\in I} X_i, Y) \to \prod_{i\in I} \operatorname{Morph}(X_i, Y)$ is an isomorphism if Y is a commutative monoid.

Because the monoid $\sum_{i\in I} X_i$ plays a particularly important role when all the X_i are commutative, we give it a special name in that situation.

Definition
Let $\{X_i\}_{i\in I}$ be a nonempty family of commutative monoids. Then the commutative monoid $\sum_{i\in I} X_i$ is called the **sum of the family of monoids** $\{X_i\}_{i\in I}$ and is often denoted by $\coprod_{i\in I} X_i$.

Returning to our discussion of finite and infinite sets, we devote the rest of this section to developing a few useful facts concerning infinite sets.

As we have already seen, a set X is infinite if and only if $\operatorname{card}(X) \geq \operatorname{card}(\mathbf{N})$. Because $\mathbf{N}$ is itself infinite, $\mathbf{N}$ is the smallest infinite set and therefore in some respects the simplest infinite set. For this reason, we begin our discussion of infinite sets by pointing out certain facts about the cardinality of the set $\mathbf{N}$.

Proposition 10.12
Let X be a set. Then $\operatorname{card}(X) = \operatorname{card}(\mathbf{N})$ under the following circumstances:

(a) $X = X_1 \cup X_2$ where $\operatorname{card}(X_1) = \operatorname{card}(\mathbf{N})$ and $\operatorname{card}(X_2) \leq \operatorname{card}(\mathbf{N})$.
(b) There exists a partition $\{X_i\}_{i\in \mathbf{N}}$ of X where each set X_i is finite and not empty.
(c) $X \cong \mathbf{N} \times \mathbf{N}$.
(d) There is a partition $\{X_i\}_{i\in N}$ of X with $\operatorname{card}(X_i) = \operatorname{card}(\mathbf{N})$ for all i in N.

(e) $X \cong \prod_{j \in J} Y_j$ where J is a finite nonempty set and $\text{card}(Y_j) = \text{card}(\mathbf{N})$ for all j in J.

As a consequence of these criteria of when a set X is isomorphic to $\mathbf{N}$ we obtain the following facts concerning arbitrary infinite sets.

Proposition 10.13
Let X be an arbitrary infinite set. Then:

(a) There exists a partition $\{X_i\}_{i \in I}$ of X with $\text{card}(X_i) = \text{card}(N)$ for all i in I.
(b) $\text{Card}(\mathbf{N} \times X) = \text{card}(X)$.

EXERCISES

(1) Suppose H is a subgroup of a group G. A subset X of G is called a **left coset** of H in G if there is an element x in G such that $X = xH$. A subset Y of G is called a **right coset** of H in G if there is an x in G such that $Y = Hx$. The subset of 2^G consisting of the left cosets of H in G is denoted by G/H and is called the **left coset space** of H in G. The subset of 2^G consisting of the right cosets of H in G is denoted by $H\backslash G$ and is called the **right coset space** of H in G. Show that the following statements are true.

(a) Let x and y be two elements of G. Then the following statements are equivalent:
  (i) x and y belong to the same left coset of H in G.
  (ii) $xH = yH$.
  (iii) $y^{-1}x$ is in H.
  Similarly, the following statements are equivalent:
  (iv) x and y are in the same right coset of H in G.
  (v) $Hx = Hy$.
  (vi) yx^{-1} is in H.

(b) The subsets G/H and $H\backslash G$ of 2^G are both partitions of G having the following properties:
  (i) For each x in G, the left coset xH is the unique element of G/H containing x. Similarly, the right coset Hx is the unique element of $H\backslash G$ containing x.
  (ii) For each x in G, we have $\text{card}(xH) = \text{card}(H) = \text{card}(Hx)$.
  (iii) $\text{card}(G/H) = \text{card}(H\backslash G)$.
  (iv) $\text{card}(H \times (G/H)) = \text{card}(G) = \text{card}(H \times (H\backslash G))$.

(c) If G is a finite group, then $\text{card}(H) \times \text{card}(G/H) = \text{card}(G) = \text{card}(H) \times \text{card}(H\backslash G)$.

(d) $G/H = H\backslash G$ if and only if H is a normal subgroup of G.

(e) If $\text{card}(G/H) = 2$, then H is a normal subgroup of G.

(2) Let X be a set and $\text{Aut}(X)$ the group of bijective maps $f: X \to X$. For each x_0 in X define $\text{Aut}_{x_0}(X)$ to be the subset of $\text{Aut}(X)$ consisting of all bijective maps $f: X \to X$ with the property $f(x_0) = x_0$. Show that:

(a) $\text{Aut}_{x_0}(X)$ is a subgroup of $\text{Aut}(X)$ which is isomorphic to the group $\text{Aut}(X - \{x_0\})$.

(b) $\mathrm{Card}(\mathrm{Aut}(X)/\mathrm{Aut}_{\infty}(X)) = \mathrm{card}(X)$.

(c) If X is a finite set with n elements, then $\mathrm{card}(\mathrm{Aut}(X)) = n!$, where $n! = 1$ if $n = 0$, and $n! = 1 \times 2 \times \cdots \times n$ for $n > 0$.

(3) Let $\{G_i\}_{i \in I}$ be a family of submonoids (subgroups) of the monoid G. Show that:

(a) $G' = \bigcap_{i \in I} G_i$ is a submonoid (subgroup) of G.

(b) If $\{G_i\}_{i \in I}$ is totally ordered under inclusion, then $G'' = \bigcup_{i \in I} G_i$ is a submonoid (subgroup) of G.

(c) Give an example of a group G which contains subgroups G_1 and G_2 such that $G_1 \cup G_2$ is not a submonoid and hence not a subgroup of G.

(4) Let X be a subset of a monoid G. Then:

(a) The subset G' of G consisting of all finite products of elements in X is a submonoid of G called the **submonoid of G generated by X.**

(b) Show that G' is the intersection of all submonoids of G containing X.

(c) The subset X of G is said to **generate** G if $G' = G$. Show that if X generates G and $f_1, f_2 : G \to H$ are two morphisms of monoids, then $f_1 = f_2$ if and only if $f_1|X = f_2|X$. [*Hint*: Use the fact that if $f_1, f_2 : G \to H$ are two morphisms of monoids, then the subset of G consisting of all x in G such that $f_1(x) = f_2(x)$ is a submonoid of G.]

(5) Let X be a subset of a group G and let X^{-1} be the subset of G consisting of all x^{-1} where x is in X.

(a) Show that the submonoid G' of G generated by $X \cup X^{-1}$ is a subgroup of G. The subgroup G' is called the **subgroup of G generated by X.**

(b) Show that the subgroup G' of G generated by X is the intersection of all subgroups of G containing X.

(c) The set X is said to **generate** G if G is the subgroup of G generated by X. Show that if X generates G and $f_1, f_2 : G \to H$ are two morphisms of groups, then $f_1 = f_2$ if and only if $f_1|X = f_2|X$. [*Hint*: use the fact that if $f_1, f_2 : G \to H$ are two morphisms of groups, then the subset of G consisting of all x in G such that $f_1(x) = f_2(x)$ is a subgroup of G.]

(d) Suppose $f : G \to H$ is a surjective morphism of groups. If the subset X of G generates G, then $f(X)$ generates H.

(6) Suppose x is an element of a group G. Show that:

(a) The subgroup of G generated by x is the subset $\{x^z\}_{z \in \mathbf{Z}}$ of G. Hence:

(b) The image of the unique morphism $f : \mathbf{Z} \to G$ of groups with the property $f(1) = x$ is the subgroup of G generated by the element x in G.

A group generated by a single element is called a **cyclic group.**

(7) (a) A group G is a cyclic group if and only if there is a surjective morphism $f : \mathbf{Z} \to G$ of groups. Hence:

(b) All cyclic groups are abelian.

(c) If Z' is a subgroup of $\mathbf{Z}$, then there is a unique n in $\mathbf{N}$ which generates Z'. [*Hint*: Use the Euclidean algorithm which states that if a and b are in $\mathbf{N}$ and $b \neq 0$, then there are q and r in N such that $a = bq + r$ with $0 \leq r < b$.]

(d) Every subgroup of a cyclic group is cyclic.

(e) Let $\mathbf{Z}n$ be the subgroup of $\mathbf{Z}$ generated by the element n in $\mathbf{N}$.

(i) If $n \neq 0$, then $\mathrm{card}(\mathbf{Z}/\mathbf{Z}n) = n$.

(ii) $n = 0$ if and only if $\mathbf{Z}/\mathbf{Z}n$ is an infinite set.
(iii) If n_1 and n_2 are in $\mathbf{N}$, then the cyclic groups $\mathbf{Z}/\mathbf{Z}n_1$ and $\mathbf{Z}/\mathbf{Z}n_2$ are isomorphic groups if and only if $n_1 = n_2$.
(iv) If G is a cyclic group, then there is a unique element n in $\mathbf{N}$ such that G is isomorphic to $\mathbf{Z}/\mathbf{Z}n$. Hence:
(v) Two cyclic groups are isomorphic if and only if their cardinalities are the same.

(f) Suppose G is a finite cyclic group with $\text{card}(G) = g$. For each positive integer n dividing g, there is one and only one subgroup H of G with $\text{card}(H) = n$.

If a group G is finite, then the number of elements in G is called **the order of G**. If G is not finite, it is said to be of infinite order. The **order of an element x** in a group G is defined to be the order of the subgroup of G generated by x.

(8) Let x be an element in a group G.
(a) x is of infinite order if and only if $x^z = 1$ implies $z = 0$ where z is an integer in $\mathbf{Z}$.
(b) The following statements are equivalent:
(i) x is of finite order.
(ii) There is a nonzero z in $\mathbf{Z}$ such that $x^z = 1$.
(iii) There is a nonzero n in $\mathbf{N}$ such that $x^n = 1$.
(c) If x is of finite order, then the order of x is the smallest n in $\mathbf{N}$ such that $x^n = 1$.

(9) (a) Show that if z_1 and z_2 are in $\mathbf{Z}$, then:
(i) $z_1|z_2$ (z_1 divides z_2) if and only if the subgroup $\mathbf{Z}z_1$ of $\mathbf{Z}$ contains the subgroup $\mathbf{Z}z_2$ of $\mathbf{Z}$. Hence:
(ii) $\mathbf{Z}z_1 = \mathbf{Z}z_2$ if and only if $|z_1| = |z_2|$.

(b) Suppose n_1 and n_2 are positive integers, that is, n_1 and n_2 are in $\mathbf{N} - \{0\}$. Then there is a unique positive integer n such that the subgroup of $\mathbf{Z}$ generated by n_1 and n_2 is $\mathbf{Z}n$. Show that n is the largest integer in $\mathbf{N} - \{0\}$ which divides both n_1 and n_2. This positive integer n is called the **greatest common divisor** of n_1 and n_2 and is denoted by $\gcd[n_1, n_2]$. Two numbers are said to be **relatively prime** if $\gcd[n_1, n_2] = 1$.

(c) Show that the positive integer n is the greatest common divisor of the positive integers n_1 and n_2 if and only if n satisfies:
(i) $n|n_1$ and $n|n_2$.
(ii) $n = z_1n_1 + z_2n_2$ where z_1 and z_2 are in $\mathbf{Z}$. Hence:

(d) The positive integers n_1 and n_2 are relatively prime if and only if there are z_1 and z_2 in $\mathbf{Z}$ such that $z_1n_1 + z_2n_2 = 1$.

(e) If n_1 and n_2 are positive integers, then there is a unique positive n such that $\mathbf{Z}n$ is the subgroup $\mathbf{Z}n_1 \cap \mathbf{Z}n_2$ of $\mathbf{Z}$. This uniquely determined positive integer n is called the **least common multiple** of n_1 and n_2 and is denoted by $\text{lcm}[n_1, n_2]$. Show that $\text{lcm}[n_1, n_2]$ is the smallest positive integer divisible by both n_1 and n_2.

(f) Show that for a pair of positive integers n_1 and n_2 we have $\gcd[n_1, n_2] \times \text{lcm}[n_1, n_2] = n_1n_2$. Hence, $\text{lcm}[n_1, n_2] = n_1n_2$ if and only if n_1 and n_2 are relatively prime positive integers.

(10) Let n_1, n_2 be positive integers and let $\gcd[n_1, n_2] = n$.

(a) Show that if $n \neq 1$, then $\mathbf{Z}/\mathbf{Z}n_1 \times \mathbf{Z}/\mathbf{Z}n_2 = G$ is not a cyclic group by showing that the set of all g in G with $ng = 0$ is a subgroup of G which is not cyclic.

(b) Show that $\mathbf{Z}(\mathrm{lcm}[n_1, n_2])$ is the kernel of the group morphism $f: \mathbf{Z} \to \mathbf{Z}/\mathbf{Z}n_1 \times \mathbf{Z}/\mathbf{Z}n_2$ given by $f(z) = (k_1(z), k_2(z))$, where $k_i: \mathbf{Z} \to \mathbf{Z}/\mathbf{Z}n_i$ are the canonical morphisms of groups. Hence, if $\gcd[n_1, n_2] = n = 1$, then the induced morphism $\mathbf{Z}/\mathbf{Z}(\mathrm{lcm}[n_1, n_2]) \to \mathbf{Z}/\mathbf{Z}n_1 \times \mathbf{Z}/\mathbf{Z}n_2$ is an isomorphism of groups. Hence:

(c) The group $\mathbf{Z}/\mathbf{Z}n_1 \times \mathbf{Z}/\mathbf{Z}n_2$ is a cyclic group if and only if $\gcd[n_1, n_2] = 1$.

(d) Suppose G_1 and G_2 are finite cyclic groups of orders n_1 and n_2, respectively.

(i) $G_1 \times G_2$ is a finite cyclic group if and only if $\gcd[n_1, n_2] = 1$.

(ii) If $\gcd[n_1, n_2] = 1$, then $G_1 \times G_2$ is a cyclic group of order $n_1 n_2$.

(iii) If $G_1 \times G_2$ is a cyclic group, then an element (g_1, g_2) in $G_1 \times G_2$ generates $G_1 \times G_2$ if and only if g_1 generates G_1 and g_2 generates G_2.

(11) Let n be a positive integer, and let $k: \mathbf{Z} \to \mathbf{Z}/\mathbf{Z}n$ be the canonical surjective morphism of groups.

(a) Show that an element x in $\mathbf{N} - \{0\}$ has the property that $k(x)$ generates $\mathbf{Z}/\mathbf{Z}n$ if and only if $\gcd[x, n] = 1$.

(b) For each positive integer d the number of distinct generators of the cyclic group $\mathbf{Z}/\mathbf{Z}d$ is denoted by $\phi(d)$. Show:

(i) If p is a prime number (that is, $p \neq 1$ and p and 1 are the only positive integers that divide p), then $\phi(p) = p - 1$. More generally, $\phi(p^n) = p^n - p^{n-1}$ for all positive integers n.

(ii) Show that if m and n are relatively prime positive integers, then $\phi(mn) = \phi(m)\phi(n)$.

(iii) Hence, if a positive number $n = p_1^{n_1} \cdots p_t^{n_t}$ where the p_i are distinct positive prime numbers and the $n_i \geq 1$, then

$$\phi(n) = (p_1^{n_1} - p_1^{n_1 - 1}) \cdots (p_t^{n_t} - p_t^{n_t - 1})$$

(iv) For each positive integer n we have $n = \Sigma_{d|n} \phi(d)$, where the sum is taken over all positive integers d that divide n. [*Hint*: For each positive integer d dividing n, let X_d be the subset of $\mathbf{Z}/\mathbf{Z}n$ consisting of all elements of order d. Show that the collection $\{X_d\}_{d|n}$ is a partition of $\mathbf{Z}/\mathbf{Z}n$ and that $\mathrm{card}(X_d) = \phi(d)$ for all $d|n$.]

(12) Throughout this exercise $\mathbf{N}$ denotes the additive monoid of nonnegative integers. The purpose of this exercise is to describe the **cyclic monoids**, that is, the monoids generated by a single element.

Let J be the complement in $\mathbf{N} \times \mathbf{N}$ of the set $\{(n, 0) | n > 0\}$. For each element (x, t) in J denote by $\mathbf{N}_{(x,t)}$ the subset of $2^{\mathbf{N}}$ defined as follows.

(i) If $t = 0$, and hence $x = 0$, define $\mathbf{N}_{(x,t)}$ to be the collection $\{X_n\}_{n \in \mathbf{N}}$ where for each n in $\mathbf{N}$ the subset X_n of $\mathbf{N}$ consists of the single element n.

(ii) If $t \neq 0$, define $\mathbf{N}_{(x,t)}$ to be the collection $\{X_i\}_{0 \leq i < x+t}$ where

$$X_i = \begin{cases} \{i\}, & \text{if } i < x \\ \{i + tn\}_{n \in \mathbf{N}}, & \text{if } x \leq i < x + t \end{cases}$$

(a) Show that each $\mathbf{N}_{(x,t)}$ is a partition of the monoid $\mathbf{N}$.

(b) Show that $\mathbf{N}_{(x,t)} = \mathbf{N}_{(x',t')}$ if and only if $x = x'$ and $t = t'$.

(c) Show that if $\mathcal{P}$ is a partition of the monoid $\mathbf{N}$, then there is a unique element (x, t) in J such that $\mathcal{P} = \mathbf{N}_{(x,t)}$.

Suppose C is a cyclic monoid.

(d) Show that there is a unique element (x, t) in J such that $C \approx N_{(x,t)}$. This uniquely determined element (x, t) of J is called the invariant of C.
(e) Show that two cyclic monoids are isomorphic if and only if they have the same invariant.
(f) Show that C is infinite if and only if its invariant is $(0, 0)$.
(g) If C is a finite monoid with invariant (x, t), then $\text{card}(C) = x + t$.
(h) Show that two cyclic monoids with the same number of elements need not be isomorphic monoids.
(i) Let (x, t) be the invariant of C. Show that C is a group (and hence a cyclic group) if and only if $x = 0$ and $t \neq 0$.

(13) Throughout this exercise $\mathbf{N}$ denotes the additive monoid of nonnegative integers. The purpose of this exercise is to study the submonoids of $\mathbf{N}$.
(a) Describe the submonoid of $\mathbf{N}$ generated by 2 and 3 and show that it is not cyclic.
(b) More generally, show that for each integer $n > 1$, the submonoid generated by $n, n+1, \ldots, 2n-1$ cannot be generated by any fewer than n elements.
(c) Show that every submonoid of $\mathbf{N}$ is finitely generated. [*Hint*: Show that every nonzero submonoid of $\mathbf{N}$ is isomorphic to a submonoid of $\mathbf{N}$ having two relatively prime elements. Then show that if N' is a submonoid of $\mathbf{N}$ having two relatively prime elements, then there is an integer n in N' such that all integers $m \geq n$ are also in N'.]
(d) Show that every submonoid of a cyclic monoid is finitely generated.
(e) Is every submonoid of a finite cyclic monoid necessarily cyclic?

(14) Let M be a multiplicative monoid. The subset $C(M)$ consisting of all x in M such that $xm = mx$ for all m in M is called the **center** of M. Show:
(a) $C(M)$ is a submonoid of M which is a commutative monoid. Moreover, $C(M) = M$ if and only if M is a commutative monoid.
(b) An invertible element x in M is in $C(M)$ if and only if the inverse x^{-1} of x is in $C(M)$. Hence;
(c) If M is a group, then $C(M)$ is a subgroup of M which is a commutative group. Moreover, every subgroup of $C(M)$ is a normal subgroup of M.
(d) If M is a group and there is a subgroup M' of $C(M)$ such that M/M' is a cyclic group, then M is a commutative group. [*Hint*: Suppose x is in M such that $k(x)$ generates M/M' where $k: M \to M/M'$ is the canonical morphism. Show that each element m in M can be written as $x^z c$ for some z in $\mathbf{Z}$ and c in M'.]
(e) Suppose $f: M \to M'$ is a surjective morphism of monoids. Show that $f(C(M)) \subset C(M')$.
(f) If M' is a submonoid of M, then $C(M) \cap M'$ is a submonoid of $C(M')$.

(15) Let $M_2(\mathbf{R})$ be the monoid whose elements are the 2×2 matrices

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

over the real numbers $\mathbf{R}$ and whose law of composition is given by the usual matrix multiplication

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} u & v \\ x & y \end{bmatrix} = \begin{bmatrix} au+bx & av+by \\ cu+dx & cv+dy \end{bmatrix}$$

(a) Show that $C(M_2(\mathbf{R}))$ is the set of all matrices

$$\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$

for all a in $\mathbf{R}$.

(b) Show that the map $\det: M_2(\mathbf{R}) \to \mathbf{R}$ given by

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

(the ordinary determinant of a matrix) is a morphism from $M_2(\mathbf{R})$ to the multiplicative monoid of $\mathbf{R}$.

(c) Show that a matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ in $M_2(\mathbf{R})$ is an invertible element of $M_2(\mathbf{R})$ if and only if

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} \neq 0$$

Hence, $\mathrm{Inv}(M_2(\mathbf{R}))$ is the set of all matrices with nonzero determinants. $\mathrm{Inv}(M_2(\mathbf{R}))$ is usually denoted by $GL_2(\mathbf{R})$ and is called the **2×2 general linear group of $\mathbf{R}$**. If we denote by $\mathbf{R}^*$ the multiplicative group of nonzero real numbers, then $\det: GL_2(\mathbf{R}) \to \mathbf{R}^*$ is a morphism of groups whose kernel is usually denoted by $SL_2(\mathbf{R})$ and is called the **2×2 special linear group**.

(d) Show that $\det: GL_2(\mathbf{R}) \to \mathbf{R}^*$ is surjective by showing that the map $f: \mathbf{R}^* \to GL_2(\mathbf{R})$ given by

$$f(a) = \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}$$

is a morphism of groups such that $\det f = \mathrm{id}_{\mathbf{R}^*}$. Also show that $SL_2(\mathbf{R}) \neq \{1\}$.

(e) Let H be the subgroup $\left\{ \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} \right\}_{a \in \mathbf{R}^*}$ of $GL_2(\mathbf{R})$. Show that $H \cap SL_2(\mathbf{R}) = \{1\}$ and that every element x in $GL_2(\mathbf{R})$ can be written in one and only one way as $x = hs$ with h in H and s in $SL_2(\mathbf{R})$. Is H a normal subgroup of $GL_2(\mathbf{R})$?

(f) Show that $C(GL_2(\mathbf{R})) = GL_2(\mathbf{R}) \cap C(M_2(\mathbf{R}))$, that is,

$$C(GL_2(\mathbf{R})) = \left\{ \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \right\}_{a \in \mathbf{R}^*}$$

Thus, $C(GL_2(\mathbf{R})) \approx \mathbf{R}^*$.

(g) Show that $SL_2(\mathbf{R}) \cap C(GL_2(\mathbf{R})) = C(SL_2(\mathbf{R}))$ and is thus a cyclic group of order 2.

(h) Show that $SL_2(\mathbf{R})$ and $C(GL_2(\mathbf{R}))$ together do not generate $GL_2(\mathbf{R})$. [*Hint*: Consider the images of $SL_2(\mathbf{R})$ and $C(GL_2(\mathbf{R}))$ under the group morphism $\det: GL_2(\mathbf{R}) \to \mathbf{R}^*$.]

(i) Show that $GL_2(\mathbf{R})$ has elements of all possible orders.

(16) Let $\mathbf{Q}$ be the additive group of rational numbers and $\mathbf{Z}$, the subgroup of $\mathbf{Q}$, of all integers. Show that the abelian group $\mathbf{Q}/\mathbf{Z}$ has the following properties:

(a) Every element of $\mathbf{Q}/\mathbf{Z}$ is of finite order.

(b) For each nonzero z in $\mathbf{Z}$, the map $f: \mathbf{Q}/\mathbf{Z} \to \mathbf{Q}/\mathbf{Z}$ given by $f(x) = zx$ is a surjective morphism of groups with $\mathrm{Ker}\, f \approx \mathbf{Z}/|z|\mathbf{Z}$.

(c) For each positive integer n in $\mathbf{N}$, there is exactly one subgroup of $\mathbf{Q}/\mathbf{Z}$ of order n and this uniquely determined subgroup is cyclic.

(17) Suppose G is a group and S a set. An **operation of G on S** is a map $f: G \times S \to S$ which, if we denote $f(g, s)$ by gs for all g in G and s in S, satisfies:

(i) $1s = s$ for all s in S.

(ii) $g_1(g_2 s) = (g_1 g_2)s$.

Suppose we are given an operation of the group G on the set S. Then:

(a) For each g in G, the map $f_g: S \to S$ given by $f_g(s) = gs$ for all s in S is an isomorphism of sets. Hence, we obtain:

(b) The map $\alpha: G \to \mathrm{Aut}(S)$ given by $\alpha(g) = f_g$ for all g in G, is a morphism of groups with kernel consisting of all g in G such that $gs = s$ for all s in S.

(c) On the other hand, given a group morphism $\alpha: G \to \mathrm{Aut}(S)$, we can define the map $f: G \times S \to S$ by $f(g, s) = \alpha(g)(s)$. Show that this is an operation of G on S. Hence:

(d) There is a natural isomorphism between the set of group morphisms $(G, \mathrm{Aut}(S))$ and the set of operations of G on S.

(18) Suppose we are given an operation of the group G on the set S. Then with each element s in S, there is associated the map $f_s: G \to S$ given by $f_s(g) = g(s)$. The image of f_s is called the **orbit of s** under the operation of G on S. For each s in S, we have:

(a) The subset G_s consisting of all g in G such that $g(s) = s$ is a subgroup of G.

(b) The partition of the map $f_s: G \to S$ is the same as the left coset space G/G_s of G_s in G. Hence, the coimage analysis of f_s gives the isomorphism of sets $j_s: G/G_s \to \mathrm{Im}\, f_s$. Thus, $\mathrm{card}(G/G_s) = \mathrm{card}(\mathrm{Im}\, f_s)$.

(c) The subset S/G of 2^S consisting of the elements $\mathrm{Im}\, f_s$ of 2^S for all s in S is a partition of S called the **orbit space** of the operation of G on S.

(d) Suppose $\mathrm{card}(G)$ and $\mathrm{card}(S)$ are both finite. From each element i in S/G choose one element s_i in i. Then $\mathrm{card}(S) = \Sigma_{i \subset S/G}\, \mathrm{card}(G/G_{s_i})$.

(e) If two elements s and s' in S are in the same orbit, then there is a g in G such that $gG_s g^{-1} = G_{s'}$.

(19) Let G be a group. Show that the map $f: G \times G \to G$ given by $(g, x) = gxg^{-1}$ for all g and x in G is an operation of the group G on the underlying set of G. For each x in G, the orbit of x under this operation is called the **conjugacy class of x** and is usually denoted by C_x. The subgroup G_x of G consisting of all g in G such that $gxg^{-1} = x$ is called the **normalizer of x** and is usually denoted by N_x.

(a) Show that $C_x = \{x\}$ if and only if x is in the center of G.

(b) Suppose G is a finite group. Then there is a finite family $\{x_i\}_{i \in I}$ of elements in G which satisfy:

(i) No x_i is in $C(G)$.

(ii) $C_{x_i} = C_{x_j}$ implies $i = j$.

(iii) If C_x is a conjugacy class of G with more than one element, then $C_x = C_{x_i}$ for some i in I.

If $\{x_i\}_{i \in I}$ is a finite family of elements of G satisfying (i), (ii), and (iii), then

$$\mathrm{card}(G) = \mathrm{card}(C(G)) + \sum_{i \in I} \mathrm{card}\left(\frac{G}{N_{x_i}}\right)$$

where each $\mathrm{card}(N_{x_i}) < \mathrm{card}(G)$. This equation is known as the **class equation** of the group G.

(20) A finite group G is called a p-group where p is a positive prime number if $\text{card}(G) = p^n$ for some n in $\mathbf{N}$. Show that if G is a p-group, then:

(a) Every subgroup and every factor group of G is a p-group.

(b) If G is not trivial, then $C(G)$ is not trivial. [*Hint*: If $\text{card}(C(G)) = 1$, then the class equation gives

$$\text{card}(G) = p^n = 1 + \sum_{i \in I} \text{card}\left(\frac{G}{N_{x_i}}\right)$$

But this is impossible since $p|\text{card}(G/N_{x_i})$ for all i in I.]

(c) If p is a positive prime number, then every group of order p^2 is abelian.

(d) If p is a positive prime number, show that $\mathbf{Z}/p\mathbf{Z} \times \mathbf{Z}/p\mathbf{Z}$ and $\mathbf{Z}/p^2\mathbf{Z}$ are nonisomorphic groups of order p^2. Prove that any group G with $\text{card}(G) = p^2$ is isomorphic to either $\mathbf{Z}/p\mathbf{Z} \times \mathbf{Z}/p\mathbf{Z}$ or $\mathbf{Z}/p^2\mathbf{Z}$.

(e) Let X be a square in the plane. Show that the set of distance-preserving maps $f: X \to X$ is a subgroup of $\text{Aut}(X)$ of order 8 which is not commutative. Because $8 = 2^3$, this shows that not all groups of order p^3 are abelian.

(f) Show that all groups of order less than 6 are abelian.

(21) In this exercise we outline a proof of the well-known Sylow theorem: Suppose G is a finite group with $n = \text{card}(G)$. If $n = p_1^{\alpha_1} \cdots p_t^{\alpha_t}$ is a prime decomposition of n, that is, the p_i are distinct positive primes and the α_i are all positive integers, then there exists a subgroup of G of order $p_i^{\alpha_i}$ for each $i = 1, \ldots, t$.

(a) First prove the following. Suppose A is a finite abelian group of order n. If p is a positive prime integer which divides n, then there is an element of order p in A. [*Hint*: Proceed by induction on $n \geq 1$. If $n = 1$, there is nothing to prove. Suppose it is true if $1 \leq n < k$. Show that it suffices to consider only the case that there is a nonzero a in A such that (a), the subgroup of A generated by a, is a proper subgroup whose order is not divisible by p. In this case there must be an element of order p in $A/(a)$ and hence a subgroup A' of A containing (a) such that $A'/(a)$ is a cyclic group of order p. Show that if an element b in A' is not in (a), then the order of b is p.]

(b) With this preliminary result out of the way, one can proceed to prove the Sylow theorem as follows. Suppose $\text{card}(G) = n$. We want to show that if $n = p^\alpha q$ where p is a prime, $\alpha > 0$, and $\gcd[p^\alpha, q] = 1$, then there is a subgroup of G of order p^α. Proceed by induction on n. If $n = 1$, there is nothing to prove. Suppose now that it is true for $1 \leq n < k$, and show that it is true for $n = k$. Consider two separate cases: (1) $p|\text{card}((C(G))$ and (2) $p \not| \text{card}(C(G))$. In the first case use the preliminary result (a). In the second case use the class equation.

(22) Describe all groups of order at most 11.

(23) An endomorphism $s: X \to X$ of a set X is said to be a **peano successor function** if it satisfies:

(a) s is injective.

(b) $X - \text{Im}\, s$ consists of a single element which we denote by x_0.

(c) A subset Y of X is all of X if x_0 is in Y and $s(Y) \subset Y$.

The aim of this exercise is to show how, starting with a set X together with a peano successor function $s: X \to X$, one can construct a monoid satisfying the axioms given for the additive monoid of nonnegative integers.

Suppose $s: X \to X$ is a peano successor function on the set X. Then s is an

element of the monoid End(X). Show that the subset N of End(X) consisting of all f in End(X) such that $fs = sf$ is a submonoid of End(X). We now outline some of the steps needed to show that the monoid N satisfies the axioms for the additive monoid of nonnegative integers.

(a) Show that the map $\phi: N \to X$ given by $\phi(f) = f(x_0)$ is an isomorphism of sets.

(b) Show that a submonoid of N is all of N if it contains the element s.

(c) Show that N is a commutative monoid by showing that s is in $C(N)$.

(d) We shall say that an element f in N is regular if $fg = fh$ implies $g = h$ for all g and h in N. Show that all elements of N are regular by showing that the set of all regular elements in N is a submonoid of N containing s.

(24) Let G be an abelian group which we write additively. Suppose g_1 and g_2 are elements of finite orders m_1 and m_2, respectively.

(a) Show that the group morphism $\mathbf{Z}/m_1\mathbf{Z} \times \mathbf{Z}/m_2\mathbf{Z} \to G$ given by $(x + m_1\mathbf{Z}, y + m_2\mathbf{Z}) \to xg_1 + yg_2$ is an injective map if $\gcd[m_1, m_2] = 1$. Hence:

(b) If the $\gcd[m_1, m_2] = 1$, then the order of $g_1 + g_2$ is $m_1 m_2$.

(c) Show that there is an element in G of order $\operatorname{lcm}[m_1, m_2]$. [*Hint*: Let $a = \gcd[m_1, m_2]$. Then what are the orders of ag_1 and $ag_1 + ag_2$?]

(25) Suppose G is a finite additive abelian group with m elements. Hence, the order of each element of G is at most m which means that there is a largest integer n which is the order of some element of G. Show that $ng = 0$ for each element g in G.

(26) Suppose G is a finite abelian group which contains a cyclic subgroup H such that G/H is cyclic and card(H) and card(G/H) are relatively prime. Show that G is a cyclic group.

(27) The group of automorphisms of the set $[1, n]$ is called the **symmetric group** on n elements and is denoted by S_n. The elements of S_n are called **permutations** of $[1, n]$. For each pair of elements $i < j$ in $[1, n]$ define the permutation $(i, j): [1, n] \to [1, n]$ by

$$(i, j)(x) = \begin{cases} j, & \text{if } x = i \\ i, & \text{if } x = j \\ x, & \text{if } x \text{ is neither } i \text{ nor } j \end{cases}$$

The permutations (i, j) in S_n are called the **transpositions**. Show:

(a) $(i, j)^{-1} = (i, j)$.

(b) If $i < j < k$ in $[1, n]$, then $(i, j)(j, k)(i, j) = (i, k)$.

(c) $(j, k) = \sigma(k-1, k)\sigma^{-1}$ where $\sigma = (j, j+1)(j+1, j+2) \cdots (k-2, k-1)$. [*Hint*: Use (a) and induction on $k - j$.]

(d) Every element of S_n is a finite product of transpositions. [*Hint*: For each σ in S_n define $\#\sigma$ to be the number of x in $[1, n]$ such that $\sigma(x) = x$. Proceed by induction on $n - \#\sigma$.]

(e) The set of transpositions $\{(i, i+1)\}_{i \in [1, n)}$ generates S_n.

(f) For each σ in S_n define

$$\operatorname{sgn} \sigma = \frac{\prod_{1 \le i < j \le n} [\sigma(j) - \sigma(i)]}{\prod_{1 \le i < j \le n} (j - i)}$$

Show that $\operatorname{sgn}\sigma = \pm 1$. Hence, we have the map $\operatorname{sgn}: S_n \to \mathrm{Inv}(\mathbf{Z})$ where $\mathrm{Inv}(\mathbf{Z})$ is the group of invertible elements in the multiplicative monoid of $\mathbf{Z}$ consisting of $\{1, -1\}$. We now want to show that $\operatorname{sgn}: S_n \to \mathrm{Inv}(\mathbf{Z})$ is a morphism of groups.

(g) Show that if τ is a transposition in S_n, then $\operatorname{sgn}\tau = -1$.

(h) Suppose $\tau = (k, k+1)$. Show that if σ is in S_n, then $\operatorname{sgn}(\sigma\tau) = -\operatorname{sgn}\sigma = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$. [*Hint*: First show that if $i<j$ and $i \neq k$ or $j \neq k+1$, then $\tau(i) < \tau(j)$. Using this, show that

$$\prod_{1\le i<j\le n} [\sigma\tau(j)-\sigma\tau(i)] = [\sigma\tau(k+1)-\sigma\tau(k)]\Bigg(\prod_{\substack{1\le i<j\le n \\ i\neq k \text{ or } \\ j\neq k+1}} [\sigma\tau(j)-\sigma\tau(i)]\Bigg)$$

$$= [\sigma(k)-\sigma(k+1)]\Bigg(\prod_{\substack{1\le i<j\le n \\ i\neq k \text{ or } \\ j\neq k+1}} [\sigma(j)-\sigma(i)]\Bigg)$$

$$= -\prod_{1\le i<j\le n} [\sigma(j)-\sigma(i)]\Bigg]$$

(i) Combining (h) with the fact that S_n is generated by the transpositions of the form $(k, k+1)$, show that the map $\operatorname{sgn}: S_n \to \mathrm{Inv}(\mathbf{Z})$ is a morphism of groups.

(j) Suppose $\sigma = \tau_1 \cdots \tau_l$ where the τ_r are transpositions. Show that $\operatorname{sgn}\sigma = 1$ if and only if l is an even number. The permutations σ with $\operatorname{sgn}\sigma = 1$ are called the **even permutations**.

(k) The set of even permutations which is denoted by A_n is a normal subgroup of S_n of index 2 since it is precisely $\operatorname{Ker}\operatorname{sgn}$. The subgroup A_n is called the **alternating subgroup** of S_n.

(28) If G is a finite group and H is a proper subgroup of G, then $G \neq \bigcup_{x\in G} xHx^{-1}$. [*Hint*: Use the fact that if x and y are in the same left coset of H, then $xHx^{-1} = yHy^{-1}$.]

(29) Let G be a finite group of order n. Prove that G is a cyclic group if G has the property that for each $d|n$, the number of elements x in G such that $x^d = 1$ is less than or equal to d. [*Hint*: For each $d|n$, show that there are at most $\phi(d)$ elements of G of order d. Using the fact that $\Sigma_{d|n}\,\phi(d) = n$, show that there must be an element in G of order n and hence that G is cyclic.]

(30) The purpose of this exercise is to outline a proof of the fact that if $\mathbf{N}$ is an additive monoid satisfying the axioms for the additive monoid of nonnegative integers, then, given any element x in a monoid X, there is one and only one morphism $f: \mathbf{N} \to X$ such that $f(1) = x$.

Suppose $\mathbf{N}$ is an additive monoid satisfying the axioms for the nonnegative integers.

(a) Show that if X is a monoid, then a map $f: \mathbf{N} \to X$ is a morphism of monoids if and only if $f(0) = 1$ and $f(n+1) = f(n)f(1)$ for all n in $\mathbf{N}$.

(b) Show that if $f, g: X \to Y$ are morphisms of monoids, then the subset X' of X consisting of all x in X such that $f(x) = g(x)$ is a submonoid of X. Use this to

show that two morphisms of monoids $f, g: \mathbf{N} \to X$ are the same if and only if $f(1) = g(1)$.

(c) Let X be a monoid and x an element in X. Show that for each n in $\mathbf{N}$, there is one and only one map $g_n: [0, n] \to X$ satisfying:
  (i) $g_n(0) = 1$.
  (ii) $g_n(1) = x$.
  (iii) For each m in $[0, n)$, we have that $g_n(m+1) = g_n(m)g_n(1)$.

(d) Let x be an element in a monoid X. Define the map $f_x: \mathbf{N} \to X$ by $f_x(n) = g_n(n)$ where g_n is the unique map $g_n: [0, n] \to X$ satisfying the conditions in (c). Show that f_x is the unique morphism with the property $f(1) = x$.

(31) The purpose of this exercise is to outline proofs of some of the claims concerning the cardinalities of subsets of the set of nonnegative integers $\mathbf{N}$.

(a) Suppose $f: [0, m) \to [0, n)$ is an injective map. Show that if $m \geq 1$ and $n \geq 1$, then there is an injective $g: [0, m) \to [0, n)$ such that $g([0, m-1)) \subset [0, n-1)$. Use this to prove by induction on n that if $\text{card}([0, m)) \leq \text{card}([0, n))$, then $m \leq n$.

(b) Suppose that $n_1 + x = n_2$. Show that $n_1 + y$ is in $[n_1, n_2)$ for each g in $[0, x)$ and that the map $f: [0, x) \to [n_1, n_2)$ given by $f(y) = n_1 + y$ for all y in $[0, x)$ is a bijective map. Use this to show that if $n_1' + x' = n_2'$ and $[n_1, n_2) \cap [n_1', n_2') = \phi$, then $\text{card}([n_1, n_2) \cup [n_1', n_2')) = \text{card}([0, x + x'))$. From this deduce that if $n_1' + x' = n_2'$, then $\text{card}([n_1, n_2) \cup [n_1', n_2')) \leq \text{card}([0, x + x'))$ and that equality holds if and only if $[n_1, n_2) \cap [n_1', n_2') = \emptyset$.

(c) Observe that if $n \geq 1$, then $[0, m) \times [0, n) = ([0, m) \times [0, n-1)) \cup ([0, m) \times \{n\})$ and also that $([0, m) \times [0, n-1)) \cap ([0, m) \times \{n\}) = \emptyset$. Use this fact to show by induction on n, that $\text{card}([0, m) \times [0, n)) = \text{card}([0, mn))$.

(d) Suppose N' is a subset of $\mathbf{N}$ which has no upper bound. Show this means that for each x in N' we have that $S_x = \{y \in N' | y > x\}$ is not empty. Using the fact that $\mathbf{N}$ and hence N' is well ordered, define the map $s: N' \to N'$ by letting $s(x)$ be the first element of S_x for each x in N'. Show that s is an injective map with the property that if y in N' is not the first element y_0 of N', then $y = s(x)$ for some x in N'. Denoting the nth iterate of s by s^n for each n in $\mathbf{N}$, show that the map $f: \mathbf{N} \to N'$ defined by $f(n) = s^n(y_0)$ is a bijective map.

(32) Let X be a set. Show that the maps $\cap: 2^X \times 2^X \to 2^X$ and $\cup: 2^X \times 2^X \to 2^X$ are monoid structures on 2^X where $\cap: 2^X \times 2^X \to 2^X$ is defined by $\cap(X', X'') = X' \cap X''$ and $\cup: 2^X \times 2^X \to 2^X$ is defined by $\cup(X', X'') = X' \cup X''$ for all subsets X' and X'' of X.

(a) Show that $(2^X, \cap)$ and $(2^X, \cup)$ are commutative monoids.

(b) Show that the map $C: 2^X \to 2^X$ given by $C(X') = X - X'$ for all X' in 2^X is an isomorphism of monoids which is its own inverse.

(c) Show that each element x in $(2^X, \cap)$ and $(2^X, \cup)$ is **idempotent**, that is, $x^2 = x$.

(33) Let X be a monoid, G a group and $f: X \to G$ a morphism of monoids. Show that if x in X is idempotent, that is, $x^2 = x$, then $f(x) = 1$. Show that if every element in X is idempotent, then the group of fractions of X is the trivial group. Use this observation to give an example of a commutative nontrivial monoid, X, that is, a monoid with more than one element whose group of fractions $G(X)$ is the trivial group. This also gives an example in which the natural morphism $h: X \to G(X)$ is not injective.

Chapter 3 CATEGORIES

The mathematical object known as a category makes precise many of the similarities the reader has no doubt observed in our summary of set theory, monoid theory, and group theory. Because categories are a useful ambience in which to view mathematical systems generally, we devote this chapter to a brief discussion of categories.

1. CATEGORIES

We have already seen, in discussing monomorphisms, epimorphisms, and isomorphisms for sets, monoids, and groups, that the definitions as well as many of the basic properties of these notions depended only on relations between the maps and morphisms rather than on the actual structure of the sets, monoids, or groups involved. In fact, experience has shown that a great many of the properties of a wide variety of mathematical systems depend only on the way one chooses to compare the objects in the system, rather than on the explicit structure of the objects themselves. Of course, in order to take full advantage of this observation, it is necessary to have a setting which makes this point explicit. This is accomplished by the notion of a category. We start our discussion of categories by pointing out certain common features of set and monoid theory which when abstracted lead to the definition of a category.

In dealing with set theory, the things of primary concern to us have been the sets themselves, the maps between sets, and the composition of maps of sets.

Because the sets themselves are the objects of study in set theory, it is reasonable to call them the objects of set theory. Usually the collection of all objects of set theory, that is, the collection of all sets, is denoted by Ob(Sets).

In addition to the objects of set theory, we have for each pair of objects X and Y in Ob(Sets) the set (X, Y) of all maps from the set X to the set Y. Because maps can only be the same if they have the same domain and range we see that $(X, Y) \cap (X', Y') = \emptyset$ unless $X = X'$ and $Y = Y'$. Thus, set theory consists not only of the collection of objects Ob(Sets) but also of the collection of sets of maps (X, Y), one for each ordered pair of objects X and Y in Ob(Sets), which has the property $(X, Y) \cap (X', Y') = \emptyset$ unless $X = X'$ and $Y = Y'$.

Next we observe that the composition of maps of sets gives rise to the maps $(U, X) \times (X, Y) \to (U, X)$ given by $(f, g) \to gf$, where gf is the composition of the map f followed by the map g. Further, this composition is associative, that is, given f in (U, X), g in (X, Y), and h in (Y, Z), then the elements $h(gf)$ and $(hg)f$ in (U, Z) are the same. In addition to being associative, the composition maps $(X, Y) \times (Y, Z) \to (X, Z)$ also have the property that given any object X in Ob(Sets) there is an element f in (X, X) such that for each object Y in Ob(Sets) we have $gf = g$ for all g in (X, Y) and $fb = b$ for all b in (Y, X). Obviously, $f = \mathrm{id}_X$.

Summarizing this discussion, we see that associated with set theory there are the following data:

(a) The collection of all sets called the objects and denoted by Ob(Sets).
(b) A collection of sets (X, Y), one for each ordered pair of objects X and Y in Ob(Sets), satisfying the condition $(X, Y) \cap (X', Y') = \emptyset$ unless $X = X'$ and $Y = Y'$. Namely, for each ordered pair of objects X and Y in Ob(Sets), the set (X, Y) is the set of all maps from the set X to the set Y.
(c) For all triples of objects U, X, Y in Ob(Sets), we have maps $(U, X) \times (X, Y) \to (U, Y)$, given by the composition of maps, satisfying:
  (i) If U, X, Y, Z are objects in Ob(Sets) and f is in (U, X), g is in (X, Y), and h is in (Y, Z), then the elements $h(gf)$ and $(hg)f$ in (U, Z) are the same.
  (ii) For each object X in Ob(Sets) there is an element f in (X, X), namely, $f = \mathrm{id}_X$, which has the property that for each object Y in Ob(Sets) we have $gf = g$ for all g in (X, Y) while $fh = h$ for all h in (Y, X).

In view of the parallels between set theory and monoid theory we developed in the first two chapters, it should not come as a surprise that associated with monoid theory is a structure very similar to the one we just pointed out for set theory.

The collection of monoids, which we denote by Ob(Monoid), constitutes the objects of monoid theory. For each ordered pair X and Y of Ob(Monoid) we denote the set of all morphisms from the monoid X to the monoid Y by (X, Y). Clearly, this collection of sets (X, Y) has the property $(X, Y) \cap (X', Y') = \emptyset$ unless $X = X'$ and $Y = Y'$.

Next, the composition of morphisms of monoids gives maps $(X, Y) \times (Y, Z) \to (X, Z)$ for all triples of objects X, Y, and Z in Ob(Monoid), namely, $(f, g) \to gf$, where gf is the composition of the morphism $f : X \to Y$ followed by $g : Y \to Z$. As in the case of sets we know these maps $(X, Y) \times (Y, Z) \to (X, Z)$ satisfy:

(a) If U, X, Y, Z are objects in Ob(Monoid) and f is in (U, X), g is in (X, Y), and h is in (Y, Z), then the elements $h(gf)$ and $(hg)f$ in (U, Z) are the same.
(b) For each object X in Ob(Monoid) there is an f in (X, X), namely, $f = \mathrm{id}_X$, such that for each object Y in Ob(Monoid) we have $gf = g$ for each g in (X, Y) while $fh = h$ for all h in (Y, X).

On the basis of these two models the reader should be just about ready to make his own definition of a category.

Definition
A **category** $\mathscr{C}$ consists of the following:

(a) A collection $\mathrm{Ob}(\mathscr{C})$ whose elements are called the **objects** of $\mathscr{C}$.
(b) A collection of sets (X, Y), one for each ordered pair of objects X and Y of $\mathscr{C}$, satisfying $(X, Y) \cap (X', Y') = \emptyset$ unless $X = X'$ and $Y = Y'$. Each element of (X, Y) is called a **morphism** from X to Y and (X, Y) is called the set of **morphisms** from X to Y. We will often denote the fact that f is in (X, Y) by writing $f: X \to Y$.
(c) For each triple X, Y, Z of objects in $\mathscr{C}$, there is a map of sets $(X, Y) \times (Y, Z) \to (X, Z)$ denoted by $(f, g) \to gf$ where gf is called the **composition of the morphisms** $f: X \to Y$ and $g: Y \to Z$. These maps $(X, Y) \times (Y, Z) \to (X, Z)$ must satisfy:
   (i) If U, X, Y, Z are objects in $\mathscr{C}$ and f is in (U, X), g is in (X, Y), and h is in (Y, Z), then the elements $(hg)f$ and $h(gf)$ in (U, Z) are the same.
   (ii) For each object X in $\mathscr{C}$, there is an f in (X, X) such that for each object Y in $\mathscr{C}$, we have $gf = g$ for all g in (X, Y) while $fh = h$ for all h in (Y, X).

In view of our previous discussion it is obvious that set theory and monoid theory are examples of categories, which we denote by Sets and Monoid, respectively. For these examples we have already seen that for each object X there is only one morphism f in (X, X) with the property that for each object Y we have $gf = g$ for all g in (X, Y) while $fh = h$ for all h in (Y, X), namely, $f = \mathrm{id}_X$. It is not difficult to see that this holds generally in categories. For suppose X is an object in an arbitrary category $\mathscr{C}$ and f and f' are two morphisms in (X, X) such that for each object Y in $\mathscr{C}$ we have $gf = g$ and $gf' = g$ for all g in (X, Y) while $fh = h$ and $f'h = h$ for all h in (Y, X). Then letting $Y = X$, it follows that $f = ff' = f'$ which is our desired result. This leads to the following.

Definition
Let $\mathscr{C}$ be a category. For each object X in $\mathscr{C}$ the **identity morphism of X**, which we denote by id_X, is defined to be the unique morphism f in (X, X) such that for each object Y in $\mathscr{C}$ we have $gf = g$ for all g in (X, Y) and $fh = h$ for all h in (Y, X).

We have already seen that for each object X in Sets or Monoid the set (X, X) together with the law of composition $(X, X) \times (X, X) \to (X, X)$ given by $(f, g) \to gf$, where gf is the usual composition of morphisms, is a monoid with identity id_X which we denoted by $\mathrm{End}(X)$ and is called the monoid of endomorphisms of X. This, too, generalizes to arbitrary categories. For it is not difficult to show that if X is an object in a category $\mathscr{C}$, then the set (X, X) together with the law of

composition $(X, X)\times(X, X) \to (X, X)$ given by $(f, g) \to gf$, where gf is the composition in $\mathscr{C}$ of the morphism f and the morphism g, is also a monoid with id_X as identity element. This suggests the following.

Definition
Let $\mathscr{C}$ be an arbitrary category. Then for each object X in $\mathscr{C}$ the monoid consisting of the set (X, X) together with the law of composition $(X, X)\times(X, X) \to (X, X)$ given by $(f, g) \to gf$ is a monoid with id_X as identity which we denote by $\mathrm{End}(X)$ and call the **monoid of endomorphisms of** X.

We devote the rest of this section to pointing out how other concepts we have discussed earlier such as groups, finite sets, etc., give rise to categories. Other interesting examples of categories are discussed later on in this chapter as well as in the exercises at the end of the chapter. Before giving these examples it is useful to have the following notation.

Let $\mathscr{C}$ be a category. We shall often denote the set of morphisms from the object X to the object Y by $\mathscr{C}(X, Y)$ or $\mathrm{Hom}_{\mathscr{C}}(X, Y)$ instead of simply by (X, Y).

Since all of the examples of categories we now consider are also examples of subcategories of the categories of Sets or Monoid, we first introduce the general notion of a subcategory of a category and then give our concrete examples.

Definition
A category $\mathscr{C}'$ is said to be a **subcategory** of the category $\mathscr{C}$ if:

(a) Each object of $\mathscr{C}'$ is also an object of $\mathscr{C}$.
(b) For all objects X and Y in $\mathscr{C}'$, we have that $\mathscr{C}'(X, Y)$ is a subset of $\mathscr{C}(X, Y)$.
(c) For each object X in $\mathscr{C}'$ the subset $\mathscr{C}(X, X)$ of $\mathscr{C}(X, X)$ contains the element id_X of $\mathscr{C}(X, X)$.
(d) Given any objects X, Y, and Z in $\mathscr{C}'$ and morphisms f in $\mathscr{C}'(X, Y)$ and g in $\mathscr{C}'(Y, Z)$, their composition gf in $\mathscr{C}'(X, Z)$ is the same as their composition gf in $\mathscr{C}(X, Y)$ when f is viewed as an element of $\mathscr{C}(X, Y) \supset \mathscr{C}'(X, Y)$ and g is viewed as an element of $\mathscr{C}(Y, Z) \supset \mathscr{C}'(Y, Z)$.

Finally, a category $\mathscr{C}'$ is called a **full subcategory** of $\mathscr{C}$ if it is a subcategory of $\mathscr{C}$ which also satisfies:

(e) $\mathscr{C}'(X, Y) = \mathscr{C}(X, Y)$ for all objects X and Y in $\mathscr{C}'$.

The reader should have no difficulty verifying the following:

Basic Properties 1.1
Let $\mathscr{C}$ be a category.

(a) $\mathscr{C}$ is a full subcategory of $\mathscr{C}$.
(b) Two categories $\mathscr{C}$ and $\mathscr{C}'$ are the same if and only if $\mathscr{C}$ is a subcategory of $\mathscr{C}'$ and $\mathscr{C}'$ is a subcategory of $\mathscr{C}$.
(c) If $\mathscr{C}'$ is a subcategory of $\mathscr{C}$ and $\mathscr{C}''$ is a subcategory of $\mathscr{C}'$, then $\mathscr{C}''$ is a subcategory of $\mathscr{C}$.
(d) If $\mathscr{C}'$ is a full subcategory of $\mathscr{C}$ and $\mathscr{C}''$ is a full subcategory of $\mathscr{C}'$, then $\mathscr{C}''$ is a full subcategory of $\mathscr{C}$.

(e) If $\mathscr{C}'$ and $\mathscr{C}''$ are full subcategories of $\mathscr{C}$, then $\mathscr{C}' = \mathscr{C}''$ if and only if $\mathrm{Ob}(\mathscr{C}') = \mathrm{Ob}(\mathscr{C}'')$.

As a consequence of this last basic property we see that in order to specify a full subcategory $\mathscr{C}'$ of a category $\mathscr{C}$ it suffices to describe which objects of $\mathscr{C}$ are in $\mathscr{C}'$. We now use this fact to describe the categories of finite sets, groups, commutative monoid, commutative groups, etc.

Example 1.2 The category of finite sets is the full subcategory of Sets whose objects are the finite sets in Sets. Therefore, the category of finite sets has all finite sets as objects, the set (X, Y) of all morphisms from the finite set X to the finite set Y is just the set of all maps from X to Y, while the composition $(X, Y) \times (Y, Z) \to (X, Z)$ for all triples of finite sets X, Y, Z is given by $(f, g) \to gf$, where gf is the usual composition of the map $f: X \to Y$ and $g: Y \to Z$.

Example 1.3 The category Group is defined to be the full subcategory of Monoid whose objects are the monoids which are groups. Therefore, the objects of Group are all groups, (X, Y) is the set of all morphisms of groups from the group X to the group Y for all objects X and Y in Groups, and the composition $(X, Y) \times (Y, Z) \to (X, Z)$ for all triples of groups X, Y, Z is given by $(f, g) \to gf$ where gf is the usual composition of the morphisms of groups $f: X \to Y$ and $g: Y \to Z$.

Example 1.4 The category Abelian Monoid is the full subcategory of Monoid whose objects are the commutative monoids. Therefore, the objects of Abelian Monoid are all commutative monoids; for each ordered pair of objects X and Y in Abelian Monoid we have (X, Y) is the set of all morphisms from the commutative monoid X to the commutative monoid Y, while the composition $(X, Y) \times (Y, Z) \to (X, Z)$ for all triples X, Y, Z of commutative monoids is given by the usual composition of morphisms of monoids.

Example 1.5 The category Abelian Group is the full subcategory of Group whose objects are the commutative groups. The category Abelian Group is usually denoted by $\mathscr{Ab}$. The reader is urged to give a detailed description of the morphisms and composition of morphisms in $\mathscr{Ab}$ as has been done in the previous examples.

2. MORPHISMS

In this section we generalize to arbitrary categories some of the notions we have already discussed for the categories of sets, monoids, groups, etc. We begin with isomorphism, epimorphism, and monomorphism.

We recall that a map $f: X \to Y$ of sets is said to be an isomorphism if and only if there is a map $g: Y \to X$ such that $gf = \mathrm{id}_X$ and $fg = \mathrm{id}_Y$. Similarly, a morphism $f: X \to Y$ of monoids is said to be an isomorphism if and only if there is a morphism $g: Y \to X$ such that $gf = \mathrm{id}_X$ and $fg = \mathrm{id}_Y$. In both cases, the only notions used in the definition of an isomorphism are morphisms, composition of morphisms, and identity morphisms. Because all of these concepts exist in any category it is natural to make the following definition.

Definition
Let $\mathscr{C}$ be a category. A morphism $f: X \to Y$ in $\mathscr{C}$ is said to be an **isomorphism** if and only if there is a morphism $g: Y \to X$ in $\mathscr{C}$ such that $gf = \mathrm{id}_X$ and $fg = \mathrm{id}_Y$. The fact that a morphism $f: X \to Y$ is an isomorphism will often be denoted by writing $f: X \approx Y$.

Just about all the familiar properties of isomorphisms in the categories of sets and monoids also hold for arbitrary categories, as we point out in the following.

Basic Properties 2.1
Let $\mathscr{C}$ be a category.

(a) If $f: X \to Y$ is an isomorphism in $\mathscr{C}$, then there is one and only one morphism $g: Y \to X$ such that $gf = \mathrm{id}_X$ and $fg = \mathrm{id}_Y$. This uniquely determined morphism g is also an isomorphism which is called the **inverse** of f and is often denoted by f^{-1}.
(b) For each object X in $\mathscr{C}$, the morphism id_X is an isomorphism which is equal to its own inverse.
(c) If $f: X \to Y$ and $g: Y \to Z$ are isomorphisms in $\mathscr{C}$, then the composition $gf: X \to Z$ is an isomorphism with $(gf)^{-1} = f^{-1}g^{-1}$.
(d) If $f: X \to Y$ and $g: Y \to Z$ are morphisms in $\mathscr{C}$ such that gf is an isomorphism, then g is an isomorphism if and only if f is an isomorphism.

We omit the proofs of these basic properties since they are the same as the corresponding properties of isomorphisms in the categories Sets and Monoid. Again in analogy with Sets and Monoid we have the following.

Definition
If X and Y are objects in a category $\mathscr{C}$, then X is said to be isomorphic to Y if there is an isomorphism $f: X \to Y$. As we have already seen, because X is isomorphic to Y if and only if Y is isomorphic to X, we will often use the symmetric expression X and Y are isomorphic instead of the asymmetric expressions X is isomorphic to Y or Y is isomorphic to X.

As an immediate consequence of this definition and the basic properties of isomorphisms, we have the following.

Basic Properties 2.2
Let X, Y, and Z be objects in a category $\mathscr{C}$. Then:

(a) X is isomorphic to X.
(b) If X is isomorphic to Y, then Y is isomorphic to X.
(c) If X is isomorphic to Y and Y is isomorphic to Z, then X is isomorphic to Z.

Having seen how the notion of isomorphism can be generalized from the categories of Sets and Monoid to arbitrary categories, the reader should have no difficulty seeing how the notions of epimorphism and monomorphism can be likewise generalized to arbitrary categories. Consequently, we just give the definitions and basic properties of these notions for categories in general. It is left to the

reader to make the appropriate connections with our discussion of these notions for the categories of Sets and Monoid.

Definitions
Let $f: X \to Y$ be a morphism in a category $\mathscr{C}$.

(a) $f: X \to Y$ is said to be an **epimorphism** if given two morphisms $g_1, g_2: Y \to Z$ in $\mathscr{C}$, we have $g_1 = g_2$ whenever $g_1 f = g_2 f$.
(b) $f: X \to Y$ is said to be a **monomorphism** if given any two morphisms $h_1, h_2: U \to X$ in $\mathscr{C}$ we have $h_1 = h_2$ whenever $fh_1 = fh_2$.

Basic Properties 2.3
Let $f: X \to Y$ and $g: Y \to Z$ be two morphisms in $\mathscr{C}$.

(a) If f and g are both epimorphisms (monomorphisms), then the composition $gf: X \to Z$ is an epimorphism (monomorphism).
(b) If $gf: X \to Z$ is an epimorphism, then so is g.
(c) If $gf: X \to Z$ is a monomorphism, then so is f.
(d) If $f: X \to Y$ is an isomorphism, then f is both a monomorphism and an epimorphism.

In connection with the last of these basic properties, it is worth observing that a morphism in a category which is both a monomorphism and epimorphism need not be an isomorphism. For we have already seen that the inclusion from the monoid **N** of nonnegative integers under addition to the group **Z** of integers under addition is a monomorphism as well as an epimorphism in the category Monoid but is not an isomorphism in Monoid since it is not a surjective morphism.

Because the category Group is a full subcategory of Monoid, it is reasonable to ask whether morphisms of Groups can be both monomorphisms and epimorphisms without being isomorphisms. We have already seen that a morphism of groups is an isomorphism if and only if it is bijective. Hence, it is natural to wonder how monomorphisms and injective morphisms of groups as well as epimorphisms and surjective morphisms of groups are connected. Using techniques similar to those used in Chapter 2, Proposition 3.6, it can be shown that a morphism of groups is a monomorphism if and only if it is injective. In addition, we leave it to the reader to show that surjective morphisms of groups are also epimorphisms. This leaves the question of whether epimorphisms of groups are necessarily surjective morphisms. In fact, it can be shown that this is indeed the case. However, we will not prove this.

The reader has undoubtedly observed that we have made no attempt to generalize the notions of surjective and injective morphisms from the categories Sets and Monoid to arbitrary categories. Once having these notions for Sets it was not difficult to transfer them to Monoid using the fact that every object in Monoid has an underlying set. However, there is nothing in the definition of a category which guarantees that each object has in any sense an underlying set. For this reason, it is more difficult to define in an arbitrary category what is meant by surjective and injective morphisms. In fact, no generally accepted way of doing this exists at the present time.

3. PRODUCTS AND SUMS

We begin by recalling some of the properties of the product $\prod_{i\in I} X_i$ of a family of sets $\{X_i\}_{i\in I}$. Associated with the product $\prod_{i\in I} X_i$ are the projection maps $\text{proj}_k: \prod_{i\in I} X_i \to X_k$, one for each k in I, which we showed to have the following property: If Y is any set and $f_i: Y \to X_i$, one for each i in I, is any family of maps from Y to the sets X_i, then there is a unique map $f: Y \to \prod_{i\in I} X_i$ such that $f_k: Y \to X_k$ is the composition $Y \xrightarrow{f} \prod_{i\in I} X_i \xrightarrow{\text{proj}_k} X_k$ for each k in I. Or stated more succinctly, if for each set Y we define the map of sets $\beta_Y: (Y, \prod_{i\in I} X_i) \to \prod_{i\in I} (Y, X_i)$ by $\beta_Y(f) = \{\text{proj}_k f\}_{k\in I}$ for all f in $(Y, \prod_{i\in I} X_i)$, then β_Y is an isomorphism of sets for each set Y. Also we showed that the product of a family of monoids has similar properties. We now show that these properties of the product of a family of sets or monoids can be used as the basis for the definition of a product of a family of objects in an arbitrary category.

Suppose $\{X_i\}_{i\in I}$ is a family of objects in a category $\mathscr{C}$. Let X be an object in $\mathscr{C}$ and $\text{proj}_k: X \to X_k$ a family of morphisms in $\mathscr{C}$, one for each k in I. These data associate with each object Y in $\mathscr{C}$ the map of sets $\beta_Y: (Y, X) \to \prod_{i\in I} (Y, X_i)$ given by $\beta_Y(f) = \{\text{proj}_k f\}_{k\in I}$ for all f in (Y, X). In view of our preliminary remarks concerning the product of a family of sets or monoids it is tempting to make the following definition.

Definition
Let $\{X_i\}_{i\in I}$ be a family of objects in a category $\mathscr{C}$. We say that an object in $\mathscr{C}$ together with a family $\{\text{proj}_k: X \to X_k\}_{k\in I}$ of morphisms in $\mathscr{C}$ is a **product** in $\mathscr{C}$ of the family $\{X_i\}_{i\in I}$ if for each object Y in $\mathscr{C}$ the map of sets

$$\beta_Y: (Y, X) \to \prod_{i\in I} (Y, X_i)$$

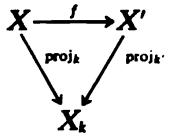
given by $\beta_Y(f) = \{\text{proj}_k f\}_{k\in I}$ for all f in (Y, X) is an isomorphism of sets.

If X together with the family of morphisms $\{\text{proj}_k: X \to X_k\}_{k\in I}$ is a product for the family $\{X_i\}_{i\in I}$, then each morphism $\text{proj}_k: X \to X_k$ is called the **kth projection morphism**.

Although our previous discussion shows that if $\{X_i\}_{i\in I}$ is a family of objects in Sets, then the set $\prod_{i\in I} X_i$ together with the usual projection maps $\text{proj}_k: \prod_{i\in I} X_i \to X_k$ is a product of $\{X_i\}_{i\in I}$ in Sets, we have not answered the following obvious question raised by this general definition of products in categories. Namely, if X is a set and $\text{proj}'_k: X \to X_k$ is a family of maps of sets such that X together with $\{\text{proj}'_k: X \to X_k\}_{k\in I}$ is also a product for $\{X_i\}_{i\in I}$, then how are the sets $\prod_{i\in I} X_i$ and X as well as the families of maps $\{\text{proj}_k: \prod_{i\in I} X_i \to X_k\}_{k\in I}$ and $\{\text{proj}'_k: X \to X_k\}_{k\in I}$ related? This question is completely answered by the following.

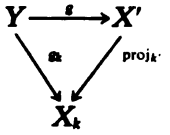
Basic Property 3.1

Let $\{X_i\}_{i\in I}$ be a family of objects in the category $\mathscr{C}$. If the object X together with the family of morphisms $\{\text{proj}_k: X \to X_k\}_{k\in I}$ as well as the object X' together with the family of morphisms $\{\text{proj}'_k: X' \to X_k\}_{k\in I}$ are both products for the family $\{X_i\}_{i\in I}$, then there is one and only one morphism $f: X \to X'$ such that the diagram



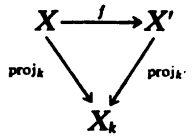
commutes for each k in I. Moreover, this uniquely determined morphism $f: X \to X'$ is an isomorphism.

PROOF: Because X' together with the family $\{\text{proj}'_k: X' \to X_k\}_{k\in I}$ of morphisms is a product for the family $\{X_i\}_{i\in I}$, we know that given any object Y in $\mathscr{C}$ and any family $\{g_i: Y \to X_i\}_{i\in I}$ of morphisms, there is a unique morphism $g: Y \to X'$ such that $\beta_Y(g) = \{g_i\}_{i\in I}$ where β_Y is the map of sets $\beta_Y: (Y, X') \to \prod_{i\in I}(Y, X_i)$ given by $\beta_Y(h) = \{\text{proj}'_k h\}_{k\in I}$ for all h in (Y, X'). Or, stated in other words, there is a unique morphism $g: Y \to X'$ which makes the diagrams



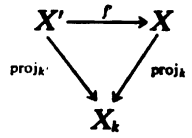
commute for all k in I.

Thus, letting $Y = X$ and $g_k: Y \to X_k$ be the morphisms $\text{proj}_k: X \to X_k$ for each k in I, there is a unique morphism $f: X \to X'$ such that $\text{proj}_k = \text{proj}'_k f$ for all k in I, or, equivalently, such that each of the diagrams



commutes. This establishes the first part of the basic property. We now show that this uniquely determined morphism $f: X \to X'$ is an isomorphism.

Because X together with the family $\{\text{proj}_k: X \to X_k\}_{k\in I}$ of morphisms is also by assumption a product for $\{X_i\}_{i\in I}$, we know by what we have just shown that there exists a unique morphism $f': X' \to X$ such that



commutes for all k in I. Since $f: X \to X'$ has the property that $\text{proj}'_k f = \text{proj}_k$ for all k in I and $f': X' \to X$ has the property $\text{proj}'_k = \text{proj}_k f'$ for all k in I, it follows that

the compositions $X \xrightarrow{f} X' \xrightarrow{f'} X$ and $X' \xrightarrow{f'} X \xrightarrow{f} X'$ have the property that $\text{proj}_k(f'f) = \text{proj}_k$ for all k in I and $\text{proj}'_k ff' = \text{proj}'_k$ for all k in I. But the fact that X together with $\{\text{proj}_k : X \to X_k\}_{k \in I}$ is a product for $\{X_i\}_{i \in I}$ implies that there is precisely one morphism $h : X \to X$ such that $\text{proj}_k h = \text{proj}_k$ for all k in I. Since id_X and $f'f$ both have this property, it follows that $f'f = \text{id}_X$. A similar argument shows that $ff' = \text{id}_{X'}$. Therefore, we have shown that the morphism $f : X \to X'$ is an isomorphism.

In essence, this basic property explains in what sense products of families of objects in a category are unique, assuming they exist. The problem of existence is not trivial. Although we know that the categories Sets, Monoids, Groups, Abelian Monoids, and Abelian Groups all have products for arbitrary families of objects (the usual products together with the usual projection morphisms), there are nonetheless categories that do not have this property. For example, if $\mathscr{C}$ is the category of finite sets and $\{X_i\}_{i \in I}$ is an infinite family of sets with each X_i having at least two elements, then the family $\{X_i\}_{i \in I}$ has no product in $\mathscr{C}$, even though it has one in the larger category Sets.

Next we show how the notion of the sum $\coprod_{i \in I} X_i$ of a family of sets or abelian monoids or groups can be generalized to arbitrary categories.

Recall that if $\{X_i\}_{i \in I}$ is a family of commutative monoids, then associated with the sum $\coprod_{i \in I} X_i$ of this family of monoids are the injection morphisms $\text{inj}_k : X_k \to \coprod_{i \in I} X_i$, one for each k in I, which have the property that if Y is any commutative monoid and $\{f_i : X_i \to Y\}_{i \in I}$ is any family of morphisms, then there is one and only one morphism $f : \coprod_{i \in I} X_i \to Y$ with the property that $f \,\text{inj}_k = f_k$ for each k in I. Or stated more succinctly, for each commutative monoid Y, the map of sets $\alpha_Y : (\coprod_{i \in I} X_i, Y) \to \prod_{i \in I} (X_i, Y)$, given by $\alpha_Y(f) = (f \,\text{inj}_k)_{k \in I}$, is an isomorphism. We now show how this property of sums of commutative monoids can be used as the basis for defining sums of a family of objects in an arbitrary category.

Suppose $\{X_i\}_{i \in I}$ is a family of objects in a category $\mathscr{C}$. Let X be an object in $\mathscr{C}$ and $\text{inj}_k : X_k \to X$ a family of morphisms, one for each k in I. These data associate with each object Y in $\mathscr{C}$ a map of sets $\alpha_Y : (X, Y) \to \prod_{i \in I} (X_i, Y)$ given by $\alpha_Y(f) = (f \,\text{inj}_k)_{k \in I}$ for each f in (X, Y). In view of our preliminary remarks concerning the sum of a family of abelian monoids, it is tempting to make the following definition.

Definition

Let $\{X_i\}_{i \in I}$ be a family of objects in a category $\mathscr{C}$. We say that an object X together with a family $\{\text{inj}_k : X_i \to X\}_{k \in I}$ of morphisms in $\mathscr{C}$ is a **sum** in $\mathscr{C}$ of the family $\{X_i\}_{i \in I}$ if for each object Y in $\mathscr{C}$, the map

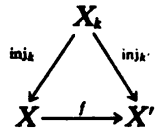
$$\alpha_Y : (X, Y) \to \prod_{i \in I} (X_i, Y)$$

given by $\alpha_Y(f) = (f \,\text{inj}_k)_{k \in I}$ is an isomorphism of sets.

If X together with the family of morphisms $\{\text{inj}_k : X_k \to X\}_{k \in I}$ is a sum of $\{X_i\}_{i \in I}$, then each morphism $\text{inj}_k : X_k \to X$ is called the ***k*th injection morphism**.

As was the case with products, one might ask how two sums for a particular family of objects in $\mathscr{C}$ are related. This is answered by the following basic property which we present without proof since the manner of proof is very similar to that used in establishing the uniqueness of products of families of objects in a category.

Basic Property 3.2
Let $\{X_i\}_{i\in I}$ be a family of objects in the category $\mathscr{C}$. If the object X together with family of morphisms $\{\text{inj}_k: X_k \to X\}_{k\in I}$ and the object X' together with the family of morphisms $\{\text{inj}'_k: X_k \to X'\}_{k\in I}$ are both sums of $\{X_i\}_{i\in I}$, then there is one and only one morphism $f: X \to X'$ such that the diagrams



commute for each k in I. Moreover, this uniquely determined morphism $f: X \to X'$ is an isomorphism.

Having explained in what sense sums for families of objects in a category are unique, it remains to discuss when sums of objects in a category actually exist. As in the case of products, this depends on the category. From our discussion concerning the sum of commutative monoids, it is obvious that the usual sum $\coprod_{i\in I} X_i$ of a family $\{X_i\}_{i\in I}$ of commutative monoids together with the usual injection morphisms $\text{inj}_k : X_k \to \coprod_{i\in I} X_i$ is a sum of $\{X_i\}_{i\in I}$ in the category Abelian Monoid. Hence, each family of objects in Abelian Monoid has a sum. Similarly, the usual sum $\coprod_{i\in I} X_i$ of a family of abelian groups $\{X_i\}_{i\in I}$ together with the usual injection morphisms $\text{inj}_k : X_k \to \coprod_{i\in I} X_i$ is a sum of $\{X_i\}_{i\in I}$ in the category Abelian Group. The reader should also check that what we defined as the sum of an indexed family of sets is the sum of that indexed family in the category Sets.

On the other hand, the full subcategory $\mathscr{C}$ of Abelian Groups consisting of all finite abelian groups does not have the property that every family of objects in $\mathscr{C}$ has a sum in $\mathscr{C}$, even though it does have one in Abelian Group. For example, if $\{X_i\}_{i\in I}$ is an infinite family of nontrivial finite abelian groups, then $\{X_i\}_{i\in I}$ does not have a sum in $\mathscr{C}$.

EXERCISES

(1) A map $f: X_1 \to X_2$ between two ordered sets X_1 and X_2 is said to be **order preserving** if $f(x_1) \geq f(x_2)$ whenever $x_1 \geq x_2$. Show:
(a) The identity map on an ordered set is order preserving.
(b) The ordinary composition of two order-preserving maps is order preserving.
(c) Show that the following data define a category $\mathscr{C}$ called the category of ordered sets.

(i) The objects of $\mathscr{C}$ are the ordered sets.
(ii) For each pair of objects X_1 and X_2 in $\mathscr{C}$, $\mathscr{C}(X_1, X_2)$ is the set of order-preserving maps from X_1 to X_2.
(iii) For all triples X_1, X_2, and X_3 in $\mathrm{Ob}(\mathscr{C})$ the composition map $\mathscr{C}(X_1, X_2) \times \mathscr{C}(X_2, X_3) \to \mathscr{C}(X_1, X_3)$ is given by $(f, g) \to gf$, the ordinary composition of maps.

(d) Show that a morphism $f: X_1 \to X_2$ of ordered sets is an isomorphism of ordered sets, that is, f is an isomorphism in the category of ordered sets if and only if (i) f as a map of sets is bijective and (ii) the map of sets $f^{-1}: X_2 \to X_1$ is order preserving.

(e) Give an example to show that a morphism $f: X_1 \to X_2$ of ordered sets which is a bijective map of sets need not be an isomorphism of ordered sets.

(f) Show that every nonempty indexed family of ordered sets has a sum and product in the category of ordered sets. Suppose each set X_i in the nonempty indexed family $\{X_i\}_{i \in I}$ is totally ordered. Is the sum or the product of the family necessarily totally ordered?

(g) Show that a morphism $f: X_1 \to X_2$ of ordered sets is a monomorphism (epimorphism) in the category of ordered sets if and only if f is injective (surjective) as a map of sets. [*Hint*: In order to show that if f is an epimorphism, then it is a surjective map, it is useful to have the following construction. Suppose Z is a nonempty ordered set and z_0 an element of Z. Consider the ordered set Z' which as a set is $Z \cup \{t\}$, where t is an element not in Z with the ordering given by:
(i) If z_1 and z_2 are in Z, then $z_1 \geq z_2$ in Z' if and only if $z_1 \geq z_2$ in Z.
(ii) If x is in Z', then $x \leq t$ if either $x = t$ or x is in Z and $x < z_0$.
(iii) If x is in Z', then $x \geq t$ if either $x = t$ or x is in Z and $x > z_0$.
After showing that Z' is an ordered set show that the following maps $f, g: Z \to Z'$ are distinct morphisms of ordered sets:

$$f(z) = z, \quad \text{for all } z \text{ in } Z$$

$$g(z) = \begin{cases} z, & \text{if } z \neq z_0 \\ t, & \text{if } z = z_0 \end{cases}$$

(2) Let G be a group. A set S together with an operation of G on S is called a G-set. If S_1 and S_2 are G-sets, then a G-morphism from S_1 to S_2 is a map of sets $f: S_1 \to S_2$ satisfying $f(gs) = gf(s)$ for all g in G and s in S_1. Show:

(a) For each G-set S, id_S is a G-morphism.

(b) If S_1, S_2, S_3 are G-sets and $f: S_1 \to S_2$ and $g: S_2 \to S_3$ are G-morphisms, then the ordinary composition of maps $gf: S_1 \to S_3$ is a G-morphism.

(c) Show that the following data define a category which is called the category of G-sets, and is denoted by G-Sets.
(i) The objects of G-Sets are the G-sets.
(ii) For each pair of objects S_1 and S_2 of G-Sets, $G\text{-Sets}(S_1, S_2)$ is the set of all G-morphisms from S_1 to S_2.
(iii) For each triple S_1, S_2, and S_3 of objects of G-Sets, the composition map $G\text{-Sets}(S_1, S_2) \times G\text{-Sets}(S_2, S_3) \to G\text{-Sets}(S_1, S_3)$ is given by $(f, g) \mapsto gf$, the ordinary composition of maps.

(d) Show that a morphism $f: S_1 \to S_2$ of G-sets is an isomorphism of G-sets, that is, f is an isomorphism in the category G-Sets if and only if $f: S_1 \to S_2$ is a bijective map of sets.

(e) Show that every indexed family $\{X_i\}_{i \in I}$ of objects of G-Sets has a sum and product in the category G-Sets.

(f) A G-set S' is said to be a **G-subset** of a G-set S if S' is a subset of S and the inclusion map $\text{inc}: S' \to S$ is a G-morphism.

(i) Suppose S' is a subset of the G-set S such that $g(s')$ is in S' for all s' in S'. Show that S' is a G-set by means of the operation $G \times S' \to S'$ given by $(g, s') = g(s')$ for all g in G and s' in S'. Also, show that the G-set S' is a G-subset of S. Thus, the G-subsets of a G-set S are nothing more nor less than the subsets S' of S satisfying $g(s')$ is in S' for each s' in S'.

(ii) Suppose S is a G-set. Show that each orbit of S is a G-subset of S and that S is isomorphic to the sum of the family of G-sets consisting of the orbits of S.

(g) Suppose G' is a subgroup of G.

(i) Show that if $X = yG'$ is a left coset of G' in G, then for each g in G, the set $g(X) = \{gx\}_{x \in X}$ is the left coset $(gy)G'$ of G' in G.

(ii) Show that the map $G \times G/G' \to G/G'$ given by $(g, X) \to g(X)$ for all g in G and X in G/G' is an operation of G on the set G/G'. This is the only way we consider G/G' a G-set.

(iii) Show that a G-set S is isomorphic to a G-set G/G' for some subgroup G' if and only if S is a nonempty set which has no G-subsets other than $\emptyset$ and S. Such a G-set S is called a simple G-set.

(iv) Show that a G-set S is simple if and only if $S \neq \emptyset$ and any G-morphism $f: S' \to S$ is surjective if $S' \neq \emptyset$.

(v) Show that if G_1 and G_2 are two subgroups of G, then the G-sets G/G_1 and G/G_2 are isomorphic if and only if there is a g in G such that $G_1 = gG_2g^{-1}$.

(h) Suppose S is a G-set. Show that there is a family $\{X_i\}_{i \in I}$ of simple G-sets such that S is a sum of the family $\{X_i\}_{i \in I}$. Also, show that if $\{Y_j\}_{j \in J}$ is another family of simple G-sets such that S is a sum of the family $\{Y_j\}_{j \in J}$, then there is an isomorphism of sets $\phi: I \to J$ such that the G-sets X_i and $Y_{\phi(i)}$ are isomorphic for all i in I.

(i) Let S be a G-set. Show that if $f: G \to S$ is a morphism of G-sets, then for each g in G the map $(gf): G \to S$ defined by $(gf)(x) = f(xg)$ is also a G-morphism. Further, if we denote the set of G-morphisms from G to S by (G, S), then the map $G \times (G, S) \to (G, S)$ given by $(g, f) \to gf$ is an operation of G on (G, S). This is the only way we consider (G, S) as a G-set.

(j) Let S be a G-set. Show that the map $\phi: (G, S) \to S$ given by $\phi(f) = f(1)$ for all f in (G, S) is an isomorphism in the category of G-sets.

(k) Show that if s is an element of the G-set S and $f: G \to S$ is the unique G-morphism such that $f(1) = s$, then $\text{Im } f$ is the orbit of s.

(l) Show that a G-set is simple if and only if there is a G-morphism $f: G \to S$ which is surjective as a map of sets.

(m) Show that a G-morphism $f: S_1 \to S_2$ of G-sets is a monomorphism in the category of G-sets if and only if f, as a map of sets, is injective.

(n) Show that a G-morphism $f: S_1 \to S_2$ of G-sets is an epimorphism in the cate-

gory of G-sets, if and only if, as a map of sets, it is surjective. [*Hint*: Show (a) Im f is a G-subset of S_2 and (b) if X is any G-subset of a G-set S, then the subset $S - X$ consisting of all elements in S but not in X is also a G-subset of S.]

(3) Suppose X is an ordered set. Show that the following data define a category which we denote by $\mathscr{C}(X)$ and call the **category of the ordered set** X.

(a) The objects of $\mathscr{C}(X)$ are the elements of X.

(b) For each pair of objects x_1 and x_2 in $\mathscr{C}(X)$, define $\mathscr{C}(X)(x_1, x_2)$ to be the ordered pair (x_1, x_2) if $x_1 \geq x_2$ in X and to be empty otherwise.

(c) Show that for each triple of objects x_1, x_2, x_3 there is one and only one map $\mathscr{C}(X)(x_1, x_2) \times \mathscr{C}(X)(x_2, x_3) \to \mathscr{C}(X)(x_1, x_3)$ and define that unique map to be the composition of morphisms in $\mathscr{C}(X)$.

(4) Suppose $\mathscr{C}$ is a category having the following properties: (a) the collection of objects of $\mathscr{C}$ is a set; (b) if C_1 and C_2 are objects of $\mathscr{C}$, then the set $\mathscr{C}(C_1, C_2)$ is empty or consists of a single element; and (c) if $\mathscr{C}(C_1, C_2)$ and $\mathscr{C}(C_2, C_1)$ are both not empty, then $C_1 = C_2$.

Let X be the set of objects of $\mathscr{C}$. Show that the relation $\geq$ in X given by $C_1 \geq C_2$ if and only if $\mathscr{C}(C_1, C_2)$ is not empty, is an order relation on X. The ordered set consisting of X together with this order relation, is called the **ordered set of** $\mathscr{C}$ and is denoted by $X(\mathscr{C})$.

(5) Show that if X is an ordered set, then $X(\mathscr{C}(X)) = X$. The reader should also convince himself that if $\mathscr{C}$ is a category satisfying the hypothesis of Exercise 4, then, although $\mathscr{C}(X(\mathscr{C}))$ is not identical to $\mathscr{C}$, it is essentially the same thing as $\mathscr{C}$.

(6) Let M be a monoid. Show that the following data define a category which we denote by $\mathscr{C}(M)$ and call the **category of the monoid** M.

(a) Ob $\mathscr{C}(M)$ is the set consisting of the single element M.

(b) The set of morphisms $\mathscr{C}(M)(M, M)$ is the set M.

(c) The composition map $\mathscr{C}(M, M) \times \mathscr{C}(M, M) \to \mathscr{C}(M, M)$ is given by $(m_1, m_2) \to m_2 \cdot m_1 = m_1 m_2$ where $m_1 m_2$ is the product in the monoid M of the elements m_2, m_1 in M.

(7) Suppose $\mathscr{C}$ is a category with one object C. Show that the set $\mathscr{C}(C, C)$ together with the law of composition given by the composition map $\mathscr{C}(C, C) \times \mathscr{C}(C, C) \to \mathscr{C}(C, C)$ is a monoid, which we denote by $M(\mathscr{C})$ and call the **monoid of the category** $\mathscr{C}$.

(8) Show:

(a) If M is a monoid, then $M(\mathscr{C}(M)) = M$.

(b) If $\mathscr{C}$ is a category with one object, then although $\mathscr{C}(M(\mathscr{C}))$ need not be $\mathscr{C}$, it is essentially the same thing as $\mathscr{C}$.

(9) Suppose $\mathscr{C}$ is a category. Show that the following data give a category $\mathscr{C}^{op}$ which is called the **opposite category** of $\mathscr{C}$.

(a) The objects of $\mathscr{C}^{op}$ are the same as the objects of $\mathscr{C}$.

(b) $\mathscr{C}^{op}(C_1, C_2) = \mathscr{C}(C_2, C_1)$ for all objects C_1 and C_2 in $\mathscr{C}$.

(c) For all triples of objects C_1, C_2, and C_3 in $\mathscr{C}^{op}$, the composition maps $\mathscr{C}^{op}(C_1, C_2) \times \mathscr{C}^{op}(C_2, C_3) \to \mathscr{C}^{op}(C_1, C_3)$ are given by $(f, g) \to g \circ f$ where $g \circ f$ in $\mathscr{C}^{op}(C_1, C_3) = \mathscr{C}(C_3, C_1)$ is the composition fg in $\mathscr{C}$ of the morphisms $C_3 \xrightarrow{g} C_2 \xrightarrow{f} C_1$ in $\mathscr{C}$.

(10) Suppose X is an ordered set and $\mathscr{C}(X)$ is the category of the ordered set X.

Show that there is a unique ordered set Y such that $\mathscr{C}(Y) = \mathscr{C}(X)^{op}$. This uniquely determined ordered set is called the **opposite of** X and is often denoted by X^{op}. What does it mean about an ordered set X that $X = X^{op}$?

(11) Suppose M is a monoid. Show that there is a unique monoid N such that $\mathscr{C}(N) = \mathscr{C}(M)^{op}$. This uniquely determined monoid is called the **opposite** of M and is written M^{op}.

(a) Show that M is a group if and only if M^{op} is a group.

(b) Show that $M = M^{op}$ if and only if M is a commutative monoid.

So far in our discussion of categories, we have not dealt at all with the problem of comparing two categories. We now describe how two categories are compared.

Suppose $\mathscr{C}_1$ and $\mathscr{C}_2$ are two categories. A **functor** F from $\mathscr{C}_1$ to $\mathscr{C}_2$ consists of the following data:

(a) A map from Ob $\mathscr{C}_1$ to Ob $\mathscr{C}_2$ which we denote by $F: \text{Ob } \mathscr{C}_1 \to \text{Ob } \mathscr{C}_2$.

(b) Maps $F: \mathscr{C}_1(C_1, C_2) \to \mathscr{C}_2(F(C_1), F(C_2))$ for each pair of objects C_1 and C_2 in $\mathscr{C}_1$ satisfying

(i) $F(\text{id}_{C_1}) = \text{id}_{F(C_1)}$ and

(ii) given $f: C_1 \to C_2$ and $g: C_2 \to C_3$ morphisms in $\mathscr{C}_1$, then $F(gf) = F(g)F(f)$.

We will usually denote the fact that F is a functor from $\mathscr{C}_1$ to $\mathscr{C}_2$ by writing $F: \mathscr{C}_1 \to \mathscr{C}_2$. It is the functors from one category to another that are the gadgets used to compare categories. We now give some examples of functors.

(12) For a category $\mathscr{C}$, show that the following data define a functor $F: \mathscr{C} \to \mathscr{C}$.

(a) The map $F: \text{Ob } \mathscr{C} \to \text{Ob } \mathscr{C}$ is the identity map.

(b) For each pair of objects C_1 and C_2 in $\mathscr{C}$, the map $F: \mathscr{C}(C_1, C_2) \to \mathscr{C}(C_1, C_2)$ is the identity map.

This functor $F: \mathscr{C} \to \mathscr{C}$ is called the **identity functor** and is usually denoted by $\text{id}_{\mathscr{C}}$.

(13) Show that the following data give a functor F from the category of all monoids, Monoid, to the category of all sets, Sets.

(a) $F: \text{Ob Monoid} \to \text{Ob Sets}$ is given by $F(M)$ is the underlying set of the monoid M.

(b) For each pair of monoids M_1, M_2, the map $F: \text{Monoid}(M_1, M_2) \to \text{Sets}(F(M_1), F(M_2))$ is given by $F(f): F(M_1) \to F(M_2)$ is the map of sets given by viewing the morphism of monoids $f: M_1 \to M_2$ simply as a map of sets.

This functor $F: \text{Monoid} \to \text{Sets}$ is called the **forgetful functor**.

(14) Show that for the categories, Group, ordered sets, and G-sets, there are functors from each of them to sets which are analogs of the forgetful functor we just defined from Monoid to Sets.

(15) Let C be an object in the category $\mathscr{C}$. For each morphism $f: X \to Y$ in $\mathscr{C}$, define the map $(C, f): (C, X) \to (C, Y)$ by $(C, f)(g) = fg$ for all g in (C, X). Show that the following data define a functor $(C, .): \mathscr{C} \to \text{Sets}$:

(a) $(C, .): \text{Ob } C \to \text{Ob Sets}$ is given by $(C, .)(X) = \mathscr{C}(C, X)$ for each object X in $\mathscr{C}$.

(b) For each pair of objects X and Y in $\mathscr{C}$, define $(C, .): (X, Y) \to ((C, X), (C, Y))$ by $(C, .)(f) = (C, f)$ for each f in (X, Y). The functor $(C, .): \mathscr{C} \to \text{Sets}$ is called the functor from $\mathscr{C}$ to Sets **represented by** C.

(16) Let $\mathscr{C}$ and $\mathscr{D}$ be categories. Show that each functor $F:\mathscr{C}^{op}\to\mathscr{D}$ can be described in terms of the categories $\mathscr{C}$ and $\mathscr{D}$ as follows:

(a) F is a map from Ob $\mathscr{C}$ to Ob $\mathscr{D}$ together with

(b) maps $F:\mathscr{C}(C_1, C_2)\to\mathscr{D}(F(C_2), F(C_1))$ for each pair of objects C_1 and C_2 in $\mathscr{C}$ satisfying:

- **(i)** $F(\mathrm{id}_C)=\mathrm{id}_{F(C)}$ for all C in $\mathscr{C}$ and
- **(ii)** For each triple C_1, C_2, C_3 of objects in $\mathscr{C}$ and morphisms $C_1\xrightarrow{f}C_2$ and $g:C_2\to C_3$, we have $F(gf)=F(f)F(g)$.

Definition

Let $\mathscr{C}$ and $\mathscr{D}$ be categories. A **contravariant functor** $F:\mathscr{C}\to\mathscr{D}$ is

(a) a map $F:\mathrm{Ob}\,\mathscr{C}\to\mathrm{Ob}\,\mathscr{D}$ together with

(b) maps $F:(C_1, C_2)\to(F(C_2), F(C_1))$ for each pair of objects C_1 and C_2 in $\mathscr{C}$ satisfying condition (b) above.

(17) Show that the contravariant functors from a category $\mathscr{C}$ to a category $\mathscr{D}$ are the same thing as the functors from the category $\mathscr{C}^{op}$ to the category $\mathscr{D}$.

We now give some examples of contravariant functors from a category $\mathscr{C}$ to a category $\mathscr{D}$. Of course, in view of Exercise 17 these are nothing more than examples of functors from $\mathscr{C}^{op}$ to $\mathscr{D}$.

(18) Show that the following data describe a contravariant functor $F:\mathscr{C}\to\mathscr{C}^{op}$.

(a) $F:\mathrm{Ob}\,\mathscr{C}\to\mathrm{Ob}\,\mathscr{C}^{op}$ is the identity map.

(b) $F:\mathscr{C}(C_1, C_2)\to\mathscr{C}^{op}(F(C_2), F(C_1))$ is the identity map.

Show that viewed as a functor from $\mathscr{C}^{op}$ to $\mathscr{C}^{op}$, the contravariant functor F is nothing more or less than the identity functor on $\mathscr{C}^{op}$.

(19) Let C be an object in a category $\mathscr{C}$. For each morphism $f:X\to Y$ in $\mathscr{C}$, define the map $(f, C):(Y, C)\to(X, C)$ by $(f, C)(g)=gf$. Show that the following data define a contravariant functor $(., C):\mathscr{C}\to\mathrm{Sets}$:

(a) $(., C):\mathrm{Ob}\,\mathscr{C}\to\mathrm{Ob}\,\mathrm{Sets}$ is given by $(., C)(X)=\mathscr{C}(X, C)$ for each object X in $\mathscr{C}$.

(b) For each pair of objects X and Y in $\mathscr{C}$ the map $(., C):\mathscr{C}(X, Y)\to((Y, C), (X, C))$ is given by $(., C)(f)=(f, C)$ for all f in $\mathscr{C}(X, Y)$.

For each object C in $\mathscr{C}$, the contravariant function $(, C):\mathscr{C}\to\mathrm{Sets}$ is called the **contravariant functor represented by** C.

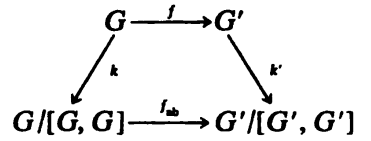
(20) Suppose G is a group. The **commutator subgroup** of G, which is usually denoted by $[G, G]$, is defined to be the subgroup of G generated by all elements of the form $xyx^{-1}y^{-1}$ with x and y elements of G. Show:

(a) $[G, G]$ is a normal subgroup of G having the following properties:

- **(i)** $[G, G]=\{1\}$ if and only if G is abelian.
- **(ii)** $G/[G, G]$ is an abelian group.
- **(iii)** A normal subgroup H of G has the property G/H is abelian if and only if $H\supset[G, G]$.
- **(iv)** The canonical epimorphism $k:G\to G/[G, G]$ has the property that for each abelian group X, the map $\phi_X:(G/[G, G], X)\to(G, X)$, given by $\phi_X(g)=gk$ for all group morphisms $g:G/[G, G]\to X$, is an isomorphism of sets.
- **(v)** If $f:G\to G'$ is a morphism of groups, then $f([G, G])\subset[G', G']$.

Hence, there is a unique morphism of groups $f_{ab}:G/[G, G]\to G'/[G', G']$

which makes the diagram



commute.

(b) Show that the following data define a functor $F:\text{Groups}\to\mathcal{Ab}$.

(i) $F:\text{Ob Groups}\to\text{Ob }\mathcal{Ab}$ is given by $F(G)=G/[G,G]$.

(ii) For each pair of groups G_1 and G_2, the map $F:(G_1,G_2)\to(F(G_1),F(G_2))$ is given by $F(f)=f_{ab}$ for all group morphisms $f:G_1\to G_2$.

(21) Suppose $F:\mathcal{C}\to\mathcal{D}$ and $G:\mathcal{D}\to\mathcal{E}$ are functors of categories. Show that the following data define a functor $GF:\mathcal{C}\to\mathcal{E}$.

(a) GF: Ob $\mathcal{C}\to$Ob $\mathcal{E}$ is given by $(GF)(C)=G(F(C))$ for all C in Ob $\mathcal{C}$.

(b) $GF:\mathcal{C}(C_1,C_2)\to\mathcal{E}(GF(C_1),GF(C_2))$ is the composition of the following maps $\mathcal{C}(C_1,C_2)\xrightarrow{F}\mathcal{D}(F(C_1),F(C_2))\xrightarrow{G}\mathcal{E}(GF(C_1),GF(C_2))$.

The functor GF is called the **composition of the functor F followed by G.**

Show that if $F:\mathcal{C}\to\mathcal{D}$, $G:\mathcal{D}\to\mathcal{E}$, and $H:\mathcal{E}\to\mathcal{F}$ are functors of categories, then $(HG)F=H(GF)$.

(22) A category $\mathcal{C}$ is said to be a **small category** if the collection Ob $\mathcal{C}$ of objects of $\mathcal{C}$ is a set.

(a) Show that if $\mathcal{C}$ and $\mathcal{D}$ are two small categories, then the collection $(\mathcal{C},\mathcal{D})$ of all functors from $\mathcal{C}$ to $\mathcal{D}$ is a set.

(b) Show that the following data define a category Cat called the **category of all small categories**:

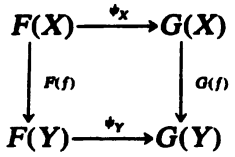
(i) Ob Cat is the collection of all small categories.

(ii) For each pair of small categories $\mathcal{C}_1$ and $\mathcal{C}_2$ in Cat we define $\text{Cat}(\mathcal{C}_1,\mathcal{C}_2)$ to be the set of all functors from $\mathcal{C}_1$ to $\mathcal{C}_2$.

(iii) The composition maps $\text{Cat}(\mathcal{C}_1,\mathcal{C}_2)\times\text{Cat}(\mathcal{C}_2,\mathcal{C}_3)\to\text{Cat}(\mathcal{C}_1,\mathcal{C}_3)$ are given by $(F,G)\to GF$, the composition of the functor F followed by G.

Definition

Suppose F and G are two functors from the category $\mathcal{C}$ to the category $\mathcal{D}$. A **morphism ψ from F to G**, which we denote by $\psi:F\to G$, consists of a family $\{\psi_x\}_{x\in\text{Ob }\mathcal{C}}$ of morphisms $\psi_X:F(X)\to G(X)$ in $\mathcal{D}$, one for each object X in Ob $\mathcal{C}$, satisfying the condition that for each morphism $f:X\to Y$ in $\mathcal{C}$, the diagram



commutes.

(23) Suppose $\mathcal{C}$ and $\mathcal{D}$ are categories and $F,G,H,I:\mathcal{C}\to\mathcal{D}$ are functors.

(a) If $\psi:F\to G$ and $\phi:G\to H$ are morphisms of functors, show that one obtains a morphism of functors $\phi\psi:F\to H$ by defining $(\phi\psi)_X:F(X)\to H(X)$ to be the

composition $F(X)\xrightarrow{\phi_X}G(X)\xrightarrow{\psi_X}H(X)$ for each X in $\mathrm{Ob}\,\mathscr{C}$. The morphism $\phi\psi: F\to H$ is called the **composition of the morphism ψ followed by ϕ.**

(b) Show that if $\alpha: F\to G$, $\beta: G\to H$, and $\gamma: H\to I$ are morphisms of functors, then $(\gamma\beta)\alpha = \gamma(\beta\alpha)$.

(24) Suppose that $\mathscr{C}$ is a small category and $\mathscr{D}$ is an arbitrary category.

(a) Show that if $F, G: \mathscr{C}\to\mathscr{D}$ are two functors, then (F, G) the collection of all morphisms from F to G is a set.

(b) The following data define a category $(\mathscr{C}, \mathscr{D})$ called the **category of functors from $\mathscr{C}$ to $\mathscr{D}$.**

(i) $\mathrm{Ob}(\mathscr{C}, \mathscr{D})$ is the collection of all functors from $\mathscr{C}$ to $\mathscr{D}$.

(ii) Given two functors F, G in $\mathrm{Ob}(\mathscr{C}, \mathscr{D})$ define $(\mathscr{C}, \mathscr{D})(F, G) = (F, G)$, the set of all morphisms from F to G.

(iii) Given a triple F, G, H in $\mathrm{Ob}(\mathscr{C}, \mathscr{D})$, the composition map $(F, G)\times(G, H)\to(F, H)$ is given by $(\alpha, \beta)\to\beta\alpha$ where $\beta\alpha$ is the composition of the morphisms of functors α followed by β.

The category $(\mathscr{C}, \mathscr{D})$ is called the **category of functors from $\mathscr{C}$ to $\mathscr{D}$.**

(25) Let $F, G: \mathscr{C}\to\mathscr{D}$ be functors. A morphism $\alpha: F\to G$ is an isomorphism if and only if there is a morphism $\beta: G\to F$ such that $\beta\alpha = \mathrm{id}_F$ and $\alpha\beta = \mathrm{id}_G$.

(a) Show that if $\alpha: F\to G$ is an isomorphism of functors, then there is only one morphism $\beta: G\to F$ such that $\beta\alpha = \mathrm{id}_F$ and $\alpha\beta = \mathrm{id}_G$. This uniquely determined morphism $\beta: G\to F$ is also an isomorphism of functors called the inverse of α and often denoted by α^{-1}.

(b) Show that a morphism $\alpha: F\to G$ of functors is an isomorphism if and only if for each X in $\mathscr{C}$ we have that $\alpha_X: F(X)\to G(X)$ is an isomorphism in $\mathscr{D}$.

(c) Suppose $F, G, H: \mathscr{C}\to\mathscr{D}$ are functors from $\mathscr{C}$ to $\mathscr{D}$ and $\alpha: F\to G$ and $\beta: G\to H$ are morphisms of functors.

(i) If α is an isomorphism, then $(\alpha^{-1})^{-1} = \alpha$.

(ii) If α and β are isomorphisms, then $\beta\alpha$ is an isomorphism.

(iii) If $\beta\alpha$ is an isomorphism, then β is an isomorphism if and only if α is an isomorphism.

(26) Let C_1, C_2 be objects in a category $\mathscr{C}$. Suppose $f: C_2\to C_1$ is a morphism in $\mathscr{C}$. Then for each X in $\mathscr{C}$ we have the map of sets $(f, X): (C_1, X)\to(C_2, X)$ given by $(f, X)(g) = gf$ for all g in (C_1, X).

(a) Show that the family $\{(f, X)\}_{X\in\mathrm{Ob}\,\mathscr{C}}$ is a morphism from the functor $(C_1, .): \mathscr{C}\to$ Sets to the functor $(C_2, .): \mathscr{C}\to$Sets. This morphism is denoted by $(f, .): (C_1, .)\to(C_2, .)$.

(b) Show that for each C in $\mathscr{C}$, the morphism $(\mathrm{id}_C, .): (C, .)\to(C, .)$ of functors is $\mathrm{id}_{(C,.)}$.

(c) Let $f: C_1\to C_2$ be a morphism in $\mathscr{C}$. Then the morphism $(f, .): (C_2, .)\to(C_1, .)$ of functors has the property that $(f, C_2)(\mathrm{id}_{C_2}) = f$. Hence, if $f, f': C_1\to C_2$ are morphisms in $\mathscr{C}$, then $(f, .) = (f', .)$ if and only if $f = f'$.

(d) Show that if $f: C_1\to C_2$ and $g: C_2\to C_3$ are morphisms in $\mathscr{C}$, then $(gf, .) = (f, .)(g, .)$.

(e) Show that if $f: C_1\to C_2$ is an isomorphism in $\mathscr{C}$, then $(f, .): (C_2, .)\to(C_1, .)$ is an isomorphism of functors with $(f, .)^{-1} = (f^{-1}, .)$.

(f) Show that a morphism $f: C_1\to C_2$ in $\mathscr{C}$ is an isomorphism if and only if the

morphism of functors $(f, .):(C_2, .)\to(C_1, .)$ is an isomorphism of functors. [*Hint*: If $(f, .):(C_2, .)\to(C_1, .)$ is an isomorphism, then $(f, C_1):(C_2, C_1)\to(C_1, C_1)$ is an isomorphism of sets. Hence, there is a $g:C_2\to C_1$ such that $(f, C_1)(g) = \mathrm{id}_{C_1}$. Show that $f:C_1\to C_2$ is an isomorphism by showing that $gf = \mathrm{id}_{C_1}$ and $fg = \mathrm{id}_{C_2}$.]

(g) Show that a morphism $f:C_1\to C_2$ in $\mathscr{C}$ is an epimorphism if and only if $(f, X):(C_2, X)\to(C_1, X)$ is an injective map of sets for each X in Ob $\mathscr{C}$.

(27) Suppose $F:\mathscr{C}\to$Sets is an arbitrary functor and C is an object in $\mathscr{C}$. Let $((C, .), F)$ denote the collection of all morphisms from $(C, .)$ to F. We want to show that the map $((C, .), F)\to F(C)$ given by $\alpha\to\alpha_C(\mathrm{id}_C)$ is bijective.

(a) Suppose $\alpha:(C, .)\to F$ is a morphism of functors. Show that for each object X in $\mathscr{C}$ and each morphism $f:C\to X$ the diagram

$$\begin{array}{ccc} (C, C) & \xrightarrow{\alpha_C} & F(C) \\ \downarrow{\scriptstyle (C, f)} & & \downarrow{\scriptstyle F(f)} \\ (C, X) & \xrightarrow{\alpha_X} & F(X) \end{array}$$

commutes. From this deduce that $\alpha_X(f) = F(f)\alpha_C(\mathrm{id}_C)$ for all f in (C, X). This result implies:

(b) If $\alpha, \beta:(C, .)\to F$ are two morphisms of functors, then $\alpha = \beta$ if and only if $\alpha_C(\mathrm{id}_C) = \beta_C(\mathrm{id}_C)$. Hence, the map $((C, .), F)\to F(C)$ given by $\alpha\to\alpha_C(\mathrm{id}_C)$ is an injective map.

(c) Suppose x is an element of $F(C)$. For each object X in $\mathscr{C}$, define a map $\alpha_X:(C, X)\to F(X)$ by setting $\alpha_X(f) = F(f)(x)$ for each f in (C, X). Show that the collection $\{a_X\}_{X\in \mathrm{Ob}\mathscr{C}}$ is a morphism $\alpha:(C, .)\to F$ with the property $\alpha_C(\mathrm{id}_C) = x$. Hence:

(d) The map $((C, .), F)\to F(C)$ given by $\alpha\to\alpha_C(\mathrm{id}_C)$ is bijective. The isomorphism of sets $((C, .), F)\to F(C)$ given by $\alpha\to\alpha_C(\mathrm{id}_C)$ is called the **Yoneda isomorphism** and is generally considered an identification. It is a basic tool in almost all work involving functors.

(28) Let C_1, C_2 be objects in a category $\mathscr{C}$. Show that for each morphism $\alpha:(C_2, .)\to(C_1, .)$ there is a unique morphism $f:C_1\to C_2$ such that $\alpha = (f, .)$.

(29) Let $\mathscr{C}$ be a category, $F, G:\mathscr{C}\to$Sets functors, and $\alpha:F\to G$ a morphism of functors. If C is an object of $\mathscr{C}$ show that the diagram

$$\begin{array}{ccc} ((C, .), F) & \longrightarrow & F(C) \\ \downarrow{\scriptstyle ((C, .), \alpha)} & & \downarrow{\scriptstyle \alpha_C} \\ ((C, .), G) & \longrightarrow & G(C) \end{array}$$

commutes where $((C, .), \alpha)(\beta) = \alpha\beta$ for all morphisms $\beta:(C, .)\to F$ and where the horizontal maps are the Yoneda isomorphisms.

Suppose $f: C \to C'$ is a morphism in $\mathscr{C}$. Show that the diagram

$$\begin{array}{ccc} ((C, .), F) & \longrightarrow & F(C) \\ \downarrow{\scriptstyle ((f, .), F)} & & \downarrow{\scriptstyle F(f)} \\ ((C', .), F) & \longrightarrow & F(C') \end{array}$$

commutes where $((f, .), F)\beta = \beta(f, .)$ for all morphisms $\beta: (C, .) \to F$ and where the horizontal morphisms are the Yoneda isomorphisms.

(30) Let $\mathscr{C}$ be a category. A functor $F: \mathscr{C} \to$ Sets is said to be a **representable functor** if there is a C in $\mathscr{C}$ such that F is isomorphic to the functor $(C, .)$.

Suppose a functor $F: \mathscr{C} \to$ Sets is representable and we are given two isomorphisms $\alpha: F \to (C, .)$ and $\alpha': F \to (C', .)$. Show that there is a unique morphism $f: C' \to C$ in $\mathscr{C}$ such that $(f, .)\alpha = \alpha'$ and this uniquely determined morphism $f: C' \to C$ is an isomorphism.

(31) Let $\mathscr{C}$ be a category and $\{F_j\}_{j \in J}$ an indexed family of functors from $\mathscr{C} \to$ Sets.

(a) Show that the following data define a functor $\prod_{j \in J} F_j: \mathscr{C} \to$ Sets:

(i) $(\prod_{j \in J} F_j)(X) = \prod_{j \in J} F_j(X)$ for each X in Ob $\mathscr{C}$.

(ii) If $f: X \to Y$ is a morphism in $\mathscr{C}$, define

$$\left(\prod F_i\right)(f): \prod_{j \in J} F_j(X) \to \prod_{j \in J} F_j(Y) \qquad \text{by} \qquad \left(\prod_{j \in J} F_j\right)(f) = \prod_{j \in J} F_j(f)$$

(b) Suppose k is in J. For each X in Ob $\mathscr{C}$ define the map $(p_k)_X: (\prod_{j \in J} F_j)(X) \to F_k(X)$ to be the kth projective map from $\prod_{j \in J} F_j(X) \to F_k(X)$. Show that the family $\{(p_k)_X\}_{X \in \text{Ob}\,\mathscr{C}}$ is a morphism of functors $p_k: \prod_{j \in J} F_j \to F_k$.

(c) Show that if $G: \mathscr{C} \to$ Sets is an arbitrary functor and $\{\alpha_j: G \to F_j\}_{j \in J}$ is an arbitrary family of morphisms, then there is a unique morphism $\alpha: G \to \prod_{j \in J} F_j$ such that $p_k\alpha = \alpha_k$ for each k in J. We denote this uniquely determined morphism α by $\prod_{j \in J} \alpha_j$.

(d) Show that given any morphism $\beta: G \to \prod_{j \in J} F_j$, $\beta = \prod_{j \in J} p_j\beta$. In view of these results, it is reasonable to call the functor $\prod_{j \in J} F_j$ the product of the indexed family $\{F_j\}_{j \in J}$ and the morphisms $p_k: \prod_{j \in J} F_j \to F_k$, the kth projection morphisms.

(32) Let $\mathscr{C}$ be a category and $\{C_i\}_{i \in I}$ an indexed family of objects in $\mathscr{C}$.

(a) Show that a family of morphisms $\{f_i: C_i \to C\}_{i \in I}$ is a sum for the family $\{C_i\}_{i \in I}$ if and only if the morphism $\Pi(f_i): (C, .) \to \prod_{i \in I}(C_i, .)$ is an isomorphism of functors.

(b) Show that if C is an object of $\mathscr{C}$, then a morphism $\alpha: (C, .) \to \prod_{i \in I}(C_i, .)$ is an isomorphism if and only if the uniquely determined family of morphisms $\{f_i: C \to C_i\}_{i \in I}$ such that $p_k\alpha = (f_k, .)$ is a sum for the indexed family $\{C_i\}_{i \in I}$.

(c) The indexed family $\{C_i\}_{i \in I}$ has a sum in $\mathscr{C}$ if and only if the functor $\Pi(C_i, .): \mathscr{C} \to$ Sets is representable.

Definition
We say that an object C in $\mathscr{C}$ is isomorphic to a sum of the family $\{C_i\}_{i\in I}$ if and only if there is a family of morphisms $\{f_i: C_i \to C\}_{i\in I}$ which is a sum for the family $\{C_i\}_{i\in I}$.
(d) An object C in $\mathscr{C}$ is isomorphic to a sum of $\{C_i\}_{i\in I}$ if and only if the functor $(C, .)$ is isomorphic to the functor $\Pi(C_i, .)$.

(33) Let $\{C_i\}_{i\in I}$ be an indexed family of objects in a category $\mathscr{C}$. Show that a family of morphisms $\{f_i: C \to C_i\}_{i\in I}$ is a product in $\mathscr{C}$ if and only if the corresponding family of morphisms $\{f_i: C_i \to C\}_{i\in I}$ in $\mathscr{C}^{op}$ is a sum in $\mathscr{C}^{op}$ for the indexed family $\{C_i\}_{i\in I}$ of objects in $\mathscr{C}^{op}$. Restate the results of Exercise **32** for products of indexed families of objects in a category $\mathscr{C}$ using contravariant functors. More generally, restate for contravariant functors the results obtained for functors in Exercises **15** through **32**.

Definitions
Suppose $F: \mathscr{C} \to \mathscr{D}$ is a functor.
(a) F is said to be a **faithful functor** if $F: (C_1, C_2) \to (F(C_1), F(C_2))$ is an injective map for each pair of objects C_1 and C_2 in $\mathscr{C}$.
(b) F is said to be a **full functor** if $F: (C_1, C_2) \to (F(C_1), F(C_2))$ is surjective for all pairs of objects C_1 and C_2 in $\mathscr{C}$.
(c) F is said to be a **fully faithful functor** if F is both full and faithful.
(d) F is said to be **dense** if for each object D in $\mathscr{D}$ there is an object C in $\mathscr{C}$ such that $D \approx F(C)$.

(34) Let $\mathscr{C}$ be the category of ordered sets and $F: \mathscr{C} \to$ Sets the forgetful functor which assigns to each ordered set X its underlying set. Show that F is faithful and dense but not full. Show that the forgetful functors from Monoid, Group, G-Sets to Sets are all faithful and dense but not in general full.
(35) Let $\mathscr{C}$ be a small category. Let $\mathscr{C} \to (\mathscr{C}^{op}, \text{Sets})$ be the functor given by $C \to (., C)$ where $(., C)$ is the functor represented by C in $\mathscr{C}^{op}$. Show that this functor $\mathscr{C} \to (\mathscr{C}^{op}, \text{Sets})$ is a fully faithful functor which need not be dense.
(36) Show that the functor Group $\to$ Ab given by $G \to G/[G, G]$ is not full or faithful but is dense.

Definitions
Let $F: \mathscr{C} \to \mathscr{D}$ be a functor of categories.
(a) F is said to be an **isomorphism of categories** if there is a functor $G: \mathscr{D} \to \mathscr{C}$ such that $GF = \mathrm{id}_{\mathscr{C}}$ and $FG = \mathrm{id}_{\mathscr{D}}$.
(b) F is said to be an **equivalence of categories** if there is a functor $G: \mathscr{D} \to \mathscr{C}$ such that $GF \approx \mathrm{id}_{\mathscr{C}}$ and $FG \approx \mathrm{id}_{\mathscr{D}}$.

(37) Let $F: \mathscr{C} \to \mathscr{D}$ be a functor of categories. Show:
(a) If F is an isomorphism of categories, then there is one and only one functor $G: \mathscr{D} \to \mathscr{C}$ such that $GF = \mathrm{id}_{\mathscr{C}}$ and $FG = \mathrm{id}_{\mathscr{D}}$. If F is an isomorphism of

categories, then the uniquely determined functor $G: \mathcal{D} \to \mathcal{C}$ such that $GF = \mathrm{id}_{\mathcal{C}}$ and $GF = \mathrm{id}_{\mathcal{D}}$ is also an isomorphism of categories which is called the inverse of F and is denoted by F^{-1}.

(b) Let Cat be the category of all small categories, that is, categories whose collection of objects is a set. Show that the following data define a functor $\alpha: \mathrm{Cat} \to \mathrm{Cat}$ which is an isomorphism of categories with $\alpha^{-1} = \alpha$.

(i) $\alpha: \mathrm{Ob\ Cat} \to \mathrm{Ob\ Cat}$ is given by $\alpha(\mathcal{C}) = \mathcal{C}^{\mathrm{op}}$ for each category $\mathcal{C}$ in Cat.

(ii) $\alpha: (\mathcal{C}_1, \mathcal{C}_2) \to (\mathcal{C}_1^{\mathrm{op}}, \mathcal{C}_2^{\mathrm{op}})$ is defined as follows: For each functor $F: \mathcal{C}_1 \to \mathcal{C}_2$, the functor $\alpha(F): \mathcal{C}_1^{\mathrm{op}} \to \mathcal{C}_2^{\mathrm{op}}$ is given by the data $\alpha(F): \mathrm{Ob}\ \mathcal{C}_1^{\mathrm{op}} \to \mathrm{Ob}\ \mathcal{C}_2^{\mathrm{op}}$ is the map $F: \mathrm{Ob}\ \mathcal{C}_1 \to \mathrm{Ob}\ \mathcal{C}_2$ (remember $\mathrm{Ob}\ \mathcal{C} = \mathrm{Ob}\ \mathcal{C}^{\mathrm{op}}$) while $\alpha(F): \mathcal{C}_1^{\mathrm{op}}(C_1, C_2) \to \mathcal{C}_2^{\mathrm{op}}(F(C_1), F(C_2))$ is the map $F: \mathcal{C}_1(C_2, C_1) \to \mathcal{C}_2(F(C_2), F(C_1))$[remember $\mathcal{C}^{\mathrm{op}}(C_1, C_2) = \mathcal{C}(C_2, C_1)$].

(c) Show that the following data define a functor $\beta: \mathrm{Monoid} \to \mathrm{Monoid}$ which is an isomorphism of categories with $\beta^{-1} = \beta$.

(i) $\beta(M) = M^{\mathrm{op}}$ for each monoid M in Ob Monoid.

(ii) If $f: M_1 \to M_2$ is a morphism of monoids, define $\beta(f): M_1^{\mathrm{op}} \to M_2^{\mathrm{op}}$ by $\beta(f)(m) = f(m)$ for all m in M_1^{op}. (Remember as sets $M_1^{\mathrm{op}} = M_1$.)

(d) Define analog isomorphisms for the categories of Groups and Ordered Sets.

(38) Show that a functor $F: \mathcal{C} \to \mathcal{D}$ of categories is an equivalence of categories if and only if it is a fully faithful dense functor.

(39) Let $\mathcal{D}$ be the full subcategory of Cat consisting of those small categories with one object.

(a) Show that the following data define a functor $F: \mathrm{Monoid} \to \mathcal{D}$:

(i) $F: \mathrm{Ob\ Monoid} \to \mathrm{Ob}\ \mathcal{D}$, given by $F(M)$, is the category $\mathcal{C}(M)$ of M for each monoid M.

(ii) If $f: M_1 \to M_2$ is a morphism of monoids, then $F(f): \mathcal{C}(M_1) \to \mathcal{C}(M_2)$ is the functor $F(f): \mathcal{C}(M_1)(M_1, M_1) \to \mathcal{C}(M_2)(M_2, M_2)$ given by $F(f)(m_1) = f(m_1)$ for all m_1 in $\mathcal{C}(M_1)(M_1, M_1)$ [remember $\mathcal{C}(M)(M, M) = M$ for all monoids M].

(b) Show that the following data define a functor $G: \mathcal{D} \to \mathrm{Monoid}$:

(i) $G: \mathrm{Ob}\ \mathcal{D} \to \mathrm{Ob\ Monoid}$ is given by $G(\mathcal{C}) = M(\mathcal{C})$, where $M(\mathcal{C})$ is the monoid of the category $\mathcal{C}$.

(ii) If $H: \mathcal{C}_1 \to \mathcal{C}_2$ is a morphism in $\mathcal{D}$, then define $G(H): M(\mathcal{C}_1) \to M(\mathcal{C}_2)$ by $G(H)(f) = H(f)$ for all f in $M(\mathcal{C}_1)$ [remember that $M(\mathcal{C}) = \mathcal{C}(X, X)$ where X is the unique object of $\mathcal{C}$ for each $\mathcal{C}$ in $\mathcal{D}$].

(c) Show that $GF = \mathrm{id}_{\mathrm{Monoid}}$ while $FG \approx \mathrm{id}_{\mathcal{D}}$. Hence, F and G are equivalences of categories.

(40) Let $\mathcal{E}$ be the full category of Cat whose objects are those small categories $\mathcal{C}$ satisfying $\mathcal{C}(C_1, C_2)$ is either empty or consists of a single element for all pairs of objects C_1 and C_2 in $\mathcal{C}$ and $C_1 = C_2$ if both $\mathcal{C}(C_1, C_2)$ and $\mathcal{C}(C_2, C_1)$ are not empty. Let $\mathcal{O}$ be the category of ordered sets. Show that $\mathcal{O}$ and $\mathcal{E}$ are equivalent categories.

(41) Let G be a group. Let $\mathcal{C}(G)$ be the category of G and $(\mathcal{C}(G), \mathrm{Sets})$ the category of all functors from $\mathcal{C}(G)$ to Sets.

(a) Let $F: \mathcal{C}(G) \to \mathrm{Sets}$ be a functor. Show that associated with F is the G-set which we denote by $\alpha(F)$ which consists of:

(i) The set $F(G)$.

(ii) For each g in G and s in $F(G)$ gs is defined to be $F(g)(s)$ where $F(g)$ is the image of g in $\mathrm{End}(F(G))$ under the map $f: \mathscr{C}(G)(G, G) \to \mathrm{Sets}(F(G), F(G))$ [remember that $G = \mathscr{C}(G)(G, G)$ as a set].

(b) Suppose $F_1, F_2: \mathscr{C}(G) \to \mathrm{Sets}$ are functors and $f: F_1 \to F_2$ is a morphism of functors. Show that the map $f: F_1(G) \to F_2(G)$ is a morphism of G-sets where $F_1(G)$ and $F_2(G)$ are considered the G-sets $\alpha(F_1)$ and $\alpha(F_2)$ described in (a).

(c) Show that the following data define a functor $\alpha: (\mathscr{C}(G), \mathrm{Sets}) \to G\text{-Sets}$ which is an equivalence of categories.

(i) $\alpha: \mathrm{Ob}(\mathscr{C}(G), \mathrm{Sets}) \to \mathrm{Ob}(G\text{-sets})$ is given by $\alpha(F)$ is the G-set described in (a).

(ii) $\alpha: (F_1, F_2) \to (\alpha(F_1), \alpha(F_2))$ is given by $\alpha(f) = f$ for all morphisms of functors $f: F_1 \to F_2$.

(42) Prove that if $\mathscr{C}$ is a category and X is in $\mathrm{Ob}(\mathscr{C})$, then (X, X) is a monoid.

(43) Let $\mathscr{C}$ and $\mathscr{D}$ be categories, let $F: \mathscr{C} \to \mathscr{D}$ be a functor, and let X be an object of $\mathscr{D}$. Show that the following set of data defines a category, which we denote by (F, X). The objects of (F, X) are all pairs (y, f) where y is an object of $\mathscr{C}$ and f is in $\mathscr{D}(F(y), X)$, that is, $f: F(y) \to X$ is a morphism from $F(y)$ to X in the category $\mathscr{D}$. If (y_1, f_1) and (y_2, f_2) are objects of (F, X), a morphism $g: (y_1, f_1) \to (y_2, f_2)$ is defined to be a morphism $g: y_1 \to y_2$ in $\mathscr{C}$ such that $f_2 F(g) = f_1$. If $g_1: (y_1, f_1) \to (y_2, f_2)$ and $g_2: (y_2, f_2) \to (y_3, f_3)$ are morphisms in (F, X), the composition is defined to be the composition $g_2 g_1$ in the category $\mathscr{C}$. When $\mathscr{C} = \mathscr{D}$ and F is the identity, we denote (F, X) by $(\mathscr{C}, X)$.

(44) Let $\mathscr{C}$ and $\mathscr{D}$ be categories, let $F: \mathscr{C} \to \mathscr{D}$ be a functor, and let X be an object of $\mathscr{D}$. Proceeding as in the foregoing exercise, construct a category (X, F) whose objects are all pairs (y, f) where y is in $\mathrm{Ob}(\mathscr{C})$ and $f: X \to F(Y)$ is a morphism in $\mathscr{D}$. When $\mathscr{C} = \mathscr{D}$ and F is the identity, we denote (X, F) by $(X, \mathscr{C})$.

(45) Let $\mathscr{C}$ be a category. Show that the following set of data defines a category which we shall call $T_2(\mathscr{C})$. The objects of $T_2(\mathscr{C})$ are the morphisms $f: X \to Y$ of $\mathscr{C}$. If $f_1: X_1 \to Y_1$ and $f_2: X_2 \to Y_2$ are objects of $T_2(\mathscr{C})$, a morphism $g: f_1 \to f_2$ in $T_2(\mathscr{C})$ is defined to be a pair of morphisms (g_1, g_2) in $\mathscr{C}$ where $g_1: X_1 \to X_2$, $g_2: Y_1 \to Y_2$, and $g_2 f_1 = f_2 g_1$. Composition of morphisms in $T_2(\mathscr{C})$ are defined in the obvious way using the composition of morphisms in $\mathscr{C}$.

(46) Let $\mathscr{C}$ be a category. Show that the following data describe a category which we shall denote by $\mathscr{C}[X]$. The objects of $\mathscr{C}[X]$ are the endomorphisms of $\mathscr{C}$, that is, morphisms $f: Y \to Y$ for all objects Y in $\mathscr{C}$. If $f_1: Y_1 \to Y_1$ and $f_2: Y_2 \to Y_2$ are objects of $\mathscr{C}[X]$, a morphism $g: f_1 \to f_2$ in $\mathscr{C}[X]$ is a morphism $g: Y_1 \to Y_2$ in $\mathscr{C}$ such that $g f_1 = f_2 g$. Composition of morphisms in $\mathscr{C}[X]$ is defined in the obvious way. Identify the category $\mathscr{C}[X]$ with a suitable subcategory of $T_2(\mathscr{C})$.

(47) Prove Basic Properties 2.1.

(48) Prove Basic Properties 2.3.

(49) Prove that products and sums do not generally exist in the category of finite sets.

(50) Prove Basic Property 3.2.

(51) Prove that the sum of an infinite family of nontrivial abelian groups does not exist in the category of finite abelian groups.

(52) Show that the following data define a functor $F:\mathrm{Sets}[X]\to\mathrm{Sets}$:

(a) If (S, f) is an object of $\mathrm{Sets}[X]$, that is, S is a set and $f:S\to S$ is an endomorphism of S, then $F((S, f))=S$.

(b) Given a morphism $g:(S_1, f_1)\to(S_2, f_2)$, that is, g is a map $S_1\to S_2$ satisfying $f_2g=gf_1$, then $F(g):F(S_1)\to F(S_2)$ is simply the map $g:S_1\to S_2$. Show that the functor F is representable.

Chapter 4 RINGS

In Chapter 2 we defined rings. In this chapter, we subject the category of rings to the same type of analysis that we applied to the categories of sets, monoids, and groups. In the course of this analysis we introduce polynomial rings over commutative rings and show in particular that the ring of polynomials over the ring of integers plays an analogous role in the category of rings to that played by **Z** in the category of groups and by **N** in the category of monoids.

1. CATEGORY OF RINGS

We now recall the definition of a ring.

Definition
A **ring** is a set R together with two laws of composition, addition written $r_1 + r_2$ and multiplication written $r_1 r_2$, which satisfy:

(a) R is a commutative group under addition with identity denoted by 0.
(b) R is a monoid under multiplication, not necessarily commutative, with identity element 1.
(c) For all elements r_1, r_2, and r_3 in R we have:
  (i) $r_1(r_2 + r_3) = r_1 r_2 + r_1 r_3$.
  (ii) $(r_1 + r_2)r_3 = r_1 r_3 + r_2 r_3$.

Finally, a ring R is said to be **commutative** if R is a commutative monoid under multiplication.

We assume the reader is familiar with the following easily verified properties.

Basic Properties 1.1
Let R be a ring. Then:

(a) $r0 = 0 = 0r$ for all r in R.
(b) $r_1(-r_2) = -(r_1r_2) = (-r_1)r_2$ for all elements r_1 and r_2 in R.

The reader should note that the definition of a ring does not preclude the possibility that $0 = 1$. A consequence of the above basic properties is that if $0 = 1$ in a ring R, then R consists solely of the element 0. Clearly, there is a ring with only one element, namely, the zero element. Such a ring is called the **trivial** or **zero ring**.

Before giving examples of rings we introduce the notion of a subring of a ring.

Definition
Let R be a ring. A **subring** R' of R is a ring which under addition is a subgroup of R and under multiplication is a submonoid of R. In particular, the identity 1 of R is contained in R'.

As an immediate consequence of this definition we have the following.

Basic Properties 1.2
Let R be a ring.

(a) If R' and R'' are two subrings of R, then $R' = R''$ if and only if their underlying sets are the same.
(b) A subset $X \subset R$ is the underlying set of a subring of R if and only if:
  (i) X is closed under addition and multiplication, that is, if x_1 and x_2 are in X, then $x_1 + x_2$ and x_1x_2 are in X.
  (ii) 0 and 1 are in X.
  (iii) If x is in X, then $-x$ is also in X.

In other words, to completely describe a subring R' of a ring R, it suffices to describe the underlying set of R'.

Example 1.3 The following are familiar examples of commutative rings:

(a) The set of integers $\mathbf{Z}$ under ordinary addition and multiplication.
(b) The set of all rational numbers $\mathbf{Q}$ under ordinary addition and multiplication.
(c) The set $\mathbf{R}$ of all real numbers under ordinary addition and multiplication.
(d) The set of all complex numbers $\mathbb{C}$ under ordinary addition and multiplication.

Also, it is not difficult to see that $\mathbf{Z}$ is a subring of $\mathbf{Q}$ which in turn is a subring of $\mathbf{R}$ which finally is a subring of $\mathbb{C}$.

Example 1.4 Let R be an arbitrary ring. Then it is not difficult to see that the 2×2 matrices over R (that is, the set of all arrays $\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$ with the r_{ij} in R) with

the usual addition and multiplication of matrices given by

$$\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} + \begin{pmatrix} r'_{11} & r'_{12} \\ r'_{21} & r'_{22} \end{pmatrix} = \begin{pmatrix} r_{11}+r'_{11} & r_{12}+r'_{12} \\ r_{21}+r'_{21} & r_{22}+r'_{22} \end{pmatrix}$$

and

$$\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}\begin{pmatrix} r'_{11} & r'_{12} \\ r'_{21} & r'_{22} \end{pmatrix} = \begin{pmatrix} r_{11}r'_{11}+r_{12}r'_{21} & r_{11}r'_{12}+r_{12}r'_{22} \\ r_{21}r'_{11}+r_{22}r'_{21} & r_{21}r'_{12}+r_{22}r'_{22} \end{pmatrix}$$

is a ring. This ring is called the ring of **two-by-two matrices** over R and is usually denoted by $M_2(R)$. The reader should not have great difficulty in showing that as long as $1 \neq 0$ in R, $M_2(R)$ is not a commutative ring.

Example 1.5 Let R be a ring. The subset $T_2(R)$ of $M_2(R)$ consisting of all elements $\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$ with $r_{12}=0$ is a subring of $M_2(R)$ called the ring of **2×2 lower triangular matrices** over R. As in Example 1.4, $T_2(R)$ is not commutative if $1 \neq 0$ in R.

Example 1.6 Let A be an abelian group and let $\mathrm{End}(A)$ be the set of all group morphisms $f: A \to A$. It is not hard to check that if f and g are in $\mathrm{End}(A)$, then the map $f+g: A \to A$ given by $(f+g)(a)=f(a)+g(a)$ for each a in A is again a morphism of groups called the sum of f and g. Further, the law of composition of $\mathrm{End}(A)$ given by $(f, g) \to f+g$ makes $\mathrm{End}(A)$ into an abelian group since (a) it is associative, (b) the zero morphism $0: A \to A$ defined $0(a)=0$ for all a in A is an identity, and (c) given a morphism $f: A \to A$, the map $(-f): A \to A$ given by $(-f)\times(a)=-(f(a))$ for all a in A is a morphism of groups with the property $f+(-f)=0$.

We already know that $\mathrm{End}(A)$ is a multiplicative monoid with id_A as identity under the law of composition given by the composition of morphisms.

Finally, it is not difficult to check that $\mathrm{End}(A)$ together with the addition and multiplication described above is a ring which is called the **ring of endomorphisms of the abelian group** A.

Having described rings, we must describe how to compare two rings. Because a ring is completely determined by its underlying set as well as its structure as an additive group and multiplicative monoid, it is clear that a morphism $f: R \to R'$ from the ring R to the ring R' should be a map from the underlying set of R to that of R' which is compatible with both the additive and multiplicative structures of R and R'. In other words, a morphism $f: R \to R'$ is a map of the underlying sets of R and R' which is at the same time a morphism of the additive group structures as well as the multiplicative monoid structures of R and R'. Stated symbolically we have the following.

Definition

A **morphism** $f: R \to R'$ from the ring R to the ring R' is a map f from the underlying set of R to the underlying set of R' satisfying, for all r_1 and r_2 in R:

(a) $f(r_1+r_2)=f(r_1)+f(r_2)$.
(b) $f(r_1r_2)=f(r_1)f(r_2)$.
(c) $f(1)=1$.

We now list some easily verified properties of morphisms of rings.

Basic Properties 1.7
Let R, R', R'', and R''' be rings.

(a) The identity map on the underlying set of the ring R is a morphism of rings which we denote by id_R.
(b) If $f: R \to R'$ and $g: R' \to R''$ are morphisms of rings, then the composition $gf: R \to R''$ of the maps f and g is a morphism of rings which we denote by gf and call the **composition** of the morphisms f and g.
(c) If $f: R \to R'$, $g: R' \to R''$, and $h: R'' \to R'''$ are morphisms of rings, then the morphisms $(hg)f: R \to R'''$ and $h(gf): R \to R'''$ are the same, that is, the composition of morphisms of rings is associative.

Our discussion shows that we can define a category called the category of rings by the following data:

(a) The objects of this category are all rings.
(b) For any two rings R and R', the set (R, R') is the set of all ring morphisms from R to R'.
(c) For each triple of rings R, R', and R'', the composition of morphisms is given by the map of sets

$$(R, R') \times (R', R'') \to (R, R'')$$

defined by $(f, g) \to gf$, the composition of ring morphisms.

The category given by these data is denoted by Rings.

Because we have defined isomorphisms, epimorphisms, and monomorphisms in arbitrary categories, we have these notions for the category Rings. Also, the fact that the objects of the category Rings have underlying sets makes it fairly clear what we mean by a morphism of rings being surjective, injective, or bijective. Either as immediate consequences of results already developed for morphisms of monoids and groups or as consequences of easy direct calculations we have the following relations between the various types of morphisms between rings.

Basic Properties 1.8
Let $f: R \to R'$ be a morphism of rings.

(a) If f is a surjective (injective) morphism, then f is an epimorphism (monomorphism).
(b) f is an isomorphism if and only if f is a bijective morphism.

These results naturally raise the question of whether a ring morphism $f: R \to R'$ which is a monomorphism (epimorphism) is necessarily an injective morphism (surjective morphism). We have already seen that monomorphisms in the categories Monoid or Groups are injective morphisms. Also, although epimorphisms in the category Group are surjective morphisms, the same is not true in the category Monoid. The situation for the category Rings is the same as that for the category Monoid. Namely, all monomorphisms in the category of Rings are injec-

tive morphisms, whereas not all epimorphisms are surjective morphisms. The next section is devoted to showing that monomorphisms of rings are injective morphisms. The fact that not all epimorphisms of rings are surjective morphisms will be shown later on in this chapter. In the meantime we end this section by pointing out the following easily verified analogs of results already established for monoids and groups.

Basic Properties 1.9
Let R and S be rings.

(a) If R is a subring of S, then the inclusion map of sets $R \to S$ is an injective morphism of rings called the **inclusion morphism** and written $\text{inc}: R \to S$.

Now suppose $f: R \to S$ is an arbitrary morphism of rings. Then:

(b) $\text{Im } f$ is a subring of S called the **image** of f.
(c) The map $f_0: R \to \text{Im } f$ is a surjective morphism of rings.
(d) The morphism $f: R \to S$ is the composition of the morphisms of rings

$$R \xrightarrow{f_0} \text{Im } f \xrightarrow{\text{inc}} S$$

Finally, for ease of reference we make the following definition.

Definition
Suppose $f: S \to T$ is a morphism of rings. For each subring R of S, the composition $R \xrightarrow{\text{inc}} S \xrightarrow{f} T$ is called the **restriction** of f to R and is denoted by $f|R$.

2. POLYNOMIAL RINGS

We recall that the proof that monomorphisms in the category Monoid are injective morphisms was based on the fact that the monoid $\mathbf{N}$ of nonnegative integers under addition has the following property: If M is an arbitrary monoid, then the map of sets $(\mathbf{N}, M) \to M$, given by $f \to f(1)$ for each morphism of monoids $f: \mathbf{N} \to M$, is an isomorphism of sets. Similarly, the proof that monomorphisms in the category Groups are injective morphisms was based on the fact that the group $\mathbf{Z}$ of all integers under addition has the property that for each group G, the map of sets $(\mathbf{Z}, G) \to G$, given by $f \to f(1)$ for each morphism of groups $f: \mathbf{Z} \to G$, is an isomorphism of sets. Our proof that monomorphisms in the category Rings are injective morphisms follows this pattern. That is, it is based on the fact that there is a ring S containing an element x having the following property: For each ring R', the map of sets $(S, R') \to R'$, given by $f \to f(x)$ for each morphism of rings $f: S \to R'$, is an isomorphism of sets. After discussing polynomial rings over commutative rings, we will see that the ring of polynomials over the integers has the property just described for the ring S.

In order to discuss rings of polynomials, it is convenient to have the following notational device.

Let $(r_i)_{i \in I}$ be a family of elements in a commutative monoid R which we write additively. The subset J of I consisting of all i in I such that $r_i \neq 0$ is called the

support of the family $(r_i)_{i\in I}$. The family $(r_i)_{i\in I}$ of elements of R is said to be **almost zero** if its support is a finite set. If $(r_i)_{i\in I}$ is an almost zero family of elements in R with support J, we denote the finite sum $\Sigma_{i\in I} r_i$ by $\Sigma_{i\in I} r_i$. Clearly, if $(r_i)_{i\in I}$ and $(r'_i)_{i\in I}$ are two almost zero families of elements in R, then $(r_i+r'_i)_{i\in I}$ is an almost zero family of elements in R and $\Sigma_{i\in I}(r_i+r'_i)=\Sigma_{i\in I} r_i+\Sigma_{i\in I} r'_i$. Finally, if R is a ring and $(r_i)_{i\in I}$ and $(x_i)_{i\in I}$ are two families of elements with $(r_i)_{i\in I}$ an almost zero family, then $(r_ix_i)_{i\in I}$ is also an almost zero family of elements in R.

Suppose R is a commutative ring. We denote by $R[\mathbf{N}]$ the set of all almost zero families $(r_n)_{n\in\mathbf{N}}$ of elements in R. It is left to the reader to check that the following maps from $R[\mathbf{N}]\times R[\mathbf{N}]$ to $R[\mathbf{N}]$ are laws of composition on $R[\mathbf{N}]$ which make $R[\mathbf{N}]$ a commutative ring. The addition law, $\text{add}: R[\mathbf{N}]\times R[\mathbf{N}]\to R[\mathbf{N}]$, is defined by $\text{add}((r_n)_{n\in\mathbf{N}}, (r'_n)_{n\in\mathbf{N}})=(r_n+r'_n)_{n\in\mathbf{N}}$ and the multiplication law, $\text{mult}: R[\mathbf{N}]\times R[\mathbf{N}]\to R[\mathbf{N}]$, is defined by $\text{mult}((r_n)_{n\in\mathbf{N}}, (r'_n)_{n\in\mathbf{N}})=(\Sigma_{j=0}^{n} r_j r'_{n-j})_{n\in\mathbf{N}}$. The zero element of the ring $R[\mathbf{N}]$ is the element $(r_n)_{n\in\mathbf{N}}$ with $r_n=0$ for all n in $\mathbf{N}$. The 1 of the ring $R[\mathbf{N}]$ is the element $(r_n)_{n\in\mathbf{N}}$ with $r_n=0$ for $n\neq 0$ and $r_0=1$.

We observe that the map $h: R\to R[\mathbf{N}]$ given by $h(r)$ is the element $(r_n)_{n\in\mathbf{N}}$ with $r_0=r$ and $r_n=0$ for $n\neq 0$, is an injective morphism of rings whose image consists precisely of the elements $(r_n)_{n\in\mathbf{N}}$ in $R[\mathbf{N}]$ with the property $r_n=0$ if $n\neq 0$. Therefore, identifying the subring of $R[\mathbf{N}]$ consisting of all $(r_n)_{n\in\mathbf{N}}$ with $r_n=0$ if $n\neq 0$ with R by means of the injective morphism $h: R\to R[\mathbf{N}]$, we have that R is a subring of $R[\mathbf{N}]$. The reader should check that as a result of this identification of R with a subring of $R[\mathbf{N}]$ we have the following rules of calculation: For r in R and $(r_n)_{n\in\mathbf{N}}$ in $R[\mathbf{N}]$ we have (**a**) $r(r_n)_{n\in\mathbf{N}}=(rr_n)_{n\in\mathbf{N}}$; (**b**) $(r_n)_{n\in\mathbf{N}}r=(r_nr)_{n\in\mathbf{N}}$; and (**c**) $r+(r_n)_{n\in\mathbf{N}}=(r'_n)_{n\in\mathbf{N}}$, where $r'_0=r+r_0$ and $r'_n=r_n$ for $n>0$.

Before we can go further in our analysis of the ring $R[\mathbf{N}]$, we need some notation. For each pair (m, n) in $\mathbf{N}\times\mathbf{N}$ we introduce the symbol $\delta_{m,n}$ which stands for the 0 of R if $m\neq n$ and for the 1 of R if $m=n$. If we let $X=(\delta_{1,n})_{n\in\mathbf{N}}$, then it is not difficult to establish:

(**a**) $X^i=(\delta_{i,n})_{n\in\mathbf{N}}$, for all i in $\mathbf{N}$.
(**b**) $rX^i=(r\delta_{i,n})_{n\in\mathbf{N}}$, for all i in $\mathbf{N}$.

From these observations it follows that for each element $(r_n)_{n\in\mathbf{N}}$ in $R[\mathbf{N}]$ we have $(r_n)_{n\in\mathbf{N}}=\Sigma_{n\in\mathbf{N}} r_nX^n$.

In practice the ring $R[\mathbf{N}]$ is generally denoted by $R[X]$, and the representation $\Sigma_{n\in\mathbf{N}} r_nX^n$ is used for the elements of $R[\mathbf{N}]$ rather than the representation $(r_n)_{n\in\mathbf{N}}$. Because of this, we recapitulate what we have already established about the ring $R[X]$ using the notation $\Sigma_{n\in\mathbf{N}} r_nX^n$ for the elements of $R[X]$.

Definition

Let R be a commutative ring. The ring $R[X]$ is called the **ring of polynomials over** R. The elements $\Sigma_{n\in\mathbf{N}} r_nX^n$ in $R[X]$ for each almost zero family $(r_n)_{n\in\mathbf{N}}$ of elements in R are called the **polynomials** over R.

Basic Properties 2.1

Let R be a commutative ring and $R[X]$ the ring of polynomials over R.

(**a**) Two elements $\Sigma_{n\in\mathbf{N}} r_nX^n$ and $\Sigma_{n\in\mathbf{N}} r'_nX^n$ in $R[X]$ are the same if and only if $r_n=r'_n$ for all n in $\mathbf{N}$.

(b) $\Sigma_{n\in\mathbf{N}} r_nX^n + \Sigma_{n\in\mathbf{N}} r'_nX^n = \Sigma_{n\in\mathbf{N}} (r_n + r'_n)X^n$.
(c) $(\Sigma_{n\in\mathbf{N}} r_nX^n)(\Sigma_{n\in\mathbf{N}} r'_nX^n) = \Sigma_{n\in\mathbf{N}}(\Sigma_{j=0}^{n} r_j r'_{n-j})X^n$.
(d) $R[X]$ is a commutative ring.
(e) The elements $\Sigma_{n\in\mathbf{N}} r_nX^n$ with $r_n = 0$ if $n \neq 0$ constitute a subring of $R[X]$ which is isomorphic to R by means of the ring morphism $r \to \Sigma_{n\in\mathbf{N}} r_nX^n$ for all r in R where $r_n = r$ if $n = 0$ and $r_n = 0$ if $n > 0$. This isomorphism is usually viewed as an identification which means that we often write simply r for the element $\Sigma_{n\in\mathbf{N}} r_nX^n$ where $r_0 = r$ and $r_n = 0$ if $n > 0$.

In order to state a fundamental property of polynomial rings which will be used in showing that ring monomorphisms are injective, we make the following observation. If S is a ring, the set of all elements x in S such that $xs = sx$ for all s in S is a commutative subring of S.

Definition
Let S be a ring. The subring of S consisting of all x in S such that $xs = sx$ for all s in S is called the **center of** S. We will denote the center of S by $C(S)$.

It is clear that a ring S is commutative if and only if $C(S) = S$.

Proposition 2.2
Let R be a commutative ring, S an arbitrary ring, and $f: R \to S$ a ring morphism such that $\operatorname{Im} f \subset C(S)$. Then for each x in S, there is a unique ring morphism $f_x : R[X] \to S$ such that $f_x|R = f$ and $f_x(X) = x$. This ring morphism $f_x : R[X] \to S$ is given by $f_x(\Sigma_{n\in\mathbf{N}} r_nX^n) = \Sigma_{n\in\mathbf{N}} f(r_n)x^n$.

PROOF: Since each element of $R[X]$ can be written uniquely as $\Sigma\, r_nX^n$, we obtain a map $f_x : R[X] \to S$ by setting $f_x(\Sigma_{n\in\mathbf{N}} r_nX^n) = \Sigma_{n\in\mathbf{N}} f(r_n)x^n$. Obviously, $f_x|R = f$ and $f_x(X) = x$. That $f_x : R[X] \to S$ is a ring morphism follows from the fact that the element x in S commutes with each element of $\operatorname{Im} f$ since $\operatorname{Im} f$ is in the center of S. The verification of this, as well as the uniqueness of the ring morphism f_x, is left to the reader.

We now wish to apply this general result to find the ring morphisms from $\mathbf{Z}[X]$ to an arbitrary ring S where $\mathbf{Z}$ is the ring of integers. To do this we first must determine the ring morphisms from $\mathbf{Z}$ to S.

Let S be an arbitrary ring. Now viewing $\mathbf{Z}$ and S as abelian groups, we know by the results of Chapter 2 that given any s in S there is one and only one morphism $f: \mathbf{Z} \to S$ of abelian groups such that $f(1) = s$, namely, the morphism given by $f(z) = zs$ for all z in $\mathbf{Z}$. Since any morphism of rings $f: \mathbf{Z} \to S$ is also a morphism of the additive groups of $\mathbf{Z}$ to S which must have the property $f(1) = 1$, it follows there is at most one morphism of rings from $\mathbf{Z}$ to S. That there is a morphism of rings from $\mathbf{Z}$ to S follows from the fact that the morphism of additive groups $f: \mathbf{Z} \to S$ given by $f(z) = z1$ is also a ring morphism. To see this, we observe that since $f: \mathbf{Z} \to S$ already has the properties $f(z_1 + z_2) = f(z_1) + f(z_2)$ and $f(1) = 1$, we only have to show that $f(z_1z_2) = f(z_1)f(z_2)$. But $f(z_1z_2) = (z_1z_2)1$ and $f(z_1)f(z_2) = (z_11)(z_21)$. Hence, we must show that $(z_1z_2)(1) = (z_11)(z_21)$ for all z_1 and z_2 in $\mathbf{Z}$. We have already seen in Chapter 2 that $(z_1z_2)(1) = z_1(z_21)$. Hence, in order to show that the map $f: \mathbf{Z} \to S$ is a morphism of rings we have to show that

$z_1(z_2 1) = (z_1 1)(z_2 1)$ for all z_1 and z_2 in $\mathbf{Z}$. This follows from the fact that for each z_2 in $\mathbf{Z}$, the maps $g : \mathbf{Z} \to S$ and $h : \mathbf{Z} \to S$ given by $g(z) = z(z_2 1)$ and $h(z) = (z 1)(z_2 1)$ are the same because they are both easily seen to be group morphisms from the additive group of $\mathbf{Z}$ to that of S satisfying $g(1) = z_2 1 = 1(z_2 1) = h(1)$. This shows that the map $f : \mathbf{Z} \to S$ given by $f(z) = z 1$ for all z in $\mathbf{Z}$ is the unique morphism of rings from $\mathbf{Z}$ to S. This proves the following.

Proposition 2.3
Given an arbitrary ring S, the map $u_S : \mathbf{Z} \to S$ given by $u_S(z) = z 1$ is the unique ring morphism from $\mathbf{Z}$ to S.

We now show that for each ring S, the image of the unique ring morphism $u_S : \mathbf{Z} \to S$ is a subring of the center of S. Suppose s is an element of S. We want to show that $s(z1) = (z1)s$ for all z in $\mathbf{Z}$. Consider the map $g : \mathbf{Z} \to S$ given by $g(z) = (z1)s$ for all z in $\mathbf{Z}$. Then $g(z_1 + z_2) = ((z_1 + z_2)1)s = (z_1 1)s + (z_2 1)s = g(z_1) + g(z_2)$. Hence, $g : \mathbf{Z} \to S$ is a morphism of abelian groups with $g(1) = s$. Similarly, the map $h : \mathbf{Z} \to S$ given by $h(z) = s(z1)$ for all z in $\mathbf{Z}$ is also a morphism of abelian groups with $h(1) = s$. Since $g(1) = h(1)$, we conclude that $g = h$. Hence, $(z1)s = s(z1)$ for all s in S and z in $\mathbf{Z}$. Therefore, Im $u_S \subset C(S)$.

Because for each ring S the image of $u_S : \mathbf{Z} \to S$ is an important subring of S, we make the following definition.

Definition
Let S be a ring and $u_S : \mathbf{Z} \to S$ the unique morphism of rings. The subring Im u_S of $C(S)$ is called the **primitive subring** of S.

As a consequence of the foregoing results, we now have the following.

Proposition 2.4
For each ring S, the map of sets $(\mathbf{Z}[X], S) \to S$ given by $f \to f(X)$ for all ring morphisms $f : Z[X] \to S$, is an isomorphism of sets.

PROOF: We first show that the map $(\mathbf{Z}[X], S) \to S$ is surjective. Suppose t is an element of S. Let $f : \mathbf{Z} \to S$ be the ring morphism $u_S : \mathbf{Z} \to S$. Since Im f is contained in $C(S)$, we know by Proposition 2.2 that there is a unique morphism $f_t : \mathbf{Z}[X] \to S$ such that $f_t|\mathbf{Z} = f$ and $f_t(X) = t$. Hence, the map $(\mathbf{Z}[X], S) \to S$ is surjective.

Suppose now $f, g : \mathbf{Z}[X] \to S$ are ring morphisms such that $f(X) = g(X)$. Since there is only one ring morphism $u_S : \mathbf{Z} \to S$ we know that $f|\mathbf{Z} = g|\mathbf{Z}$. Hence, again by Proposition 2.2, the fact that $f(X) = g(X)$ and $f|\mathbf{Z} = g|\mathbf{Z}$ implies that $f = g$. Therefore, the map $(\mathbf{Z}[X], S) \to S$ is injective as well as surjective and hence is an isomorphism of sets.

Finally, we use this proposition to establish the following result which supplied the motivation for this entire section.

Proposition 2.5
In the category Rings every monomorphism of rings $f : S \to T$ is an injective morphism.

PROOF: Suppose $f: S \to T$ is a monomorphism. We want to show that if s_1 and s_2 are elements of S such that $f(s_1) = f(s_2)$, then $s_1 = s_2$. By our previous result we know that there are morphisms of rings $f_1, f_2: \mathbf{Z}[X] \to S$ such that $f_1(X) = s_1$ and $f_2(X) = s_2$. Since $f(s_1) = f(s_2)$, the compositions $\mathbf{Z}[X] \xrightarrow{ff_1} T$ and $\mathbf{Z}[X] \xrightarrow{ff_2} T$ have the property $ff_1(X) = ff_2(X)$ because $ff_1(X) = f(s_1) = f(s_2) = ff_2(X)$. This implies that $ff_1 = ff_2$, because our previous result showed that morphisms from $\mathbf{Z}[X]$ to T are completely determined by their values on X. Because $f: S \to T$ is a monomorphism, the fact that $ff_1 = ff_2$ implies that $f_1 = f_2$. This in turn implies $s_1 = f_1(X) = f_2(X) = s_2$. Thus, we have our desired result that if $f: S \to T$ is a monomorphism, then f is an injective morphism.

3. ANALYSES OF RING MORPHISMS

In this section we continue our discussion of some of the general properties of rings that are direct analogs of results already considered for monoids and groups. For instance, analyses of morphisms of rings, partitions of rings, and various isomorphism theorems will be discussed. Because most of the proofs for these results can be obtained by direct application of results already obtained for monoids and groups, few proofs will actually be given. Those that are given will be mainly for the purpose of illustrating how the appropriate results for monoids and groups can be applied to rings. It is hoped that the reader will find it a useful exercise to supply the missing proofs.

We have already seen that if $f: R \to S$ is a morphism of rings, then $\operatorname{Im} f$ is a subring of S and $f: R \to S$ is the composition of the morphisms $R \xrightarrow{f_0} \operatorname{Im} f \xrightarrow{\text{inc}} S$ where $f_0: R \to \operatorname{Im} f$ is a surjective morphism and $\text{inc}: \operatorname{Im} f \to S$ is an injective morphism. In analogy with the situation for monoids we make the following definition.

Definition
Let $f: R \to S$ be a morphism. Then the factorization

$$R \xrightarrow{f_0} \operatorname{Im} f \xrightarrow{\text{inc}} S$$

of f is called the **image analysis of** f.

More generally, any factorization

$$R \xrightarrow{g} R' \xrightarrow{h} S$$

of f with g a surjective morphism of rings and h an injective morphism of rings is called an **analysis of** f.

On the basis of our experience with monoids and groups one should expect that any two analyses of a morphism of rings are essentially the same. That this is indeed the case will follow from the following general considerations.

Proposition 3.1
Suppose R, S, and T are rings and $f: R \to S$ and $g: S \to T$ are maps of the underlying sets of the rings involved such that the map $gf: R \to T$ is a morphism of rings.

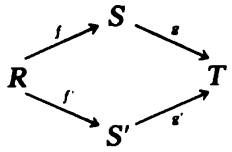
(a) If $f: R \to S$ is a surjective morphism of rings, then $g: S \to T$ is also a morphism of rings.
(b) If $g: S \to T$ is an injective morphism of rings, then $f: R \to S$ is also a morphism of rings.

PROOF: (a) This can be proved directly from first principles or else derived from the analogous result for monoids. We will take the second path. Because the composition $gf: R \to T$ is a morphism of rings, it is certainly a morphism of the additive group of R to the additive group of T. Similarly, $f: R \to S$ is a surjective morphism of the additive group of R to that of S. Hence, by our previous results concerning monoids, we know that the map $g: S \to T$ is also a morphism of the additive group of S to the additive group of T. A similar argument also shows that $g: S \to T$ is a morphism of the multiplicative monoid of S to that of T. Therefore, the map $g: S \to T$ is a morphism of rings because it is both a morphism of the additive groups of S and T and a morphism of the multiplicative monoids of S and T.

(b) This can be established in a manner similar to part (a) and is left as an exercise.

As for monoids and groups, we have, as a direct consequence of this result, the following.

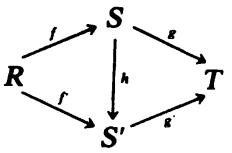
Proposition 3.2
Suppose we are given a commutative diagram of morphisms of rings



satisfying:

(a) f is a surjective morphism.
(b) g' is an injective morphism.

Then there is one and only one morphism of rings $h: S \to S'$ such that the diagram



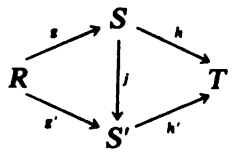
commutes.

By way of application of this result we show in what sense any two analyses of a morphism of rings are the same.

Proposition 3.3
If

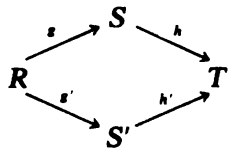
$$R \xrightarrow{g} S \xrightarrow{h} T$$
$$R \xrightarrow{g'} S' \xrightarrow{h'} T$$

are analyses of the same morphism of rings $f: R \to T$, then there is a unique morphism of rings $j: S \to S'$ such that the diagram

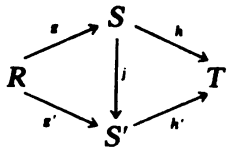


commutes. This uniquely determined morphism $j: S \to S'$ is an isomorphism of rings.

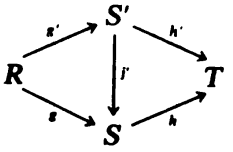
PROOF: Because $R \xrightarrow{g} S \xrightarrow{h} T$ and $R \xrightarrow{g'} S' \xrightarrow{h'} T$ are analyses of the same ring morphism $f: R \to T$, it follows that



is a commutative diagram satisfying (a) g and g' are surjective morphisms and (b) h and h' are injective morphisms. Hence, by our previous result there are unique ring morphisms $j: S \to S'$ and $j': S' \to S$ such that the diagrams

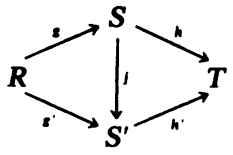


and

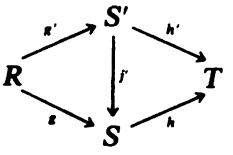


commute. If we show that this uniquely determined morphism $j: S \to S'$ is an isomorphism, we will have established our desired result.

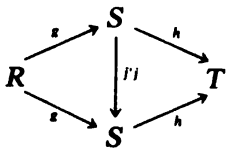
It follows easily from the commutativity of the diagrams



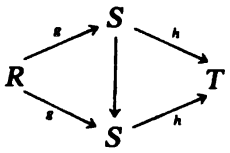
and



that the diagram



commutes. But by our previous proposition we know that there is only one morphism $S \to S$ which makes the diagram



commute because g is a surjective morphism and h is an injective morphism. Therefore, the fact that the identity morphism $\mathrm{id}_S: S \to S$ in addition to $j'j: S \to S$ has this property implies that $j'j = \mathrm{id}_S$. A similar argument shows that $jj' = \mathrm{id}_{S'}$. Therefore, we have established that $j: S \to S'$ is an isomorphism of rings, which completes the proof of the proposition.

Having introduced the general notion of an analysis of a morphism of rings, we now discuss the analog for ring morphisms of the coimage analyses of a morphism of monoids or groups.

Suppose we are given a morphism of rings $f: R \to S$. Then viewed as a map of sets, f has the coimage analysis $R \xrightarrow{k_{\mathrm{Coim}\, f}} \mathrm{Coim}\, f \xrightarrow{j_f} S$. Since $f: R \to S$ is also a morphism from the additive group of R to the additive group of S, we know that $\mathrm{Coim}\, f$ has a unique structure as a commutative group such that the maps $k_{\mathrm{Coim}\, f}: R \to \mathrm{Coim}\, f$ and $j_f: \mathrm{Coim}\, f \to S$ are morphisms of groups.

Because $f: R \to S$ is also a morphism from the multiplicative monoid of R to that of S, $\mathrm{Coim}\, f$ has a unique monoid structure such that the maps $k_{\mathrm{Coim}\, f}: R \to \mathrm{Coim}\, f$ and $j_f: \mathrm{Coim}\, f \to S$ are also morphisms of monoids. Thus, there are uniquely determined laws of composition + and × on $\mathrm{Coim}\, f$ such that $\mathrm{Coim}\, f$ under + is a commutative group and under × is a monoid such that the maps $k_{\mathrm{Coim}\, f}: R \to \mathrm{Coim}\, f$ and $j_f: \mathrm{Coim}\, f \to S$ are simultaneously morphisms for both the additive and multiplicative structures on R, $\mathrm{Coim}\, f$, and S. Hence, if we show that $\mathrm{Coim}\, f$ with these laws of composition is a ring, then we will have that the maps $k_{\mathrm{Coim}\, f}: R \to \mathrm{Coim}\, f$ and $j_f: \mathrm{Coim}\, f \to S$ are morphisms of rings with the properties (**a**) $k_{\mathrm{Coim}\, f}: R \to \mathrm{Coim}\, f$ is a surjective morphism of rings; (**b**) $j_f: \mathrm{Coim}\, f \to S$ is an injective morphism of rings; and (**c**) $f = j_f k_{\mathrm{Coim}\, f}$. The fact that $\mathrm{Coim}\, f$ with + and × as defined above is actually a ring follows from the following general property.

Basic Properties 3.4

Suppose $f: R \to X$ is a surjective map of sets with R a ring.

(**a**) If + and × are two maps from $X \times X \to X$ such that

$$f(r_1 + r_2) = f(r_1) + f(r_2)$$

and

$$f(r_1r_2) = f(r_1) \times f(r_2)$$

for all pairs of elements r_1 and r_2 in R, then X together with $+$ and $\times$ is a ring such that $f: R \to X$ is a surjective morphism of rings.

(b) X has at most one ring structure such that the surjective map of sets $f: R \to X$ is a morphism of rings.

PROOF: Left as an exercise.

Summarizing, we have the following.

Proposition 3.5

Let $f: R \to S$ be a morphism of rings. Then the set Coim f has a unique structure as a ring such that the maps $k_{\text{Coim}\, f}: R \to \text{Coim}\, f$ and $j_f: \text{Coim}\, f \to S$ are ring morphisms. This uniquely determined ring structure on Coim f is given by

(a) $[r_1] + [r_2] = [r_1 + r_2]$
(b) $[r_1][r_2] = [r_1r_2]$

for all elements r_1 and r_2 in R where $[r]$ stands for the unique element of Coim f containing the element r in R. This naturally suggests the following.

Definitions

Let $f: R \to S$ be a morphism of rings. The ring consisting of the set Coim f with the laws of composition given by

$$[r_1] + [r_2] = [r_1 + r_2]$$

and

$$[r_1][r_2] = [r_1r_2]$$

for all r_1 and r_2 in R, is called the **coimage of f** and is denoted by Coim f. Moreover, the factorization

$$R \xrightarrow{k_{\text{Coim}\, f}} \text{Coim}\, f \xrightarrow{j_f} S$$

of f into the surjective morphism of rings $k_{\text{Coim}\, f}: R \to \text{Coim}\, f$ and the injective morphism of rings $j_f: \text{Coim}\, f \to S$ is called the **coimage analysis of f.**

As with monoids and groups, one of the most important consequences of the existence of the coimage analysis of a morphism is that it enables us to describe all surjective ring morphisms $f: R \to S$ for a fixed ring R essentially in terms of the ring R itself. This observation is based on the easily verified fact that a morphism of rings $f: R \to S$ is a surjective morphism if and only if the injective morphism $j_f: \text{Coim}\, f \to S$ is surjective and hence an isomorphism. Thus, we see that a surjective morphism of rings $f: R \to S$ is essentially the same as the surjective morphism $k_{\text{Coim}\, f}: R \to \text{Coim}\, f$, because there is a unique isomorphism Coim $f \to S$ (why unique?), namely, the morphism $j_f: \text{Coim}\, f \to S$, such that $f = j_f k_{\text{Coim}\, f}$. Therefore, if we can determine which partitions of the ring R come from morphisms of rings $f: R \to S$, we will have essentially described all surjective morphisms with domain R.

4. IDEALS

Suppose $f: R \to S$ is a morphism of rings. Then the coimage analysis $R \xrightarrow{k_{\text{Coim}\, f}} \text{Coim}\, f \xrightarrow{i_f} S$ of this morphism of rings is also the coimage analysis of $f: R \to S$ viewed just as a morphism from the additive group of R to that of S. Hence, if we let I be the subgroup $f^{-1}(0)$ of the additive group of R, we see that, as an additive group, $\text{Coim}\, f = R/I$ and the map $k_{\text{Coim}\, f}: R \to \text{Coim}\, f$ is the map $k_{R/I}: R \to R/I$ given by the canonical morphisms of the group R to its factor group R/I.

Now the fact that $f: R \to S$ is also a morphism of rings shows that if x is in I, then rx and xr are also in I for all r in R since $f(rx) = f(r)f(x) = 0 = f(x)f(r) = f(xr)$. Hence, the subgroup $I = f^{-1}(0)$ also satisfies the condition $rI \subset I$ and $Ir \subset I$ for all r in R. Finally, we point out that the multiplication in $\text{Coim}\, f = R/I$ is given by $(r_1 + I)(r_2 + I) = r_1 r_2 + I$ for all r_1 and r_2 in R. Summarizing, we have the following.

Basic Properties 4.1
Let $f: R \to S$ be a morphism of rings.

(a) $I = f^{-1}(0)$ is a subgroup of the additive group of R satisfying the conditions $rI \subset I$ and $Ir \subset I$ for all r in R.
(b) As an additive group, $\text{Coim}\, f = R/I$, while the multiplication in the ring Coim f is given by

$$(r_1 + I)(r_2 + I) = (r_1 r_2 + I)$$

for all r_1 and r_2 in R.
(c) The morphism of rings $k_{\text{Coim}\, f}: R \to \text{Coim}\, f$ is the same as the map $k_{R/I}: R \to R/I$ where $k_{R/I}$ is the canonical morphism from the additive group of R to its factor group R/I.

Moreover, it is not difficult to check the following additional property.

Basic Properties 4.2
Let R be a ring and I a subgroup of R satisfying the conditions $rI \subset I$ and $Ir \subset I$ for all r in R. Then:

(a) The partition R/I of the additive group of R is also a partition of the multiplicative monoid of R, or, what is the same thing, $(r_1 + I)(r_2 + I) \subset r_1 r_2 + I$ for all r_1 and r_2 in R.
(b) The abelian group R/I together with the multiplication given by $(r_1 + I) \times (r_2 + I) = r_1 r_2 + I$ is a ring since the canonical surjective morphism of groups $k_{R/I}: R \to R/I$ has the properties
  (i) $k_{R/I}(r_1 + r_2) = k_{R/I}(r_1) + k_{R/I}(r_2)$ and
  (ii) $k_{R/I}(r_1 r_2) = k_{R/I}(r_1) k_{R/I}(r_2)$
  for all r_1 and r_2 in R.
(c) This ring structure on R/I is the unique ring structure which makes the canonical surjective map

$$k_{R/I}: R \to R/I$$

given by $k_{R/I}(r) = r + I$ for all r in R a morphism of rings.

(d) Finally, the morphism of rings $k_{R/I}: R \to R/I$ has the property: $k_{R/I}^{-1}(0) = I$.

These results suggest the following.

Definitions
Let R be a ring.

(a) A subgroup I of the additive group of R is said to be an **ideal of R** if $rI \subset I$ and $Ir \subset I$ for all r in R.
(b) If I is an ideal in R, then we denote by R/I the ring which, as an additive group, is the group R/I and whose multiplication is given by

$$(r_1 + I)(r_2 + I) = (r_1 r_2 + I)$$

for all r_1 and r_2 in R. The ring R/I is called the **factor ring** of R by the ideal I.
(c) The map $k_{R/I}: R \to R/I$ given by $k_{R/I}(r) = r + I$ for all r in R is a surjective morphism of rings which we call the **canonical morphism from R to R/I.**
(d) If $f: R \to S$ is a morphism of rings, then the ideal $f^{-1}(0)$ of R is called the **kernel of f** and is often denoted by Ker f.

The reader should have no difficulty in establishing the following properties of ideals and factor rings which are exact analogs of what has already been established for normal subgroups and factor groups of groups.

Proposition 4.3
Suppose $f: R \to S$ is a morphism of rings with $I = \text{Ker } f$.

(a) For each subset $r + I$ of R in R/I, the subset $f(r + I)$ of S consists of the single element $f(r)$. Thus, we obtain a map $j_f: R/I \to S$ given by $j_f(r + I) = f(r)$ for all r in R, which is an injective morphism of rings.
(b) The composition

$$R \xrightarrow{k_{R/I}} R/I \xrightarrow{j_f} S$$

is nothing more than the coimage analysis of the morphism f. Thus:
(c) $f: R \to S$ is:
  (i) injective if and only if Ker $f = 0$;
  (ii) surjective if and only if $j_f: R/I \to S$ is an isomorphism;
  (iii) an isomorphism if and only if Ker $f = 0$ and f is surjective.
(d) If $f: R \to S$ is a surjective morphism and $g: R \to T$ is an arbitrary morphism of rings, then there exists a morphism $h: S \to T$ of rings which makes the diagram

$$\begin{array}{ccc} R & \xrightarrow{f} & S \\ \downarrow{\scriptstyle id_R} & & \downarrow{\scriptstyle h} \\ R & \xrightarrow[g]{} & T \end{array}$$

commute if and only if Ker $f \subset$ Ker g. Moreover, if Ker $f \subset$ Ker g, then there is only one morphism $h: S \to T$ of rings such that $hf = g$.
(e) If $f: R \to S$ and $g: R \to T$ are two surjective morphisms of rings, then Ker $f =$ Ker g if and only if there is an isomorphism of rings $h: S \to T$ such that $hf = g$.

(f) If J is an ideal of S, then $f^{-1}(J)$ is an ideal of R containing I. Further, $f^{-1}(J)$ is the kernel of the composition of morphisms

$$R \xrightarrow{f} S \xrightarrow{k_{S/J}} S/J$$

As is the case with groups, we obtain much more detailed information when dealing with surjective morphisms instead of arbitrary morphisms of rings. This point is made explicit in the following.

Proposition 4.4
Suppose $f: R \to S$ is a surjective morphism of rings with kernel I.

(a) If I_1 is an ideal of R, then $f(I_1)$ is an ideal of S.
(b) If I_1 is an ideal of R, then $f^{-1}(f(I_1)) = I_1 + I$.
(c) If I_1 and I_2 are two ideals of R, then $f(I_1) = f(I_2)$ if and only if the ideals $f^{-1}(f(I_1))$ and $f^{-1}(f(I_2))$, both of which contain I, are the same.
(d) Hence, if we denote the set of all ideals of R containing I by $\mathcal{U}$ and the set of all ideals of S by $\mathcal{J}$, then the maps $\mathcal{U} \to \mathcal{J}$ and $\mathcal{J} \to \mathcal{U}$ given by $I_1 \to f(I_1)$ and $J \to f^{-1}(J)$, respectively, are isomorphisms of sets which are inverses of each other.
(e) For each ideal J of S, there is a unique morphism of rings $h: R/f^{-1}(J) \to S/J$ which makes the diagram

$$\begin{array}{ccc} R & \xrightarrow{f} & S \\ \downarrow{\scriptstyle k_{R/f^{-1}(J)}} & & \downarrow{\scriptstyle k_{S/J}} \\ R/f^{-1}(J) & \xrightarrow{h} & S/J \end{array}$$

commute, and this unique morphism $h: R/f^{-1}(J) \to S/J$ is an isomorphism.

Specializing these results to the case when R is a ring, I an ideal in R, and the morphism of rings is the canonical surjective morphism $k_{R/I}: R \to R/I$, we obtain the following.

Corollary 4.5
Let I be an ideal in the ring R and $k_{R/I}: R \to R/I$ the canonical surjective morphism. Suppose I_1 is an ideal of R.

(a) $k_{R/I}(I_1)$ is the ideal, $(I + I_1)/I$ of R/I. Moreover:
(b) The ideal $I + I_1$ of R is the kernel of the composition $R \to R/I \to R/I/(I + I_1)/I$. Hence:
(c) There is a unique ring morphism $h: R/I + I_1 \to R/I/(I + I_1)/I$ which makes the diagram

$$\begin{array}{ccc} R & \xrightarrow{k_{R/I}} & R/I \\ \downarrow{\scriptstyle k_{R/(I+I_1)}} & & \downarrow{\scriptstyle k_{R/I/(I+I_1)/I}} \\ R/(I+I_1) & \xrightarrow{h} & R/I/(I+I_1)/I \end{array}$$

commute. This uniquely determined morphism h is an isomorphism which we consider an identification.

(d) If $I_1 \supset I$, then $I + I_1 = I_1$ and so the above isomorphism h takes the form $h: R/I_1 \to R/I/I_1/I$.

5. PRODUCTS OF RINGS

The object of this section is to show that every nonempty indexed family of rings has a product in the category Rings.

Let $\{R_i\}_{i\in I}$ be a nonempty family of rings. Because each R_i is an abelian group under addition and a monoid under multiplication, the set $\prod_{i\in I} R_i$ has a structure of an additive abelian group and a multiplicative monoid, namely, that given by the product of the indexed family $\{R_i\}_{i\in I}$ of abelian groups and the product of the indexed family $\{R_i\}_{i\in I}$ of multiplicative monoids (see Chapter 2, Section 10). Explicitly, the addition in $\prod_{i\in I} R_i$ is given by $\{r_i\}_{i\in I} + \{r_i'\}_{i\in I} = \{r_i + r_i'\}_{i\in I}$ and the multiplication in $\prod_{i\in I} R_i$ is given by $\{r_i\}_{i\in I}\{r_i'\}_{i\in I} = \{r_i r_i'\}_{i\in I}$ for all $\{r_i\}_{i\in I}$ and $\{r_i'\}_{i\in I}$. The reader can easily check that the set $\prod_{i\in I} R_i$ with this addition and multiplication is a ring. This suggests the following.

Definition

Let $(R_i)_{i\in I}$ be a nonempty indexed family of rings. The ring whose underlying set is $\prod_{i\in I} R_i$ and whose laws of composition are given by

$$\{r_i\}_{i\in I} + \{r_i'\}_{i\in I} = \{r_i + r_i'\}_{i\in I}$$

$$\{r_i\}_{i\in I}\{r_i'\}_{i\in I} = \{r_i r_i'\}_{i\in I}$$

is called the **product of the family** $\{R_i\}_{i\in I}$ of rings. This ring is denoted by $\prod_{i\in I} R_i$.

One defense for this terminology is that $\prod_{i\in I} R_i$ is a product for the nonempty family of rings $\{R_i\}_{i\in I}$ in the category Rings. To see this we must define ring morphisms $\text{proj}_k: \prod_{i\in I} R_i \to R_k$ for each k in I and show that the family $\{\text{proj}_k: \prod_{i\in I} R_i \to R_k\}$ of ring morphisms is a product in the category Rings for the nonempty family $\{R_i\}_{i\in I}$ of rings. Obvious candidates for the ring morphisms $\text{proj}_k: \prod_{i\in I} R_i \to R_k$ are the maps $\prod_{i\in I} R_i \to R_k$ given by $\{r_i\}_{i\in I} \to r_k$ for each $\{r_i\}_{i\in I}$. It is easily checked that these maps are surjective ring morphisms. This suggests the following.

Definition

Let $\{R_i\}_{i\in I}$ be a nonempty family of rings. The ring morphism $\text{proj}_k: \prod_{i\in I} R_i \to R_k$ given by $\text{proj}_k(\{r_i\}_{i\in I}) = r_k$ for each k in I is called the **kth projection morphism**.

The fact that the family of ring morphisms $\{\text{proj}_k: \prod_{i\in I} R_i \to R_k\}_{k\in I}$ is a product for the nonempty indexed family $\{R_i\}_{i\in I}$ of rings is the substance of the following.

Basic Property 5.1
Let $\{R_i\}_{i\in I}$ be a nonempty indexed family of rings and S an arbitrary ring. Then the map

$$\beta_S: \left(S, \prod_{i\in I} R_i\right) \to \prod_{i\in I} (S, R_i)$$

defined by $\beta_S(f) = \{\text{proj}_k f\}_{k\in I}$ for each ring morphism $f: S \to \prod_{i\in I} R_i$ is an isomorphism of sets.

Thus, we see that every nonempty indexed family of rings has a product in the category Rings. This naturally raises the question whether every nonempty indexed family of rings has a sum in the category Rings. Although the answer to the question is always affirmative, we shall only show, in the exercises for this chapter, that every nonempty indexed family of commutative rings has a sum in the category of commutative rings.

EXERCISES

(1) Suppose R is a commutative ring and M is a monoid. Let $R[M]$ be the set of all maps $f: M \to R$ with the property that the set of all m in M such that $f(m) \neq 0$ is a finite subset of M.

(a) Show that if f and g are in $R[M]$, then the map $f+g: M \to R$ defined by $(f+g)(m) = f(m) + g(m)$ for all m in M is in $R[M]$. Prove that the map $R[M] \times R[M] \to R[M]$ given by $(f, g) \to f+g$ is a law of composition which makes $R[M]$ an abelian group whose identity is the map $0: M \to R$ given by $0(m) = 0$ for all m in M.

(b) Show that if f and g are in $R[M]$, then the map $fg: M \to R$ defined by $fg(m) = \Sigma_{m_i m_j = m} f(m_i)g(m_j)$ [where $\Sigma_{m_i m_j = m} f(m_i)g(m_j)$ stands for the sum of the finite number of nonzero terms $f(m_i)g(m_j)$ obtained by letting (m_i, m_j) range over all the distinct ordered pairs of elements of M such that $m_i m_j = m$]. Prove that the map $R[M] \times R[M] \to R[M]$ given by $(f, g) \to fg$ is a law of composition which makes $R[M]$ a monoid whose unit is the map $1: M \to R$ given by $1(1) = 1$ and $1(m) = 0$ if $m \neq 1$ in M.

(c) Prove that $R[M]$ together with the addition and multiplication just defined is a ring. This ring is called **the monoid ring of M over R.**

(d) If we denote by $\Sigma_{m\in M} r_m m$ the map $f: M \to R$ in $R[M]$ such that $f(m) = r_m$ for all m in M, then we see that the ring $R[M]$ can be described as follows:

(i) $R[M]$ consists of all sums $\Sigma_{m\in M} r_m m$ with the elements r_m in R having the property $r_m = 0$ except for a finite set of m in M.

(ii) $\Sigma_{m\in M} r_m m = \Sigma_{m\in M} r'_m m$ if and only if $r_m = r'_m$ for all m in M.

(iii) $\Sigma_{m\in M} r_m m + \Sigma_{m\in M} r'_m m = \Sigma_{m\in M} (r_m + r'_m) m$.

(iv) $(\Sigma_{m\in M} r_m m)(\Sigma_{m\in M} r'_m m) = \Sigma_{m\in M} (\Sigma_{m_i m_j = m} r_{m_i} r_{m_j}) m$.

(v) The zero element of $R[M]$ is the element $\Sigma_{m\in M} r_m m$ with $r_m = 0$ for all m in M.

(vi) The identity of $R[M]$ is the element $\Sigma_{m\in M} r_m m$ with $r_1 = 1$ and $r_m = 0$ for $m \neq 1$.

(e) Show that the map $\phi: R \to R[M]$ given by $\phi(x) = \Sigma_{m \in M} r_m m$ with $r_1 = x$ and $r_m = 0$, for all $m \neq 1$, is an injective morphism of rings whose image consists precisely of the elements $\Sigma_{m \in M} r_m m$ of $R[M]$ satisfying $r_m = 0$ for $m \neq 1$. If we identify, as we usually shall, the element x in R with the element $\phi(x)$ in $R[M]$, we see that R can be viewed as a subring of $R[M]$ and $\phi: R \to R[M]$ becomes the inclusion map.

(f) Show that the map $\psi: M \to R[M]$ given by $\psi(y) = \Sigma_{m \in M} r_m m$ where $r_m = 0$ if $m \neq y$ and $r_m = 1$ if $m = y$ is an injective morphism from the monoid M to the multiplicative monoid of the ring $R[M]$. Thus, if we identify, as we usually shall, the element y in M with $\psi(y)$ in $R[M]$, then M can be viewed as a submonoid of the multiplicative monoid of $R[M]$ and $\psi: M \to R[M]$ becomes the inclusion map.

(g) Show that the ring $R[M]$ is commutative if and only if M is commutative.

(h) Show that the subring R of $R[M]$ is contained in the center of $R[M]$.

(i) Show that if $\mathbf{N}$ is the monoid of nonnegative integers under multiplication, then $R[\mathbf{N}]$ is the ring of polynomials $R[X]$ over R.

(2) Suppose R is a commutative ring. An **R-algebra** Λ is a morphism of rings $f: R \to \Lambda$ such that the image of f is contained in the center of Λ. If $f_1: R \to \Lambda_1$ and $f_2: R \to \Lambda_2$ are two R-algebras, then an R-algebra morphism from f_1 to f_2 is a ring morphism $g: \Lambda_1 \to \Lambda_2$ such that $gf_1 = f_2$.

(a) Show that if $f: R \to \Lambda$ is an R-algebra, then $\mathrm{id}_\Lambda: \Lambda \to \Lambda$ is an R-algebra morphism.

(b) Show that if $f_1: R \to \Lambda_1$, $f_2: R \to \Lambda_2$, and $f_3: R \to \Lambda_3$ are R-algebras and $g_1: \Lambda_1 \to \Lambda_2$ and $g_2: \Lambda_2 \to \Lambda_3$ are R-algebra morphisms, then the usual composition $g_2 g_1: \Lambda_1 \to \Lambda_3$ of ring morphisms is an R-algebra morphism.

(c) Show that the following data define a category which we denote by R-Alg and call the category of R-algebras.

(i) The objects of R-Alg are the R-algebras.

(ii) If $f_1: R \to \Lambda_1$ and $f_2: R \to \Lambda_2$ are two R-algebras, then $R\text{-Alg}(f_1, f_2)$ is the set of R-algebra morphisms from f_1 to f_2.

(iii) If $f_i: R \to \Lambda_i$, $i = 1, 2, 3$ are R-algebras, then the composition map $R\text{-Alg}(f_1, f_2) \times R\text{-Alg}(f_2, f_3) \to R\text{-Alg}(f_1, f_3)$ is defined by $(g_1, g_2) \mapsto g_2 g_1$, the ordinary composition of ring morphisms.

(3) Show that the category of Rings is isomorphic to the category of $\mathbf{Z}$-algebras where $\mathbf{Z}$ is the ring of integers.

(4) Suppose R is a commutative ring, M a monoid, and $f: R \to \Lambda$ an R-algebra. We want to determine the R-algebra morphisms from the R-algebra $\mathrm{inc}: R \to R[M]$ to the R-algebra $f: R \to \Lambda$.

Associated with each R-algebra morphism $g: R[M] \to \Lambda$ is the morphism of monoids $g|M: M \to \Lambda$ where Λ is considered a multiplicative monoid. Hence, we have a map $R\text{-Alg}(R[M], \Lambda) \to \mathrm{Monoid}(M, \Lambda)$ given by $g \to g|M$ for each R-algebra morphism $g: R[M] \to \Lambda$. We now outline a proof that this map $R\text{-Alg}(R[M], \Lambda) \to \mathrm{Monoid}(M, \Lambda)$ is an isomorphism of sets.

(a) Show that if $g_1, g_2: R[M] \to \Lambda$ are two R-algebra morphisms, then $g_1 = g_2$ if and only if $g_1|M = g_2|M$.

(b) Suppose that we are given a morphism of monoids $h: M \to \Lambda$ (remember that Λ is being considered a multiplicative monoid). Show that the map $g: R[M] \to \Lambda$

given by $g(\Sigma_{m\in M} r_m m) = \Sigma_{m\in M} f(r_m)h(m)$ is an R-algebra morphism such that $g|M = h$.

In the next set of exercises we use the notion of a monoid ring to discuss **polynomial rings in several variables**, not just one variable.

Let $\mathbf{N}$ be the additive monoid of nonnegative integers. We have already seen that if R is a commutative ring, then the R-algebra $R[\mathbf{N}]$ is isomorphic to the R-algebra $R[X]$, the ring of polynomials over R. The isomorphism $R[N]\to R[X]$ is given by $\Sigma_{n\in N} r_n \to \Sigma_{n\in N} r_n X^n$. We now discuss **polynomial rings in two variables.**

(5) Let $\mathbf{N}_1$ and $\mathbf{N}_2$ denote two copies of the additive monoid of nonnegative integers $\mathbf{N}$. Suppose R is a commutative ring. Then $R[\mathbf{N}]$ is a commutative ring so we can form the ring $R[\mathbf{N}_1][\mathbf{N}_2]$, which is called the ring of polynomials over R in two variables.

(a) Show that the subset M of $R[\mathbf{N}_1][\mathbf{N}_2]$ consisting of all products $n_1 n_2$ with n_1 in $\mathbf{N}_1$ and n_2 in $\mathbf{N}_2$ is a submonoid of the multiplicative monoid of $R[\mathbf{N}_1][\mathbf{N}_2]$.

(b) By the previous exercise, we know that there is a unique R-algebra morphism $h: R[M]\to R[\mathbf{N}_1][\mathbf{N}_2]$ such that $h|M: M\to R[\mathbf{N}_1][\mathbf{N}_2]$ is the inclusion morphism of monoids. Prove that $h: R[M]\to R[\mathbf{N}_1][\mathbf{N}_2]$ is an isomorphism of R-algebras.

(c) Let $\mathbf{N}_1 \times \mathbf{N}_2$ be the sum of the monoids $\mathbf{N}_1$ and $\mathbf{N}_2$. Show that the map $\mathbf{N}_1 \times \mathbf{N}_2 \to M$ given by $(n_1, n_2)\to n_1 n_2$ is an isomorphism of monoids.

(d) Show that there is a unique morphism of R-algebras $f: R[\mathbf{N}_1 \times \mathbf{N}_2]\to R[M]$ such that $f((n_1, n_2)) = n_1 n_2$ in M for all (n_1, n_2) in $\mathbf{N}_1 \times \mathbf{N}_2$ and that f is an isomorphism of R-algebras. This isomorphism is usually considered an identification of R-algebras.

(e) The composition $R[\mathbf{N}_1 \times \mathbf{N}_2] \xrightarrow{f} R[M] \xrightarrow{h} R[\mathbf{N}_1][\mathbf{N}_2]$ is an isomorphism of R-algebras which is also usually considered an identification of R-algebras.

In dealing with the polynomial ring in two variables $R[\mathbf{N}_1][\mathbf{N}_2]$ over R, the elements n of N_1 are often denoted by X_1^n and the elements n of N_2 are often denoted by X_2^n. Obviously, $X_1^n X_1^{n'} = X_1^{n+n'}$ while $X_2^n X_2^{n'} = X_2^{n+n'}$. Also one usually denotes the R-algebra $R[N_1][N_2]$ by $R[X_1][X_2]$. Clearly, in this notation the submonoid M of $R[X_1][X_2]$ consists of all possible products $X_1^{n_1} X_2^{n_2}$ and is called the **submonoid of monomials** of R. Finally, the R-algebra $R[\mathbf{N}_1 \times \mathbf{N}_2]$ is denoted by $R[X_1, X_2]$. Using the identification of $R[\mathbf{N}_1 \times \mathbf{N}_2]$ with $R[M]$ the elements of $R[X_1, X_2]$ are usually written as $\Sigma_{(n_1,n_2)\in \mathbf{N}\times\mathbf{N}} r_{n_1 n_2} X_1^{n_1} X_2^{n_2}$. The identification $R[X_1, X_2]\to R[X_1][X_2]$ then takes the form

$$\sum_{(n_1,n_2)\in \mathbf{N}\times\mathbf{N}} r_{n_1,n_2} X_1^{n_1} X_2^{n_2} \mapsto \sum_{n_2\in \mathbf{N}} \left(\sum_{n_1\in \mathbf{N}} r_{n_1,n_2} X_1^{n_1}\right) X_2^{n_2}$$

We have already seen that the polynomial ring $R[X]$ has the property that given any **commutative** R-algebra $h: R\to\Lambda$ (that is, an R-algebra $f: R\to\Lambda$ with Λ a commutative ring), the map R-Alg$(R[X], \Lambda)\to\Lambda$ given by $g\mapsto g(X)$ is an isomorphism of sets. In fact, given any λ in Λ, the unique R-algebra morphism $g: R[X]\to\Lambda$ such that $g(X) = \lambda$ is given by $g: \Sigma_{n\in N} r_n X^n \to \Sigma_{n\in N} r_n \lambda^n$. If we follow the usual convention of denoting an element $\Sigma\, r_n X^n$ of $R[X]$ by $f(X)$, then $\Sigma\, r_n \lambda^n$ is denoted by $f(\lambda)$. Hence, in this notation the unique R-algebra morphism $g: R[X]\to\Lambda$ such that $g(X) = \lambda$ is given by $f(X)\mapsto f(\lambda)$ for all $f(X)$ in $R[X]$.

We now want to describe the R-algebra morphisms from $R[X_1, X_2]$ to an arbitrary commutative R-algebra $h: R\to\Lambda$.

(6) Let $h: R\to\Lambda$ be a commutative R-algebra. Show that the map of sets

R-Alg$(R[X_1, X_2], \Lambda) \to \Lambda \times \Lambda$ given by $g \to (g(X_1), g(X_2))$ is an isomorphism of sets by showing:

(a) If $g_1, g_2 : R[X_1, X_2] \to \Lambda$ are two R-algebra morphisms, then $g_1 = g_2$ if and only if $g_1(X_1) = g_2(X_1)$ and $g_1(X_2) = g_2(X_2)$.

(b) If (λ_1, λ_2) is in $\Lambda \times \Lambda$, then we know that the map $g : R[X_1, X_2] \to \Lambda$ given by $g(\Sigma_{(n_1,n_2)\in \mathbf{N}\times\mathbf{N}} r_{n_1,n_2} X_1^{n_1} X_2^{n_2}) = \Sigma_{(n_1,n_2)\in \mathbf{N}\times\mathbf{N}} r_{n_1,n_2} \lambda_1^{n_1} \lambda_2^{n_2}$ is a morphism of R-algebras with the property that $g(X_1) = \lambda_1$ and $g(X_2) = \lambda_2$.

As in the case of one variable, if we denote an element $\Sigma_{(n_1,n_2)\in \mathbf{N}\times\mathbf{N}} r_{n_1,n_2} X_1^{n_1} X_2^{n_2}$ by $f(X_1 X_2)$ and $\Sigma_{(n_1,n_2)\in \mathbf{N}\times\mathbf{N}} r_{n_1,n_2} \lambda_1^{n_1} \lambda_2^{n_2}$ by $f(\lambda_1, \lambda_2)$ for each (λ_1, λ_2) in $\Lambda \times \Lambda$, then given any pair of elements λ_1, λ_2 in Λ, the unique R-algebra morphism $g : R[X_1, X_2] \to \Lambda$ such that $g(X_1) = \lambda_1$ and $g(X_2) = \lambda_2$ is given by $f(X_1, X_2) \mapsto f(\lambda_1, \lambda_2)$ for all $f(X_1, X_2)$ in $R[X_1, X_2]$.

We now want to define a polynomial ring over a commutative ring R in any number of variables.

Let I be a nonempty set and let $\mathbf{N}_i$ be a copy of the additive monoid of nonnegative integers for each i in I. We recall the definition of the standard sum $\coprod_{i\in I} \mathbf{N}_i$ of the indexed family $\{\mathbf{N}_i\}_{i\in I}$ in the category of commutative monoids. As a set, $\coprod_{i\in I} \mathbf{N}_i$ is the subset of $\prod_{i\in I} \mathbf{N}_i$ consisting of all $\{n_i\}_{i\in I}$ with the property that the set of all i in I such that $n_i \neq 0$ is a finite subset of I. Addition in $\coprod_{i\in I} \mathbf{N}_i$ is given by $\{n_i\}_{i\in I} + \{n_i'\}_{i\in I} = \{n_i + n_i'\}_{i\in I}$. Also for each $k \in I$, we have the morphism of monoids $\text{inj}_k : \mathbf{N}_k \to \coprod_{i\in I} \mathbf{N}_i$ given by $\text{inj}_k(n) = \{n_i\}_{i\in I}$ satisfying $n_i = 0$ if $i \neq k$ and $n_k = n$. Finally, we recall that the family $\{\text{inj}_k : \mathbf{N}_k \to \coprod_{i\in I} \mathbf{N}_i\}_{i\in I}$ of morphisms is a sum in the category of commutative monoids. That is, given any commutative monoid M, the map $(\coprod_{i\in I} \mathbf{N}_i, M) \to \prod_{i\in I} (\mathbf{N}_i, M)$ given by $g \mapsto \{g \text{ inj}_k\}_{k\in I}$ is an isomorphism of sets.

We have also seen that the additive monoid $\mathbf{N}$ has the property that the map $(\mathbf{N}, M) \to M$ given by $g \to g(1)$ is an isomorphism of sets. Hence, if for each i in I we denote by M_i a copy of M, we have that the map $(\coprod_{i\in I} \mathbf{N}_i, M) \to \prod_{i\in I} M_i$ given by $g \to \{g \text{ inj}_k(1)\}_{k\in I}$ is an isomorphism of sets.

(7) Let I be a nonempty set and $\{X_i\}_{i\in I}$ an indexed family of distinct symbols X_i. Let M be the set of all symbols $\prod_{i\in I} X_i^{n_i}$ with the $n_i \in \mathbf{N}$ having the property that the set of all i in I with $n_i \neq 0$ is a finite subset of I.

(a) Show that M together with the multiplication given by $\prod_{i\in I} X_i^{n_i} \cdot \prod_{i\in I} X_i^{n_i'} = \prod_{i\in I} X_i^{n_i + n_i'}$ is a commutative monoid. This monoid M is called **monoid of monomials of the family of symbols** $\{X_i\}_{i\in I}$.

(b) Show that the map $\coprod \mathbf{N}_i \to M$ given by $\{n_i\}_{i\in I} \to \prod_{i\in I} X_i^{n_i}$ is an isomorphism of monoids.

(c) For each k in I denote by x_k^n the element $\prod_{i\in I} X_i^{n_i}$ in M with the property that $n_i = 0$ if $i \neq k$ and $n_i = n$ if $i = k$. Show that for each commutative monoid A, the map $(M, A) \to \prod_{i\in I} A_i$, where each $A_i = A$, given by $g \to (g(X_i))_{i\in I}$ is an isomorphism of sets. Let $\{a_i\}_{i\in I}$ be an element of $\prod_{i\in I} A_i$ and $g : M \to A$ the unique morphism such that $g(X_i) = a_i$. Then $g(\prod_{i\in I} X^{n_i})$ is usually written as $\prod_{i\in I} a_i^{n_i}$.

Definition
Let $\{X_i\}_{i\in I}$ be a nonempty family of distinct symbols and R a commutative ring. The **ring of polynomials in the variables** $\{X_i\}_{i\in I}$ **over** R is the commutative R-algebra $R[M]$ where M is the **monoid of monomials** of the indexed set $\{X_i\}_{i\in I}$. This ring is usually denoted by $R[X_i]_{i\in I}$. The elements of $R[X_i]_{i\in I}$ are often denoted by $f(X_i)$.

(8) Suppose $f: R \to \Lambda$ is an arbitrary commutative R-algebra. Show that the map $R\text{-Alg}(R[X_i]_{i\in I}, \Lambda) \to \prod_{i\in I} \Lambda_i$, where each $\Lambda_i = \Lambda$, given by $g \to \{g(X_i))_{i\in I}$ is an isomorphism of sets by showing:

(a) If $g_1, g_2: R[X_i]_{i\in I} \to \Lambda$ are two R-algebra morphisms, then $g_1 = g_2$ if and only if $g_1(X_i) = g_2(X_i)$ for all i in I.

(b) If $\{\lambda_i\}_{i\in I}$ is an element of $\prod_{i\in I} \Lambda_i$, then the map $R[X_i]_{i\in I} \to \prod_{i\in I} \Lambda_i$ given by

$$\sum_{(n_i)\in \amalg \mathbf{N}_i} r_{(n_i)} \prod_{i\in I} X_i^{n_i} \to \sum_{(n_i)\in \amalg \mathbf{N}_i} f(r_{(n_i)}) \prod_{i\in I} \lambda_i^{n_i}$$

is a morphism of R-algebras with the property that $X_i \to \lambda_i$ for all $i \in I$.

Definitions
Let S be a subset of a ring Λ. The **subring of** Λ **generated by** S is the intersection of all the subrings of Λ containing S. If the subring of Λ generated by S is all of Λ, then S is said to generate Λ.

Suppose $h: R \to \Lambda$ is an R-algebra. If S is a subset of Λ, then the ***R*-subalgebra of** Λ **generated by** S is the R-algebra given by $h': R \to \Lambda'$ where Λ' is the subring of Λ generated by S and Im h, and $h': R \to \Lambda'$ is given by $h'(r) = h(r)$ for all r in R. The subset S is said to generate the R-algebra Λ if $\Lambda' = \Lambda$.

(9) Let $h: R \to \Lambda$ be a commutative R-algebra. Let $\{\lambda_i\}_{i\in I}$ be a family of elements of Λ and $R[X_i]_{i\in I}$ the polynomial ring in the variables X_i. Prove that the morphism of R-algebras $g: R[X_i]_{i\in I} \to \Lambda$ given by $g(f(X_i)) = f(\lambda_i)$ for all $f(X_i)$ in $R[X_i]_{i\in I}$ has the property that Im g is the R-subalgebra of Λ generated by the family $\{\lambda_i\}_{i\in I}$ of elements of Λ. This subalgebra generated by $\{\lambda_i\}_{i\in I}$ is sometimes also denoted by $R[\lambda_i]_{i\in I}$.

(10) Show that if $h: R \to \Lambda$ is a commutative R-algebra, then there is a family of variables $\{X_i\}_{i\in I}$ such that there is a surjective morphism of R-algebras $R[X_i]_{i\in I} \to \Lambda$.

(11) Let $\{X_i\}_{i\in I}$ be a family of variables over the commutative ring R. Let $\{J_1, J_2\}$ be a partition of I.

(a) Let M_k be the monoid of monomials in $\{X_i\}_{i\in J_k}$ for $k = 1, 2$ and M the monoid of monomials for $\{X_i\}_{i\in I}$. Define for each $k = 1, 2$ the map $M_k \to M$ by $\prod_{j\in J_1} X_j^{n_j} \to \prod_{i\in I} X_i^{n_i}$ where $n_i = n_j$ for all i in J_k and $n_i = 0$ if i is not in J_k. Show that the map $M_k \to M$ is an injective morphism of monoids for $k = 1, 2$. Usually one identifies M_k with its image in M by means of the injective morphism $M_k \to M$ just described.

(b) Let $g: R[X_j]_{j\in J_1} \to R[X_i]_{i\in I}$ be the unique morphism of R-algebras which has the property that $g|M_1: M_1 \to R[M]$ is the composition $M_1 \to M \xrightarrow{\text{inc}} R[M]$. Show that g is an injective morphism of R-algebras. Usually one identifies $R[X_j]_{j\in J_1}$

with its image in $R[X_i]_{i\in I}$ by means of the injective morphism g just described.

(c) Since $R[X_j]_{j\in J_1}$ is a subring of the commutative ring $R[X_i]_{i\in I}$, we can view $R[X_i]_{i\in I}$ as an $R[X_j]_{j\in J_1}$-algebra. Show that the unique $R[X_j]_{j\in J_1}$ morphism $R[X_j]_{j\in J_1}([X_k]_{k\in J_2})\to R[X_i]_{i\in I}$ which extends the composition $M_2\to M\to R[X_i]_{i\in I}$ is an isomorphism of $R[X_j]_{j\in J_1}$-algebras and hence of R-algebras. This isomorphism is usually considered an identification.

(12) Let $\mathscr{C}$ be the full subcategory of R-Alg consisting of the commutative R-algebras. Let $\{X_i\}_{i\in I}$ and $\{Y_j\}_{j\in J}$ be disjoint indexed families of variables over R. Let $K = I \amalg J$, the sum of the sets I and J, and let $\{Z_k\}_{k\in K}$ be the family of variables given by $Z_k = X_k$ if $k\in I$ and $Z_k = Y_k$ if k is in J.

(a) Show that the natural injective morphisms $g: R[X_i]_{i\in I}\to R[Z_k]_{k\in K}$ and $h: R[Y_j]_{j\in J}\to R[Z_k]_{k\in K}$ given by $g(X_i) = X_i$ in $R[Z_k]_{k\in K}$ and $h(Y_j) = Y_j$ for j in J are a sum in the category $\mathscr{C}$ for the pair of R-algebras $R[X_i]_{i\in I}$ and $R[Y_j]_{j\in J}$.

(b) Suppose A is an ideal in $R[X_i]$ and B is an ideal in $R[Y_j]$. Let A' and B' be the images respectively of A and B in $R[Z_k]_{k\in K}$. Show that the subset C of $R[Z_k]$, consisting of all finite sums of elements of the form $f(X_k)a' + g(X_k)b'$ with a' in A', b' in B', and $f(Z_k)$, $g(Z_k)$ arbitrary elements of $R[Z_k]_{k\in K}$ is an ideal in $R[Z_k]_{k\in K}$.

(c) Show that the morphisms of R-algebras $R[X_i]_{i\in I}\to R[Z_k]_{k\in K}$ and $R[Y_j]_{j\in J}\to R[Z_k]_{k\in K}$ induce morphisms of R-algebras $R[X_i]_{i\in I}/A\to R[Z_k]_{k\in K}/C$ and $R[Y_j]_{j\in J}/B\to R[Z_k]_{k\in K}/C$ and that these morphisms are a sum of the R-algebras $R[X_i]_{i\in I}/A$ and $R[Y_j]_{j\in J}/B$ in the category $\mathscr{C}$.

(d) Show that any two R-algebras in $\mathscr{C}$ have a sum in $\mathscr{C}$.

(e) Show that every finite family of R-algebras in $\mathscr{C}$ has a sum in $\mathscr{C}$.

(f) Show that every family of R-algebras in $\mathscr{C}$ has a sum in $\mathscr{C}$.

(13) Let $\mathscr{C}$ be the category of commutative monoids. Show that the following data define a functor G: Sets$\to\mathscr{C}$.

(a) G: Ob Sets$\to$ Ob $\mathscr{C}$ is given by $G(X) = \coprod_{x\in X} \mathbf{N}_x$ where each $\mathbf{N}_x = \mathbf{N}$, the additive monoid of nonnegative integers.

(b) Given a map $f: X\to Y$ of sets, $G(f): \coprod_{x\in X} \mathbf{N}_x\to \coprod_{y\in Y} \mathbf{N}_y$ is the unique morphism of monoids such that for each u in X the composition $G(f)\,\mathrm{inj}_u: \mathbf{N}_u\to \coprod_{y\in Y} \mathbf{N}_y$ is the morphism $g_u: \mathbf{N}\to \coprod_{y\in Y} \mathbf{N}_y$ given by $g_u(n) = \{m_y\}_{y\in Y}$ where $m_y = 0$ if $y\neq f(u)$ and $m_y = n$ if $y = f(u)$.

Let $F: \mathscr{C}\to$ Sets be the forgetful functor. Then for each commutative monoid M and each set X define the map of sets $\psi_{M,X}: \mathscr{C}(G(X), M)\to$ Sets $(X, F(M))$ by $\psi_{M,X}(\alpha)(x) = \alpha(\{n_y\}_{y\in X})$ where $n_y = 0$ if $y\neq x$ and $n_x = 1$. Show that:

(c) Each of the maps $\psi_{M,X}$ is an isomorphism of sets.

(d) If $f: X\to Y$ is a map of sets, then the diagram

$$\begin{array}{ccc} \mathscr{C}(G(Y), M) & \xrightarrow{\psi_{Y,M}} & \mathrm{Sets}(Y, F(M)) \\ \downarrow{\scriptstyle (G(f), M)} & & \downarrow{\scriptstyle (f, F(M))} \\ \mathscr{C}(G(X), M) & \xrightarrow{\psi_{X,M}} & \mathrm{Sets}(X, F(M)) \end{array}$$

commutes.

(e) If $g: L \to M$ is a morphism of monoids, then the diagram

$$\begin{array}{ccc} \mathscr{C}(G(X), L) & \xrightarrow{\psi_{X,L}} & \mathrm{Sets}(X, F(L)) \\ \downarrow{\scriptstyle (G(X), g)} & & \downarrow{\scriptstyle (X, F(g))} \\ \mathscr{C}(G(X), M) & \xrightarrow{\psi_{X,M}} & \mathrm{Sets}(X, F(M)) \end{array}$$

commutes.

Exercise 13 is an example of a very general and important concept.

Definition

Let $F: \mathscr{C} \to \mathscr{D}$ be a functor of categories. A functor $G: \mathscr{D} \to \mathscr{C}$ is said to be a **left adjoint of** F if there is for each pair of objects C in $\mathscr{C}$ and D in $\mathscr{D}$ a map of sets $\psi_{D,C}: \mathscr{C}(G(D), C) \to \mathscr{D}(D, F(C))$ satisfying:

(a) Each $\psi_{C,D}$ is an isomorphism of sets.

(b) If $f: D \to D'$ is a morphism in $\mathscr{D}$, then for each C in $\mathscr{C}$ the diagram

$$\begin{array}{ccc} \mathscr{C}(G(D'), C) & \xrightarrow{\psi_{D',C}} & \mathscr{D}(D', F(C)) \\ \downarrow{\scriptstyle (G(f), C)} & & \downarrow{\scriptstyle (f, F(C))} \\ \mathscr{C}(G(D), C) & \xrightarrow{\psi_{D,C}} & \mathscr{D}(D, F(C)) \end{array}$$

commutes.

(c) If $g: C \to C'$ is a morphism in $\mathscr{C}$, then for each D in $\mathscr{D}$, the diagram

$$\begin{array}{ccc} \mathscr{C}(G(D), C) & \xrightarrow{\psi_{D,C}} & \mathscr{D}(D, F(C)) \\ \downarrow{\scriptstyle (G(D), g)} & & \downarrow{\scriptstyle (D, F(g))} \\ \mathscr{C}(G(D), C') & \xrightarrow{\psi_{D,C'}} & \mathscr{D}(D, F(C')) \end{array}$$

commutes.

Given a pair of functors $F: \mathscr{C} \to \mathscr{D}$ and $G: \mathscr{D} \to \mathscr{C}$ we say that G is a **right adjoint of** F if F is a left adjoint of G.

(14) Suppose $F: \mathscr{C} \to \mathscr{D}$ is a functor of categories.

(a) Show that if $G, G': \mathscr{D} \to \mathscr{C}$ are both right (left) adjoint of F, then G and G' are isomorphic functors.

(b) Suppose $G: \mathscr{D} \to \mathscr{C}$ is a left adjoint of F. Show:

(i) If $f: C \to C'$ is a monomorphism in $\mathscr{C}$, then $F(f): F(C) \to F(C')$ is a monomorphism $\mathscr{D}$. [*Hint*: Use the fact that the isomorphisms $\psi_{D,C}: \mathscr{C}(G(D), C) \to \mathscr{D}(D, F(C))$ have the property that for all D in $\mathscr{D}$, the diagrams

$$\begin{array}{ccc} \mathscr{C}(G(D, C) & \longrightarrow & \mathscr{D}(D, F(C)) \\ \downarrow{\scriptstyle (G(D), f)} & & \downarrow{\scriptstyle (D, F(f))} \\ \mathscr{C}(G(D), C') & \longrightarrow & \mathscr{D}(D, F(C')) \end{array}$$

commute.

(ii) In a similar way show that if $\{f_i : C \to C_i\}_{i \in I}$ is a product for the indexed family $\{C_i\}_{i \in I}$ of objects in $\mathscr{C}$, then $\{F(f_i) : F(C) \to F(C_i)\}_{i \in I}$ is a product in $\mathscr{D}$ for the family $\{F(C_i)\}_{i \in I}$ of objects in $\mathscr{D}$.

(iii) Show that if $g : D \to D'$ is an epimorphism in $\mathscr{D}$, then $G(g) : G(D) \to G(D')$ is an epimorphism in $\mathscr{C}$.

(iv) Show that if $\{g_i : D_i \to D\}_{i \in I}$ is a sum for the family $\{D_i\}_{i \in I}$ of objects in D, then $\{G(g_i) : G(D_i) \to G(D)\}$ is a sum for the family $\{G(D_i)\}_{i \in I}$ in $\mathscr{C}$.

(15) Let R be a commutative ring and $\mathscr{C}$ the category of commutative R-algebras. Show that the following data define a functor $G : \mathrm{Sets} \to \mathscr{C}$.

(a) $G : \mathrm{Ob\ Sets} \to \mathrm{Ob}\ \mathscr{C}$ is given by $G(I) = R[X_i]_{i \in I}$.

(b) If $f : I \to J$ is a map of sets, then $G(f) : R[X_i]_{i \in I} \to R[X_j]_{j \in J}$ is the unique R-algebra morphism having the property $G(f)(X_i) = X_{f(i)}$ for all i in I.
Show that the forgetful functor $F : \mathscr{C} \to \mathrm{Sets}$ is a right adjoint of G.

(16) Let Λ be an arbitrary ring and n a nonzero positive integer.

(a) Show that the following data define a ring which we denote by $M_n(\Lambda)$ and call the **ring of $n \times n$ matrices over Λ**.

(i) As a set $M_n(\Lambda)$ consists of all square arrays $(\lambda_{ij})_{(i,j) \in [1,n] \times [1,n]}$, that is

$$(\lambda_{ij}) = \begin{pmatrix} \lambda_{11}, \ldots, \lambda_{1n} \\ \vdots \\ \lambda_{n1}, \ldots, \lambda_{nn} \end{pmatrix}$$

(ii) Addition in $M_n(\Lambda)$ is given by $(\lambda_{ij}) + (\lambda'_{ij}) = (\lambda_{ij} + \lambda'_{ij})$.

(iii) Multiplication in $M_n(\Lambda)$ is given by

$$(\lambda_{ij}) \cdot (\lambda_{jk}) = \left(\sum_{j=1}^{n} \lambda_{ij} \lambda_{jk} \right)_{(i,k) \in [1,n] \times [1,n]}$$

(b) Show that if $f : \Lambda \to \Lambda'$ is a morphism of rings, then the map $M_n(f) : M_n(\Lambda) \to M_n(\Lambda')$ defined by $M_n(f)(\lambda_{ij}) = (f(\lambda_{ij}))$ is a morphism of rings having the following properties:

(i) f is injective if and only if $M_n(f)$ is injective.

(ii) f is surjective if the only if $M_n(f)$ is surjective.

(iii) f is an isomorphism of rings if and only if $M_n(f)$ is an isomorphism of rings.

(iv) Show that if $\mathrm{Ker}\ f = I$, then $\mathrm{Ker}\ (M_n(f))$ consists of precisely all (λ_{ij}) with λ_{ij} in I for all (i, j) in $[1, n] \times [1, n]$.

(c) Show that the map $\Lambda \to M_n(\Lambda)$ given by $\lambda \to (x_{ij})$ where $x_{ii} = \lambda$ for all $i = 1, \ldots, n$ and $x_{ij} = 0$ if $i \neq j$ is an injective morphism of rings.

(17) Suppose $I \subset M_2(\Lambda)$ is an ideal of $M_2(\Lambda)$. Show that:

(a) If $\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$ is in I, then each of the terms $\begin{pmatrix} \lambda_{11} & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & \lambda_{12} \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ \lambda_{21} & 0 \end{pmatrix}$, and $\begin{pmatrix} 0 & 0 \\ 0 & \lambda_{22} \end{pmatrix}$ as well as the terms $\begin{pmatrix} \lambda\lambda_{11} & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} \lambda_{11}\lambda & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & \lambda\lambda_{12} \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & \lambda_{12}\lambda \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ \lambda\lambda_{21} & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ \lambda_{21}\lambda & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 0 & \lambda\lambda_{22} \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 0 & \lambda_{22}\lambda \end{pmatrix}$ are in I for all λ in Λ.

(b) Show that the set of all λ in Λ with the property that there is a (λ_{ij}) in I such that $\lambda = \lambda_{ij}$ for some i, j in $[1, 2]\times[1, 2]$ is an ideal in Λ which we denote by I'.

(c) Show that I is the set of all (λ_{ij}) in $M_2(\Lambda)$ such that each λ_{ij} is in I'.

(d) Show that the map of sets Ideals$(M_2(\Lambda))\to$Ideals (Λ) given by $I\to I'$ is a bijective map.

(e) Show that if $g: M_2(\Lambda)\to\Gamma$ is a surjective ring morphism with Ker $g = I$, then Γ is isomorphic to the ring $M_2(\Lambda/I')$.

(f) Show that the center of $M_2(\Lambda)$ is the subset of $M_2(\Lambda)$ consisting of all elements $\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}$ with λ in the center of Λ.

(g) A ring Λ is said to be a **simple ring** if it is not the zero ring and (0) and Λ are the only ideals of Λ. Show that $M_2(\Lambda)$ is a simple ring if and only if Λ is a simple ring.

(18) Generalize the results of Exercise 17 to arbitrary $M_n(\Lambda)$.

(19) Show that if Λ is a nonzero commutative ring, then Λ and $M_n(\Lambda)$ are isomorphic rings if and only if $n = 1$. Does the same thing hold if Λ is not commutative?

(20) Suppose R is a nonzero commutative ring.

(a) Show that the following conditions are equivalent:

(i) R is a simple ring.

(ii) If x is in R and $x\neq 0$, then there is a y in R such that $xy = 1$.

A nonzero commutative ring satisfying either of these conditions is called a **field**.

(b) Let $\mathbf{Z}$ be the ring of integers.

(i) Show that ideals of $\mathbf{Z}$ are precisely the subgroups of $\mathbf{Z}$.

(ii) Show that $\mathbf{Z}/n\mathbf{Z}$ is a field if and only if n is a prime integer.

(c) Show that a commutative ring R is a field if and only if $M_n(R)$ is a simple ring for all n.

(21) Let Λ be a ring. Denote by $T_n(\Lambda)$ the subset of $M_n(\Lambda)$ consisting of all (λ_{ij}) in $M_n(\Lambda)$ with $\lambda_{ij} = 0$ if $i < j$.

(a) Show that $T_n(\Lambda)$ is a subring of $M_n(\Lambda)$.

(b) Show that the subset I of $T_n(\Lambda)$ consisting of all (λ_{ij}) such that $\lambda_{ii} = 0$ for all i is a proper ideal of $T_n(\Lambda)$.

(c) Describe the ring $T_n(\Lambda)/I$.

(d) Show that $T_n(\Lambda)$ is not simple even if Λ is simple provided $n > 1$.

(22) Show that a ring Λ which has the property $\lambda^2 = \lambda$ for all λ in Λ is a commutative ring.

(23) Let G be a finite group and R a commutative ring. Show that the center of $R[G]$ is not R. [*Hint*: Consider the element $\Sigma_{g\in G} g$ in $R[G]$.]

(24) Prove Basic Properties 1.1.

(25) Prove that $M_2(R)$ is not a commutative ring if R is not the zero ring.

(26) Write out a detailed proof of Proposition 2.2.

(27) Write out a detailed proof of Basic Properties 3.4.

(28) Write out a detailed proof of Proposition 4.3.

(29) Write out a detailed proof of Proposition 4.4.

Let $\mathscr{C}$ be a category. Suppose that for every pair of objects X and Y of $\mathscr{C}$, the set of morphisms $\mathscr{C}(X, Y)$ is an abelian group. If $f: X\to Y$ and $g: X\to Y$ are in

$\mathscr{C}(X, Y)$, denote by $f+g: X \to Y$ their sum in $\mathscr{C}(X, Y)$. Suppose, in addition, that for every triple of objects X, Y, Z of $\mathscr{C}$, and every pair of morphisms $f: X \to Y$ and $g: X \to Y$, we have $(f+g)h = fh + gh$ for all h in $\mathscr{C}(Z, X)$ and $h'(f+g) = h'f + h'g$ for all h' in $\mathscr{C}(Y, Z)$. In this case, $\mathscr{C}$ is called a **preadditive category**.

(30) Prove that if $\mathscr{C}$ is a preadditive category and $0: X \to Y$ is the zero element of the group $\mathscr{C}(X, Y)$, then $0f = 0$ and $g0 = 0$ for all f in $\mathscr{C}(Z, X)$ and all g in $\mathscr{C}(Y, Z)$.

(31) It was remarked at the end of Chapter 3 that if X and Y were abelian groups and if $f: X \to Y$ and $g: X \to Y$ were group morphisms, then $f+g: X \to Y$ defined by setting $(f+g)(x) = f(x) + g(x)$ is again a group morphism from X to Y. Prove that the category of abelian groups, which we denote by $\mathscr{Ab}$, is a preadditive category, where the group operation in $\mathscr{Ab}(X, Y)$ is that described above.

(32) Let R be a ring. Show that the following data define a preadditive category which we denote by $\mathscr{C}(R)$ and call the **category of the ring** R.

- **(i)** Ob $\mathscr{C}(R)$ is the set consisting of the single element R.
- **(ii)** The set of morphisms $\mathscr{C}(R)(R, R)$ is the set R.
- **(iii)** The composition map $(R, R) \times (R, R) \to (R, R)$ is given by $(r_1, r_2) \to r_2 \circ r_1 = r_1 r_2$ where $r_1 r_2$ is the product in R of the elements r_1 and r_2 in R.
- **(iv)** The addition map $(R, R) \times (R, R) \to (R, R)$ is given by $(r_1, r_2) \mapsto r_1 + r_2$ where $r_1 + r_2$ is the sum in R of the elements r_1 and r_2 in R.

(33) Let $\mathscr{C}$ and $\mathscr{D}$ be preadditive categories. A functor $F: \mathscr{C} \to \mathscr{D}$ is said to be additive if for every pair of objects X and Y of $\mathscr{C}$, $F(f+g) = F(f) + F(g)$ for every pair of morphisms $f: X \to Y$ and $g: X \to Y$.

(a) Show that if $F: \mathscr{C} \to \mathscr{D}$ and $G: \mathscr{C} \to \mathscr{D}$ are additive functors and $\phi: F \to G$ and $\psi: F \to G$ are two morphisms from F to G, then $\phi + \psi: F \to G$ is a morphism of functors where $(\phi + \psi)(X): F(X) \to G(X)$ is defined to be $\phi(X) + \psi(X)$ for every object X of $\mathscr{C}$.

(b) Let $\mathscr{C}$ be a small preadditive category and $\mathscr{D}$ any preadditive category. We denote $(\mathscr{C}, \mathscr{D})^+$ the full subcategory of $(\mathscr{C}, \mathscr{D})$ whose objects are the additive functors from $\mathscr{C}$ to $\mathscr{D}$. For each pair of objects F, G in $(\mathscr{C}, \mathscr{D})^+$, define addition in (F, G) as in part (a). Prove that with this addition, $(\mathscr{C}, \mathscr{D})^+$ is a preadditive category.

(34) Let $\mathscr{C}$ be a preadditive category and let $f_1: X \to X$, $f_2: Y \to Y$ be objects of $\mathscr{C}[X]$. If $g_1: f_1 \to f_2$ and $g_2: f_1 \to f_2$ are morphisms of $\mathscr{C}[X]$, then g_1 and g_2 are morphisms from X to Y such that $g_1 f_1 = f_2 g_1$ and $g_2 f_1 = f_2 g_2$. Prove that the morphism $g_1 + g_2$ from X to Y in $\mathscr{C}$ also has the property that $(g_1 + g_2) f_1 = f_2 (g_1 + g_2)$ so that $g_1 + g_2$ is a morphism from f_1 to f_2 in $\mathscr{C}[X]$. With this addition of morphisms in $\mathscr{C}[X]$, prove that $\mathscr{C}[X]$ is a preadditive category.

(35) Let R be a commutative ring, let $\mathscr{C}(R)$ be the category of the ring R, and let $\mathscr{Ab}$ be the category of abelian groups. Denote the preadditive category $(\mathscr{C}(R), \mathscr{Ab})^+$ by $\mathrm{Mod}(R)$. Similarly, denote the preadditive category $(\mathscr{C}(R[X]), \mathscr{Ab})^+$ by $\mathrm{Mod}(R[X])$, where $R[X]$ is the polynomial ring over R.

(a) Prove that the following data define a functor T from $(\mathrm{Mod}(R))[X]$ to $\mathrm{Mod}(R[X])$.

- **(i)** If $\phi: F \to F$ is an object of $(\mathrm{Mod}(R))[X]$, let $T(\phi)$ be the functor from $\mathscr{C}(R[X])$ to $\mathscr{Ab}$ defined by:
  - (a) $T(\phi)(R[X]) = F(R)$.
  - (b) $T(\phi)(\Sigma r_n X^n): F(R) \to F(R)$ is the morphism defined by $a \to$

$\Sigma F(r_n)\phi(R)^n(a)$ for all a in $F(R)$, where $\phi(R)^n$ means the composition of the endomorphism $\phi(R): F(R) \to F(R)$ with itself n times, and $F(r_n): F(R) \to F(R)$ is the morphism defined by the functor F.

(ii) If $g: \phi_1 \to \phi_2$ is a morphism in $(\mathrm{Mod}(R))[X]$, define $T(g): T(\phi_1) \to T(\phi_2)$ by letting $T(g)(R[X]): T(\phi_1)(R[X]) \to T(\phi_2)(R[X])$ be the morphism $g(R): F_1(R) \to F_2(R)$ where $\phi_1: F_1 \to F_1$ and $\phi_2: F_2 \to F_2$.

(b) Prove that the functor T defined above is additive.

(c) Prove that the functor T is an isomorphism of categories.

(36) Let Group be the category of all groups and $\mathscr{A}$ the full subcategory of Group consisting of the abelian groups. Show that the inclusion functor $i: \mathscr{A} \to$ Group has a left adjoint.

(37) Let $\mathscr{C}$ be the category of commutative monoids and $\mathscr{A}$ the full subcategory of $\mathscr{C}$ consisting of the abelian groups. Show that the inclusion functor $i: \mathscr{A} \to \mathscr{C}$ has a left adjoint.

(38) Show that the following data define a functor $F:$ Rings $\to$ Monoid:

(a) The map $F:$ Ob Rings $\to$ Ob Monoid is given by $F(R)$ is the multiplicative monoid of the ring R for each ring R.

(b) For each ring morphism $f: R_1 \to R_2$ we define $F(f): F(R_1) \to F(R_2)$ to be the map f viewed as a morphism of the multiplicative monoid of R_1 to that of R_2. Prove that the functor $F:$ Rings $\to$ Monoid has a left adjoint.

(39) Generalize Exercise 38 to the category of R-algebras for any commutative ring R.

PART TWO

Chapter 5 UNIQUE FACTORIZATION DOMAINS

In this chapter we will be mainly concerned with examples and properties of commutative rings with which the reader is for the most part familiar. For example, the basic properties of $\mathbf{Z}$, the ring of integers, and $K[X]$, the ring of polynomials over a field K, are discussed, including the fact that they are unique factorization domains. We will also show that the ring $R[X]$ of polynomials over a ring R is a unique factorization domain if and only if the ring R is a unique factorization domain. From these sample results it is obvious that one of our major preoccupations in this chapter is the question of when commutative rings are unique factorization domains.

The reader who is at all familiar with the notion of a ring being a unique factorization domain should have no difficulty seeing that this idea is intimately connected with the general one of divisibility in a ring. For instance, we usually say that a nonzero integer $n \neq 1$ in the ring $\mathbf{Z}$ of all integers is a prime if and only if ± 1 and $\pm n$ are the only integers which divide n. Further, the fact that every integer can be written (in an essentially unique way) as a finite product of primes is also a statement concerning how integers divide each other. Because for rings generally, and not just for integers, questions of divisibility are related to unique factorization, we begin this chapter by studying divisibility in commutative rings. Related matters such as unique factorization and rings of quotients will be taken up later on.

Because we are only interested in commutative rings in this chapter, we make the blanket assumption that unless stated to the contrary all rings are commutative. We remind the reader that since we are assuming that our rings are commuta-

tive, an ideal in a ring R is simply a subgroup I of R satisfying the condition that $rI \subset I$ for all r in R.

1. DIVISIBILITY

We begin by recalling what it means for one element in a ring to divide another.

Definition
Let x and y be elements in a ring R. We say that x **divides** y if there is an element z in R such that $xz = y$. We often denote the fact that x divides y by writing $x|y$.

We leave it to the reader to verify the following.

Basic Properties 1.1
Let R be a ring.

(a) For each element x in R we have that $x|x$.
(b) If x, y, and z are elements in R such that $x|y$ and $y|z$, then $x|z$.
(c) For a fixed element x in R, the set of all elements in R divisible by x is the set Rx consisting of all elements of the form rx with r in R.
(d) For each x in R, the set Rx of all elements of R divisible by x is the unique ideal J of R satisfying:
  (i) x is in J.
  (ii) If I is an ideal of R containing x, then $J \subset I$.

For each element x in R, because the set Rx of all elements in R divisible by x is an ideal in R, it is reasonable to expect that ideals of this type play an important role in studying divisibility. For this reason we give such ideals a special name.

Definitions
For each x in R, the ideal Rx is called the ideal or the **principal ideal generated by the element** x. We will often use the notation (x) for the ideal Rx generated by x.

An ideal I in R is called a **principal ideal** if there is an element $x \in I$ such that $Rx = I$.

The reader should have no difficulty verifying the following.

Basic Properties 1.2
Let x and y be elements in a ring R.

(a) $x|y$ if and only if $(x) \supset (y)$.
(b) $(x) = (y)$ if and only if $x|y$ and $y|x$.
(c) For an element x in R, the following statements are equivalent:
  (i) $x|1$.
  (ii) $(x) = R = (1)$.
  (iii) x is an invertible element in the multiplicative monoid of R.
  (iv) $x|y$ for all y in R.
  (v) $(xy) = (y)$ for all y in R.

In a ring R because we have $x|y$ if and only if $(x) \supset (y)$, we see that the study of the way the elements of R divide each other is the same thing as studying the order relation given by inclusion on the set of principal ideals of R. Stated more symbolically, if we denote the set of all principal ideals of R by $PI(R)$, then the map $f: R \to PI(R)$ given by $f(x) = Rx$ for all x in R is a surjective map with the property that $x|y$ if and only if $f(x) \supset f(y)$.

But this is not the only relationship between R and $PI(R)$. For it is not difficult to show that $PI(R)$ has a (unique) commutative monoid structure such that with this monoid structure the map $f: R \to PI(R)$ is a morphism from the multiplicative monoid of R to $PI(R)$. The existence of this monoid structure in $PI(R)$ is based on the following general definition.

Definition
Let I and J be ideals in the ring R. The set of all finite sums $\Sigma x_i y_i$ with x_i in I and y_i in J is an ideal in R which we denote by IJ and call the **product of I and J.**

It is easily checked that the product of ideals in a ring has the following set of properties.

Basic Properties 1.3
Let I_1, I_2, and I_3 be ideals in a ring R. Then:

(a) $I_1 I_2 = I_2 I_1$.
(b) $RI_1 = I_1 R = I_1$.
(c) $I_1(I_2 I_3) = (I_1 I_2) I_3$.
(d) If $I_1 \supset I_2$, then $I_3 I_1 \supset I_3 I_2$.
(e) The product of two principal ideals is again a principal ideal because $(x)(y) = (xy)$ for all x and y in R.

Thus, we see that the product of ideals defined above makes the set of all ideals in R a commutative, multiplicative monoid with the principal ideal $R = (1)$ as identity element. Because the product of principal ideals is again a principal ideal, we see that the set $PI(R)$ of all principal ideals in R is a submonoid of the monoid of all ideals in R. It is obvious that this is the unique monoid structure on $PI(R)$ which makes the surjective map $f: R \to PI(R)$ a monoid morphism from the multiplicative monoid of R to $PI(R)$.

Definition
If R is a ring, we denote by $PI(R)$ the commutative monoid whose elements are the principal ideals of R and whose multiplication is given by $(x)(y) = (xy)$ for all x and y in R.

Because the morphism $f: R \to PI(R)$ from the multiplicative monoid of R to $PI(R)$ is surjective, we know that the canonical morphism $j_f: \text{Coim } f \to PI(R)$ given by $j_f([x]) = f(x)$ for all x in R is an isomorphism of monoids. Hence, any description of the monoid Coim f gives an alternate description of $PI(R)$. So we now turn our attention to studying Coim f.

We have already shown that if $x|1$, then $f(xy) = Rxy = Ry = f(y)$ for all y in R. Hence, it is obvious that the set of all elements x in R such that $x|1$ plays an

important role in describing Coim f. Because this type of element plays a significant role in all of ring theory, not just commutative ring theory, we make the general statement.

Definition
Let R be an arbitrary, not necessarily commutative, ring. An element x in R is called a **unit** in R if x is an invertible element of the multiplicative monoid of R, that is, there is a y such that $xy = 1 = yx$. The group $\mathrm{Inv}(R)$, which is the submonoid of the multiplicative monoid of R consisting of all units or invertible elements in R, is called the **group of units** of R and is denoted by $U(R)$.

Returning to the morphism of monoids $f: R \to PI(R)$ in the case when R is commutative, we see that since $f(uy) = f(y)$ for all u in $U(R)$ we have $xU(R) \subset [x]$ for each x in R where $[x]$ is the unique element of Coim f containing x. Although it is not true for arbitrary rings R that $xU(R) = [x]$ for all x in R (see the exercises for an example), it is true for rings R which are integral domains. We recall the following.

Definition
An element x in a ring R is said to be **regular** if $xy = 0$ implies $y = 0$. A ring R is said to be an **integral domain** if $R \neq (0)$ and every nonzero element in R is regular.

Clearly, every subring of an integral domain is also an integral domain.

As an easy consequence of this definition we have the following.

Basic Properties 1.4
Let R be a ring.

(a) An element x in R is regular if and only if $xy_1 = xy_2$ implies $y_1 = y_2$.
(b) The set of all regular elements in R is a submonoid of the multiplicative monoid of R.
(c) Because each unit in R is regular, we have that $U(R)$, the group of units in R, is a submonoid of the monoid of regular elements of R.
(d) If x and y are elements of R such that x is regular and $(x) = (y)$, then y is regular and there is a unit u in R such that $ux = y$.
(e) If x and y are elements in an integral domain, then $(x) = (y)$ if and only if there is a unit u in R such that $ux = y$.

PROOF:(a)–(c) Left as exercises.

(d) Because we are assuming that $(x) = (y)$, we know there are elements u_1 and u_2 in R such that $u_1x = y$ and $u_2y = x$. Hence, $u_2u_1x = x$ or equivalently $x(u_2u_1 - 1) = 0$. The fact that x is regular implies $u_1u_2 - 1 = 0$. Thus, u_1 and u_2 are units in R. Therefore, $y = u_1x$ where u_1 is a unit in R. This also implies y is regular, since both u_1 and x are regular and the product of regular elements is regular [see parts (b) and (c)].

(e) Follows trivially from (d).

As a consequence of this discussion we have the following description of the coimage of the morphism from the multiplicative monoid of an integral domain R to $PI(R)$.

Proposition 1.5
Let R be an integral domain and $f: R \to PI(R)$ the surjective morphism from the multiplicative monoid of R to $PI(R)$ given by $f(x) = Rx$.

(a) The elements of Coim f consist of the subsets of R of the form $xU(R)$ for all x in R.
(b) The multiplication in Coim f is given by $(x_1U(R)) \cdot (x_2U(R)) = x_1x_2U(R)$ for all x_1 and x_2 in R.

2. INTEGRAL DOMAINS

Before giving examples to illustrate some of the material of the preceding section, we recall some of the basic facts concerning fields.

Definition
A ring R is called a **field** if $R \neq (0)$ and every nonzero element of R is a unit in R, or what amounts to the same thing, $U(R) = R - \{0\}$.

Basic Properties 2.1
Suppose R is a ring and $R \neq (0)$.

(a) If R is a field, then R is an integral domain.
(b) R is a field if and only if (0) and R are the only ideals of R.
(c) If S is a nontrivial ring and R is a field, then every morphism of rings $f: R \to S$ is an injective morphism.

PROOF: (a) Already proven.

(b) Suppose R is a ring such that (0) and R are the only ideals in R. Let x be a nonzero element in R. Then the ideal (x) is not the zero ideal and so must be the whole ring R. This means that there is a y in R such that $yx = 1$ or equivalently, x is a unit in R. Hence, every nonzero element in R is a unit in R which means that R is a field. The fact that if R is a field, then (0) and R are the only ideals of R is obvious.

(c) Follows trivially from (b).

Example 2.2 We have already shown that the ring $\mathbf{Z}$ of all integers is an integral domain (see Chapter 2, Basic Properties 9.2). Also, $U(\mathbf{Z})$ consists of 1 and -1 and is thus isomorphic to the group $\mathbf{Z}/2\mathbf{Z}$.

Example 2.3 Let m_1 and m_2 be any two nonzero elements of $\mathbf{Z}$ which are not units. Then the ring $\mathbf{Z}/(m_1m_2)$ is not an integral domain because $m_1 + (m_1m_2)$ and $m_2 + (m_1m_2)$ are nonzero elements whose product is zero.

Proposition 2.4
Let R be a ring. Then the ring $R[X]$ of polynomials over R is an integral domain if and only if R is an integral domain. In particular, if R is a field, then $R[X]$ is an integral domain.

PROOF: Because R is a subring of $R[X]$, it is an integral domain if $R[X]$ is an integral domain.

Suppose R is an integral domain. Let $\Sigma_{i\in\mathbb{N}} r_iX^i$ and $\Sigma_{i\in\mathbb{N}} r'_iX^i$ be two nonzero elements in $R[X]$. Then there are nonzero integers n_1 and n_2 such that $r_{n_1}\neq 0$ while $r_i=0$ for all $i>n_1$ and $r'_{n_2}\neq 0$ while $r'_i=0$ for all $i>n_2$. From these remarks it follows that the product

$$\left(\sum_{i\in\mathbb{N}} r_iX^i\right)\left(\sum_{i\in\mathbb{N}} r'_iX^i\right)=\sum_{i\in\mathbb{N}}\left(\sum_{k=0}^{i} r_kr'_{i-k}\right)X^i$$

is not zero. For the coefficient of $X^{n_1+n_2}$ in $\Sigma_{i\in\mathbb{N}}(\Sigma_{k=0}^{i} r_kr'_{i-k})X^i$ is $\Sigma_{k=0}^{n_1+n_2} r_kr'_{n_1+n_2-k}$ which under our hypothesis, equals $r_{n_1}r'_{n_2}$. But $r_{n_1}r'_{n_2}$ is not zero because r_{n_1} and r'_{n_2} are nonzero elements in the integral domain R.

In the proof of the last proposition we made implicit use of the notion of the degree of a polynomial. Before going on with our next example, we give an explicit formulation of this useful notion.

Definition
Let $R[X]$ be the ring of polynomials over a ring R. The **degree** of a nonzero polynomial $\Sigma_{i\in\mathbb{N}} r_iX^i$ is defined to be the largest nonnegative integer i such that $r_i\neq 0$. We shall usually denote the degree of a nonzero polynomial Σr_iX^i by $\deg(\Sigma_{i\in\mathbb{N}} r_iX^i)$. If the degree of a polynomial $\Sigma_{i\in\mathbb{N}} r_iX^i$ is n, then r_n is called the **leading coefficient** of the polynomial.

The argument we just used to show that if R is an integral domain then so is $R[X]$ can also be used to establish the following.

Basic Property 2.5
Let $f(X)=\Sigma_{i\in\mathbb{N}} r_iX^i$ and $g(X)=\Sigma_{i\in\mathbb{N}} s_iX^i$ be two nonzero polynomials in $R[X]$ of degrees m and n, respectively. Then:

(a) $f(X)$ is in the subring R of $R[X]$ if and only if $\deg(f(X))=0$.
(b) If $f(X)g(X)\neq 0$, then

$$\deg(f(X)g(X))\leq \deg(f(X))+\deg(g(X))$$

(c) $\deg(f(X)g(X))=\deg(f(X))+\deg(g(X))$ if and only if $r_m\cdot s_n\neq 0$ where r_m is the leading coefficient of $f(X)$ and s_n is the leading coefficient of $g(X)$. Hence:
(d) If $\deg(f(X)g(X))=\deg(f(X))+\deg(g(X))$, then the leading coefficient of $f(X)g(X)$ is the product of the leading coefficients of $f(X)$ and $g(X)$.
(e) If R is an integral domain, then $\deg(f(X)g(X))=\deg(f(X))+\deg(g(X))$.

Proposition 2.6
Let $R[X]$ be the ring of polynomials over the integral domain R. Then $U(R[X])=U(R)$. In particular, if R is a field, then $U(R[X])$ is precisely the set of nonzero elements of R.

PROOF: Because R is a subring of $R[X]$, it follows trivially that $U(R)\subset U(R[X])$. Suppose now that $f(X)=\Sigma_{i\in\mathbb{N}} r_iX^i$ is a unit in $R[X]$. Then there is a $g(X)$ in $R[X]$ such that $f(X)g(X)=1$. Because R is an integral domain, we know that $\deg(f(X)g(X))=\deg(f(X))+\deg(g(X))$. Because $\deg(1)=0=\deg(f(X))+\deg(g(X))$, it follows that $\deg(f(X))=0$ and $\deg(g(X))=0$. Hence,

$f(X)$ and $g(X)$ are in R, which means that $f(X)$ is in $U(R)$. Therefore, $U(R[X]) \subset U(R)$ which completes the proof of the fact that $U(R) = U(R[X])$.

Because a field is an integral domain, we know that every subring of a field is an integral domain. We now show that if R is an integral domain, then R is a subring of a field. We do this by constructing a particular field containing R called the field of quotients of R for each integral domain R. Not only does the field of quotients of an integral domain show that every integral domain is a subring of a field, but it is also a very useful tool for studying the integral domain itself. In particular, as we shall see later on, the field of quotients of an integral domain R is useful in studying unique factorization domains. The reader should observe that the following construction of the field of quotients of an arbitrary integral domain is modeled on the construction of ordinary rational numbers from the ring of integers as well as our construction of the ring of all integers $\mathbf{Z}$ from the nonnegative integers $\mathbf{N}$.

We recall that the field of rational numbers $\mathbf{Q}$ consists of fractions n_1/n_2 where n_1 and n_2 are integers with $n_2 \neq 0$, subject to the condition that two fractions n_1/n_2 and n_1'/n_2' are equal if and only if $n_1 n_2' = n_2 n_1'$. Moreover, the addition and multiplication in $\mathbf{Q}$ are given by the formulas

$$\frac{n_1}{n_2} + \frac{n_1'}{n_2'} = \frac{n_1 n_2' + n_2 n_1'}{n_2 n_2'}$$

$$\frac{n_1}{n_2} \cdot \frac{n_1'}{n_2'} = \frac{n_1 n_1'}{n_2 n_2'}$$

We now generalize this construction to an arbitrary integral domain.

Suppose R is an integral domain. Then the subset $S = R - \{0\}$ of R is a submonoid of the multiplicative monoid of R.

Consider the addition and multiplication on $R \times S$ given by

$$(r, s) + (r', s') = (rs' + r's, ss')$$

$$(r, s)(r', s') = (rr', ss')$$

It is easy to see that $R \times S$ is a commutative monoid under addition with identity $(0, 1)$ and a commutative monoid under multiplication with identity $(1, 1)$. The multiplication also distributes over addition so that $R \times S$ is almost a commutative ring. The only way it fails to be a ring is that elements do not, in general, have additive inverses.

From our experience with the integers and rational numbers, it seems reasonable to consider the following relation I on $R \times S$. Namely, $(r_1, s_1) I (r_2, s_2)$ if $r_1 s_2 = r_2 s_1$. It is a routine matter to check that I is an equivalence relation not only on $R \times S$ considered as a set, but also on the additive and multiplicative monoid structures of $R \times S$. Therefore, $R \times S/I$ has induced additive and multiplicative monoid structures having the following properties:

(a) Under addition, $R \times S/I$ is an abelian group whose zero element is $k(0, 1)$ where $k : R \times S \to R \times S/I$ is the canonical surjective map.
(b) Multiplication distributes over addition in $R \times S/I$; hence:
(c) $R \times S/I$ is a commutative ring.

We denote the element $k(r, s)$ in $R \times S/I$ by r/s. It is easily checked that the elements r/s of $R \times S/I$ have the following familiar properties:

(a) $r/s = 0$ if and only if $r = 0$.
(b) $r/s = r'/s'$ if and only if $rs' = r's$.
(c) $r/s + r'/s' = (rs' + r's)/ss'$.
(d) $r/s \cdot r'/s' = rr'/ss'$.
(e) $1/1$ is the identity.

We now show that these properties imply that $R \times S/I$ is a field. First of all the identity element $1/1$ is not 0. Second, if $r/s \neq 0$, then $r \neq 0$ which means that r is in S and hence s/r is in $R \times S/I$. But $(r/s)(s/r) = rs/rs = 1/1$. Hence, if $r/s \neq 0$, then r/s is a unit with s/r as the inverse. Thus, $R \times S/I$ is a field which we shall denote by $Q(R)$.

We now show that $Q(R)$ is a field which contains R. To do this we consider the map $f: R \to Q(R)$ given by $f(r) = r/1$ for all r in R which is easily seen to be an injective morphism of rings, since Ker $f = 0$. Thus, we can view R as a subring of $Q(R)$ by identifying the element $r/1$ in $Q(R)$ with the element r in R for each r in R. From now on we will consider R a subring of $Q(R)$ by means of this identification.

Definition
For each integral domain R, the field $Q(R)$ containing the ring R as a subring is called the **field of quotients** of R. If $R = \mathbf{Z}$, then $Q(\mathbf{Z})$ is denoted more simply by $\mathbf{Q}$ and is called the field of **rational numbers**.

Proposition 2.7
Let R be an integral domain.

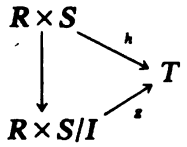
(a) Let T be an arbitrary ring (not necessarily commutative) with $1 \neq 0$. If $f: R \to T$ is a morphism of rings such that $f(r)$ is invertible in T for all nonzero elements r in R [that is, $f(r)$ is in $U(T)$ for all $r \neq 0$], then f is injective and there is a unique morphism of rings $g: Q(R) \to T$ such that $g|R = f$. The morphism $g: Q(R) \to T$ is given by $g(r/s) = f(r)f(s)^{-1}$ where $f(s)^{-1}$ is the inverse of $f(s)$ in T. Finally, $g: Q(R) \to T$ is an injective morphism.
(b) The inclusion morphism $R \to Q(R)$ is an epimorphism in the category of all rings as well as in the category of commutative rings.
(c) $R = Q(R)$ if and only if R is a field.

PROOF: (a) Let $f: R \to T$ be a morphism of rings such that for each $r \neq 0$ we have that $f(r)$ is invertible in T. Since $1 \neq 0$ in T, it follows that $f(r) \neq 0$ if $r \neq 0$.

Because R is an integral domain, we know that $S = R - \{0\}$ is a submonoid of the multiplicative monoid of R. Because $f(r)$ is invertible in T for each nonzero r in R, we have that $f(S)$ is contained in $U(T)$. From the fact that $f: R \to T$ is a morphism of the multiplicative monoid of R to that of T, it follows that $f(S)$ is a commutative submonoid of the group $U(T)$. We have already seen in Chapter 2, Lemma 8.5 that under these circumstances the subset $f(S)f(S)^{-1}$ of $U(T)$ consisting of all products $f(s_1)f(s_2)^{-1}$ is a commutative subgroup of $U(T)$.

The fact that this subgroup of $U(T)$ is commutative enables us to show that

the elements in the image of the map $h: R \times S \to T$ given by $h(r, s) = f(r)f(s)^{-1}$ commute with each other. From this fact it follows that if (r_1, s_1) and (r_2, s_2) are in $R \times S$, then $(r_1, s_1)I(r_2, s_2)$ if and only if $h((r_1, s_1)) = h((r_2, s_2))$. Therefore, there is a unique morphism of rings $g: R \times S/I \to T$ such that the diagram



commutes where $R \times S \to R \times S/I$ is the canonical surjective morphism. This establishes the first part of (a). For we have (1) $R \times S/I = Q(R)$, (2) $g(r/s) = f(r)f(s)^{-1}$ because $g(r/s) = h((r, s)) = f(r)f(s)^{-1}$, and (3) $g(r) = g(r/1) = f(r)$ for all r in R, which shows that $g|R = f$. Because it is obvious that g is an injective morphism, in order to finish the proof of (a) it only remains to show that $g: Q(R) \to T$ is the only morphism from $Q(R)$ to T whose restriction to R is f. This is implied by the fact that the inclusion morphism $R \to Q(R)$ is an epimorphism in the category of rings, a result we now prove.

(b) Suppose $f: Q(R) \to T$ is a morphism of rings. Then $f(r/s) = f(r/1 \cdot 1/s) = f(r/1) \cdot f(1/s)$. Because $1 = s \cdot (1/s) = (1/s) \cdot s$, it follows that $1 = f(1) = f(s) \times f(1/s) = f(1/s)f(s)$. Hence, $f(1/s)$ is the inverse of $f(s)$ in T. This shows that f is completely determined by its restriction to R or, what is the same thing, $R \to Q(R)$ is an epimorphism.

(c) Left as an exercise.

We end this section by pointing out certain consequences of this proposition.

Suppose R is a subring of the integral domain R'. Because $Q(R')$ is a field, the inclusion morphism $R \to Q(R')$ is the composition of the inclusion morphisms $R \to R'$ and $R' \to Q(R')$ and has the property that if r is a nonzero element of R, then r is invertible in $Q(R')$. Thus, by our proposition, there is a unique morphism $g: Q(R) \to Q(R')$ such that $g|R$ is the inclusion morphism $R \to Q(R')$. It is easily checked that the injective morphism $g: Q(R) \to Q(R')$ is given by $g(r/s) = r/s$ for all r and s in R with $s \neq 0$. This enables us to consider $Q(R)$ as a subring of $Q(R')$ by identifying the quotient r/s in $Q(R)$ with the same quotient r/s in $Q(R')$. Hence, from now on we shall use this identification to consider $Q(R)$ a subring of $Q(R')$ whenever R is a subring of R'.

This convention has the following consequence. Suppose R is an integral domain and R' is a subring of $Q(R)$ containing R, that is, $R \subset R' \subset Q(R)$. Then by what we have just agreed upon, we have $Q(R) \subset Q(R') \subset Q(Q(R))$. But $Q(R) = Q(Q(R))$, because $Q(R)$ is a field. Therefore, we have the following.

Proposition 2.8
Suppose R is an integral domain and R' is a subring of $Q(R')$ containing R. Then $Q(R) = Q(R')$.

Finally, the fact that for each integral domain the inclusion morphism $R \to Q(R)$ is an epimorphism in the category of rings shows that in the categories of rings and commutative rings, epimorphisms need not be surjective. To see this,

consider the injective epimorphism $\mathbf{Z} \to \mathbf{Q}$. If this morphism were also surjective, it would be an isomorphism which would mean that $\mathbf{Z}$ is a field. But this is certainly not the case since $U(\mathbf{Z}) = \pm 1$ which is very different from $\mathbf{Z} - \{0\}$.

3. UNIQUE FACTORIZATION DOMAINS

In this section we discuss the general notion of a ring being a unique factorization domain. Because we will be dealing only with rings that are integral domains, we assume once and for all that unless stated to the contrary all the rings in this section are commutative and integral domains.

Probably the best-known example of a unique factorization domain is the ring $\mathbf{Z}$ of integers. We usually say that a nonzero, nonunit number p in $\mathbf{Z}$ is a prime if ± 1 and $\pm p$ are the only integers dividing p. The prime numbers in $\mathbf{Z}$ have the following well-known properties which are usually summarized by saying that $\mathbf{Z}$ is a unique factorization domain: (a) Given any nonzero n in $\mathbf{Z}$ different from ± 1, there is a nonempty finite family of prime elements $(p_i)_{i \in I}$ such that $n = \prod_{i \in I} p_i$, the product of the p_i, and (b) if $(p_i)_{i \in I}$ and $(p_j)_{j \in J}$ are two nonempty finite families of prime elements in $\mathbf{Z}$ such that $\prod_{i \in I} p_i = \prod_{j \in J} p_j$, then card (I) = card (I) and there is an isomorphism of sets $f: I \to J$ such that $p_i = u_i p_{f(i)}$ where u_i is a unit in $\mathbf{Z}$, that is, $u_i = \pm 1$, for all i in I.

Using this description of the fact that the ring $\mathbf{Z}$ is a unique factorization domain as a starting point, it is natural to consider the following conditions on an arbitrary integral domain R as a description of when such a ring should generally be considered a unique factorization domain:

There is a set $\mathcal{P}$ of nonzero elements of R which are not units satisfying:

(i) If p is in $\mathcal{P}$, then up is in $\mathcal{P}$ for all units u in R.

(ii) If n is a nonzero element of R which is not a unit, then there is a finite nonempty family $(p_i)_{i \in I}$ of elements in $\mathcal{P}$ such that $n = \prod_{i \in I} p_i$.

(iii) If $(p_i)_{i \in I}$ and $(p_j)_{j \in J}$ are two nonempty finite families of elements in P such that $\prod_{i \in I} p_i = \prod_{j \in J} p_j$, then card (I) = card (J) and there is an isomorphism of sets $f: I \to J$ such that $p_i = u_i p_{f(i)}$ where u_i is a unit in R for all i in I.

It is interesting to note that the above conditions on the subset $\mathcal{P}$ of R completely determines the subset $\mathcal{P}$ as we see from the following.

Proposition 3.1
Let $\mathcal{P}$ be a subset of a ring R consisting of nonzero, noninvertible elements of R which satisfies conditions (i), (ii), and (iii) just given. Then the following statements are equivalent for a nonzero, noninvertible element r in R:

(a) r is in $\mathcal{P}$.
(b) If $x|r$, then x is either a unit or $ux = r$ with u a unit in R.
(c) If $r|r_1r_2$ and $r \not| r_1$, then $r|r_2$.

PROOF: (a) implies (b). Suppose r is in $\mathcal{P}$ and $x|r$, that is, $xy = r$ for some y in R. We want to show that either x or y is a unit. We prove this by contradiction.

Assume that neither x nor y is a unit in R. Then we know that there are nonempty finite families $(p_i)_{i\in I}$ and $(p_j)_{j\in J}$ of elements in $\mathcal{P}$ such that $x = \prod_{i\in I} p_i$ and $y = \prod_{j\in J} p_j$. Because $xy = r$ and $xy = (\prod_{i\in I} p_i)(\prod_{j\in J} p_j)$, we have $r = \prod_{i\in I} p_i \prod_{j\in J} p_j$. From condition (iii) (as well as from the fact that $r \in \mathcal{P}$), we see that $1 = \text{card}(I) + \text{card}(J)$. This means that either card (I) or card (J) must be zero, and this is a contradiction.

Because (b) obviously implies (a), the equivalence of (a) and (b) is established.

(a) implies (c). Assume r is in $\mathcal{P}$ and suppose $r|r_1r_2$, that is, $rr' = r_1r_2$ for some r' in R. We want to show that r then divides r_1 or r_2. If either r_1 or r_2 is zero or a unit, the result is trivial, so we can assume that neither r_1 nor r_2 is zero or a unit. Let $(p_i)_{i\in I}$ and $(p_j)_{j\in J}$ be nonempty finite families of elements in $\mathcal{P}$ such that $r_1 = \prod_{i\in I} p_i$ and $r_2 = \prod_{j\in J} p_j$. Because $rr' = \prod_{i\in I} p_i \prod_{j\in J} p_j$ and r is in $\mathcal{P}$, it follows that r' is not a unit (why?), and so $r' = \prod_{k\in K} p_k$ where $(p_k)_{k\in K}$ is a nonempty family of elements in $\mathcal{P}$. Hence, $r \prod_{k\in K} p_k = \prod_{i\in I} p_i \prod_{j\in J} p_j$ which by condition (iii) for $\mathcal{P}$ means that $ur = p_l$ for some l in I or J and u a unit in R. Thus, $r|p_l$ and, because $p_l|r_1$ or $p_l|r_2$ depending on whether l is in I or J, it follows that $r|r_1$ or $r|r_2$. Therefore, if r is in $\mathcal{P}$ and $r|r_1r_2$, then $r|r_1$ or $r|r_2$.

(c) implies (a). Suppose a nonzero, noninvertible element r in R has the property that if $r|r_1r_2$, then $r|r_1$ or $r|r_2$. It is not difficult to show by induction that this condition implies that if $r|r_1r_2\cdots r_n$, then $r|r_i$ for some $1 \le i \le n$. Because $r \neq 0$ and is not a unit, we know that there is a nonempty finite family $(p_i)_{i\in I}$ of elements in $\mathcal{P}$ such that $r = \prod_{i\in I} p_i$. Thus, $r|p_i$ for some i in I. Because p_i is in $\mathcal{P}$, we know by the equivalence of (a) and (b) that $r = up_i$ with u a unit in R, because $r|p_i$ and r is not a unit. From this it follows that r is in $\mathcal{P}$, because p_i is in $\mathcal{P}$ and thus up_i is in $\mathcal{P}$ for any unit u in R.

Because elements in a ring R satisfying either (b) or (c) in the above proposition play an important role in studying unique factorization domains, we make the following definition.

Definitions

Let R be a commutative ring (not necessarily an integral domain), and let r be a nonzero, noninvertible element of R. Then:

(a) r is said to be **irreducible** if whenever $r = r_1r_2$, either r_1 or r_2 is a unit.
(b) r is said to be a **prime** element if whenever $r|r_1r_2$ and $r \nmid r_1$, then $r|r_2$.

We now list some easily verified properties.

Basic Properties 3.2

Let R be an integral domain.

(a) A nonzero, noninvertible element r in R is irreducible if and only if R is the only principal ideal of R containing (r) properly [that is, R is the only principal ideal different from (r) which contains (r)]. Hence, r is irreducible if and only if ur is irreducible for all units u in R.

(b) A nonzero element r in R is a prime element in R if the ring R/Rr is an integral domain. Hence, r is a prime element if and only if ur is prime for all units u in R.

(c) Suppose r is a prime element in R. Then:
  (i) r is an irreducible element of R.
  (ii) Suppose $r_1, \ldots, r_n$ is a finite set of prime elements in R and $r \mid \prod_{i=1}^{n} r_i$. Then $r = ur_i$ for some $i = 1, \ldots, n$ and some unit u in R.

(d) Let $(r_i)_{i \in I}$ and $(r_j)_{j \in J}$ be two nonempty finite families of prime elements in R. If $\prod_{i \in I} r_i = \prod_{j \in J} r_j$, then card (I) = card (J) and there is an isomorphism of sets $f : I \to J$ such that $r_i = u_i r_{f(i)}$ with u_i a unit in R for all i in I.

The reader should observe that although we have shown that every prime element in an integral domain is irreducible, we have not claimed that every irreducible element is prime. An example is given in the exercises which shows that irreducible elements need not be prime. Later on we explain what additional property an irreducible element must have in order to be a prime element.

Summarizing our discussion so far, we have the following.

Proposition 3.3
Let R be an integral domain. Then every nonzero element which is not a unit in R can be written as a finite product of prime elements if and only if there exists a subset $\mathcal{P}$ of R satisfying:

(a) If r is in $\mathcal{P}$, then:
  (i) $r \neq 0$ and r is not a unit.
  (ii) ur is in $\mathcal{P}$ for each unit u in R.

(b) Every nonzero, noninvertible element in R can be written as a finite product of elements in $\mathcal{P}$.

(c) If $(r_i)_{i \in I}$ and $(r_j)_{j \in J}$ are nonempty finite families of elements in $\mathcal{P}$ such that $\prod_{i \in I} r_i = \prod_{j \in J} r_j$, then card (I) = card (J) and there is an isomorphism of sets $f : I \to J$ such that $r_i = u_i r_{f(i)}$ with u_i a unit in R for all i in I.

Further, if it is true that every nonzero, nonunit element of R can be written as a finite product of prime elements, then the set $\mathcal{P}$ is precisely the set of prime elements in R. This result suggests the following.

Definition
An integral domain R is a **unique factorization domain** if every nonzero, noninvertible element in R can be written as a finite product of prime elements. We denote the fact that R is a unique factorization domain by writing R is a UFD.

4. DIVISIBILITY IN UFD'S

Having generalized the notion of unique factorization domain from the ring $\mathbf{Z}$ of integers to arbitrary integral domains, we now show that some other familiar notions concerning the divisibility of integers can also be generalized to arbitrary in-

tegral domains. We start with the notions of the greatest common divisor and least common multiple of a finite nonempty family of integers. The reader should have no difficulty in convincing himself that the following definitions for arbitrary integral domains give the familiar concepts when specialized to **Z**.

Definition
Let $r_1, \ldots, r_n$ be a finite, nonempty set of elements in an integral domain R.

(a) An element r in R is said to be a **greatest common divisor** for the set $r_1, \ldots, r_n$ if and only if:
  (i) $r|r_i$ for all $i = 1, \ldots, n$.
  (ii) If $x|r_i$ for all $i = 1, \ldots, n$, then $x|r$. We shall denote the fact that r is a greatest common divisor of $r_1, \ldots, r_n$ by writing $r = \gcd[r_1, \ldots, r_n]$.

(b) An element r in R is said to be a **least common multiple** of $r_1, \ldots, r_n$ if and only if:
  (i) $r_i|r$ for all $i = 1, \ldots, n$.
  (ii) If $r_i|x$ for all $i = 1, \ldots, n$, then $r|x$. We shall denote the fact that r is a least common multiple of $r_1, \ldots, r_n$ by writing $r = \operatorname{lcm}[r_1, \ldots, r_n]$.

Before discussing these ideas further, it is convenient to introduce the notion of the ideal generated by a family $\{x_i\}_{i \in I}$ of elements in an arbitrary commutative ring R (that is, R need not be an integral domain). It is easily seen that if $\{x_i\}_{i \in I}$ is a family of elements in R, then the set of all elements of the form $\Sigma_{i \in I} r_i x_i$, where $\{r_i\}_{i \in I}$ is an almost zero family of elements in R, is an ideal in R containing the element x_i for each $i \in I$.

Definition
Suppose $\{x_i\}_{i \in I}$ is a family of elements in an arbitrary commutative ring R. Then the ideal consisting of all elements of the form $\Sigma_{i \in I} r_i x_i$, where $\{r_i\}_{i \in I}$ is an almost zero family of elements in R, is called the **ideal generated by the family** $\{x_i\}_{i \in I}$ and is denoted by $(x_i)_{i \in I}$ or $\Sigma_{i \in I} R x_i$.

It is left to the reader to establish the following characterization of the ideal generated by a family of elements in a ring.

Basic Property 4.1
Suppose $\{x_i\}_{i \in I}$ is a family of elements in the arbitrary commutative ring R. Then an ideal J in R is the ideal generated by $\{x_i\}_{i \in I}$ if and only if:

(a) x_i is in J for each i in I.
(b) If J' is another ideal containing x_i for all i in I, then $J' \supset J$.

Returning to our discussion of greatest common divisors and least common multiples, the reader should have no difficulty establishing the following.

Basic Properties 4.2
Let $r_1, \ldots, r_n$ be a finite number of elements in a ring R.

(a) An element r is a $\gcd[r_1, \ldots, r_n]$ if and only if Rr is the smallest principal

ideal containing the ideal generated by $r_1, \ldots, r_n$. More specifically:

(i) $Rr \supset \sum_{i=1}^{n} Rr_i$.

(ii) If $Rx \supset \sum_{i=1}^{n} Rr_i$, then $Rx \supset Rr$.

(b) If r is a gcd $[r_1, \ldots, r_n]$, then r' is a gcd $[r_1, \ldots, r_n]$ if and only if $Rr = Rr'$.

(c) The set $r_1, \ldots, r_n$ has a greatest common divisor if and only if there is a principal ideal I such that:

(i) $I \supset \sum_{i=1}^{n} Rr_i$.

(ii) If J is another principal ideal such that $J \supset \sum_{i=1}^{n} Rr_i$, then $J \supset I$.

Moreover, if I_1 and I_2 are two principal ideals satisfying conditions (i) and (ii), then $I_1 = I_2$. Hence, two elements x_1 and x_2 are both gcd$[r_1, \ldots, r_n]$ if and only if x_1 is a gcd$[r_1, \ldots, r_n]$ and $x_1 = ux_2$ where u is a unit of R.

(d) An element $r = \text{lcm}[r_1, \ldots, r_n]$ if and only if $(r) = (r_1) \cap (r_2) \cap \cdots \cap (r_n)$.

(e) If $x_1 = \text{lcm}[r_1, \ldots, r_n]$, then $x_2 = \text{lcm}[r_1, \ldots, r_n]$ if and only if $Rx_1 = Rx_2$. Hence, x_1 and x_2 are both lcm$[r_1, \ldots, r_n]$ if and only if x_1 is a lcm$[r_1, \ldots, r_n]$ and $x_1 = ux_2$ where u is a unit of R.

(f) The set $r_1, \ldots, r_n$ has a least common multiple if and only if $(r_1) \cap \cdots \cap (r_n)$ is a principal ideal.

(g) Every nonempty finite family of elements in R has a least common multiple (greatest common divisor) if and only if every pair of elements in R has a least common multiple (greatest common divisor).

In arbitrary integral domains, although not every pair of nonzero elements need have a least common multiple or greatest common divisor (see the exercises for examples), all unique factorization domains do have this property as we now proceed to show.

Suppose R is a UFD and that $\mathcal{P}$ is the set of prime elements of R. We define a relation A on $\mathcal{P}$ by setting $p_1 A p_2$ if and only if there is a unit u in R such that $p_1 = up_2$. The reader can easily check that A is an equivalence relation on $\mathcal{P}$. Let I be the set of equivalent classes of A. Because each element i in I is a nonempty subset of $\mathcal{P}$, we may choose, for each i in I, an element p_i of $\mathcal{P}$ such that p_i is in i.

Definition

Let R be a UFD, $\mathcal{P}$ the set of prime elements of R, and I the set of equivalence classes of the equivalence relation A on $\mathcal{P}$ defined by setting $p_1 A p_2$ if and only if there is a unit a in R such that $p_1 = up_2$. A family of prime elements $\{p_i\}_{i \in I}$ of R is a **representative family of primes** if for each i in I, the element p_i is in i.

Basic Properties 4.3

Let R be a UFD and $\{p_i\}_{i \in I}$ a representative family of primes.

(a) If p is any prime element of R, then there is a unique i in I such that $p = up_i$ where u is a unit of R.

(b) If p_i and p_j are elements of the representative family $\{p_i\}_{i \in I}$, then $p_i = up_j$ for some unit u in R if and only if $i = j$ and $u = 1$.

(c) If r is a nonzero element of R, then there exists a unique almost zero family $\{n_i(r)\}_{i\in I}$ of elements of $\mathbf{N}$ such that $r = u \prod_{i\in I} p_i^{n_i(r)}$ where u is a unit in R. Moreover, if $\{q_i\}_{i\in I}$ is another representative family of prime elements in R, then $r = u' \prod_{i\in I} q_i^{n_i(r)}$ where u' is a unit in R. Hence, the almost zero family $\{n_i(r)\}_{i\in I}$ of $\mathbf{N}$ depends only on the element r because it is independent of the particular choice of representative family of primes used to obtain it.

This last property suggests the following.

Definition
Let R be a UFD and r a nonzero element of R. An almost zero family $\{n_i(r)\}_{i\in I}$ of elements of $\mathbf{N}$ is called the **prime exponents** for r if for some, and hence any, representative family of primes $\{p_i\}_{i\in I}$ for R we have $r = u \prod_{i\in I} p_i^{n_i(r)}$ where u is a unit in R.

Basic Properties 4.4
Let R be a UFD and let r_1 and r_2 be nonzero elements of R.

(a) r_1 is a unit in R if and only if $n_i(r) = 0$ for all i in I where $\{n_i(r)\}_{i\in I}$ is the prime exponents of R.
(b) $\{n_i(r_1 r_2)\}_{i\in I} = \{n_i(r_1) + n_i(r_2)\}_{i\in I}$.
(c) $r_1 | r_2$ if and only if $n_i(r_1) \le n_i(r_2)$ for all i in I.
(d) Given any almost zero family $\{n_i\}_{i\in I}$ of elements in $\mathbf{N}$, there is a nonzero element r in R such that $n_i(r) = n_i$ for all i in I.
(e) $r_1 = ur_2$ where u is a unit in R or equivalently $(r_1) = (r_2)$ if and only if $n_i(r_1) = n_i(r_2)$ for all i in I.

Using prime exponents for nonzero elements of a UFD, we can prove the following.

Proposition 4.5
Let $r_1, \ldots, r_t$ be nonzero elements of a ring R which is a UFD.

(a) For each i in I, let $M_i = \max(n_i(r_1), \ldots, n_i(r_t))$. An element r in R is a $\operatorname{lcm}[r_1, \ldots, r_t]$ if and only if $n_i(r) = M_i$ for all i in I. Hence, the set of elements $r_1, \ldots, r_t$ has a least common multiple in R.
(b) For each i in I, let $m_i = \min(n_i(r_1), \ldots, n_i(r_t))$. An element r in R is a $\gcd[r_1, \ldots, r_t]$ if and only if $n_i(r) = m_i$ for all i in I. Hence, the set of elements $r_1, \ldots, r_t$ has a greatest common divisor in R.

PROOF: Obviously $\{M_i\}_{i\in I}$ and $\{m_i\}_{i\in I}$ are almost zero families of elements in $\mathbf{N}$. Thus, by Properties 4.4 we know that there are elements r and r' in R such that $n_i(r) = M_i$ for all i in I and $n_i = (r') = m_i$ for all i in I.

(a) Suppose r is an element of R such that $n_i(r) = M_i$ for all i in I. Because $n_i(r) \ge n_i(r_k)$ for each $k = 1, \ldots, t$ and all i in I, it follows from Properties 4.4 that each $r_k | r$ for $k = 1, \ldots, t$. Moreover, if x is a nonzero element of R such that $r_k | x$ for each $k = 1, \ldots, t$, then again by 4.4, we know that $n_i(r_k) \le n_i(x)$ for each

$k = 1, \ldots, t$ and all i in I. Therefore, $n_i(x) \geq M_i = n_i(r)$ for all i in I. Thus, $r|x$, which shows that r is a lcm $[r_1, \ldots, r_t]$.

(b) Left as an exercise because it is entirely analogous to that for part (a).

Because in all unique factorization domains every pair of nonzero elements has a least common multiple, it is natural to ask how close an integral domain is to being a UFD if every pair of nonzero elements in it has a least common multiple. Although this condition does not quite guarantee that an integral domain is a UFD, it does imply that every irreducible element in the ring is a prime element. After establishing this fact, we will discuss what further conditions the ring must satisfy in order to guarantee that it is indeed a unique factorization domain.

We recall the following.

Definition

Two nonzero elements r_1, r_2 in a ring R are said to be **relatively prime** if $\gcd(r_1, r_2) = 1$, that is, R is the only principal ideal containing r_1 and r_2. The following are the properties of relatively prime elements that we shall need.

Basic Properties 4.6

Let x and y be nonzero elements in R.

(a) Assume x is not a unit. Then x is irreducible if and only if y is relatively prime to x whenever x does not divide y.

(b) Suppose x and y are relatively prime and x and y have a lcm, that is, $(x) \cap (y)$ is principal. Then $xy = \text{lcm}\,[x, y]$ or equivalently $(x) \cap (y) = (xy)$.

PROOF: **(a)** See Basic Properties 3.2.

(b) Suppose x and y are relatively prime. Because xy is obviously in $(x) \cap (y)$, to show that $(xy) = (x) \cap (y)$ it suffices to show that $x|z$ and $y|z$ implies $xy|z$.

By assumption, x and y have a lcm which we will denote by s. Because $x|xy$ and $y|xy$, we have that $s|xy$, that is, $ts = xy$. On the other hand, $s = t_1 x$ and $s = t_2 y$. Therefore, $xy = ts = tt_1 x = tt_2 y$, which implies that $t|x$ and $t|y$. But this means that t is a unit in R because $\gcd[x, y] = 1$. Therefore, $s = t^{-1}xy$ which shows that $xy = \text{lcm}\,[x, y]$ if x and y are relatively prime elements in R which have a lcm.

As a consequence of these observations we have the following characterization of prime elements.

Proposition 4.7

Let x be an element of the ring R which is neither zero nor a unit. Then x is a prime element if and only if x is irreducible and the ideal $(x) \cap (y)$ is principal for all y in R, that is, the pair x, y have a lcm for all y in R.

PROOF: Suppose x is a prime element in R. We want to show that x is irreducible and $(x) \cap (y)$ is principal for all y in R. Because we have already shown that every prime element is irreducible, it only remains to establish the second condition.

Let y be in R. If y is in (x), then $(y) \subset (x)$ and so $(x) \cap (y) = (y)$ which is principal.

Suppose now that y is not in (x). Let z be in $(x) \cap (y)$, that is, $x|z$ and $y|z$. Then $z = yv$ and thus $x|yv$. Because x is a prime and $x \nmid y$, the fact that $x|yv$ implies that $x|v$, that is, $xw = v$. Therefore, $z = yv = yxw$ which means that $xy|z$. Hence, we have shown that if y is not in (x), then $(x) \cap (y) \subset (xy)$. Since $(xy) \subset (x) \cap (y)$ we have that $(x) \cap (y) = (xy)$ if y is not in (x). Therefore, if x is a prime, we have shown that $(x) \cap (y)$ is a principal ideal for all y in R, which completes the proof of the proposition in one direction.

Suppose now that x is irreducible and $(x) \cap (y)$ is principal for all y in R. We want to show this implies that x is a prime. Assume that $x|yz$ and $x \nmid y$. Since x is irreducible, the fact that y is not in (x) implies that x and y are relatively prime (see Basic Properties 4.6). But we have also seen (Basic Properties 4.6) that if x and y are relatively prime and $(x) \cap (y)$ is principal, then $(xy) = (x) \cap (y)$. This implies that $xy|yz$ since yz is in (x) by assumption and in (y) by definition. The fact that $xy|yz$ implies that $x|z$. Hence, if $x|yz$ and $x \nmid y$, then $x|z$, which means that x is a prime element in R. This completes the proof of the proposition.

As an immediate consequence of this characterization of prime elements we have the following.

Corollary 4.8
If the ring R has the property that the intersection of any two principal ideals is principal, then every irreducible element in R is prime.

Summarizing our discussion so far we have the following.

Proposition 4.9
A ring R is a unique factorization domain if and only if every nonzero, nonunit element can be written as a finite product of irreducible elements and the intersection of any two principal ideals is principal.

We shall present another version of this description of unique factorization domains which is given solely in terms of the structure of the principal ideals in the ring. This new description will be used in the next section to prove that principal ideal domains are unique factorizatien domains. To do this, we develop the general notion of the ascending chain condition for a set of subsets of a set because our new description of unique factorization domains utilizes this extremely important general concept.

Proposition 4.10
Let $\mathcal{S}$ be a nonempty set of subsets of a set X. Then the following statements about $\mathcal{S}$ are equivalent:

(a) If

$$X_1 \subset X_2 \subset X_3 \subset \cdots \subset X_i \subset \cdots$$

is any ascending chain of subsets of X in $\mathcal{S}$, then there is an integer n such that $X_i = X_n$ for all $i \geq n$.

(b) Every nonempty subset $\mathcal{T}$ of $\mathcal{S}$ contains a maximal element; that is, there is an element X_0 in $\mathcal{T}$ with the property that if X_i is in $\mathcal{T}$ and $X_i \supset X_0$, then $X_i = X_0$.

PROOF: (**a**) implies (**b**). Suppose $\mathcal{T}$ is a nonempty subset of $\mathcal{S}$ which has no maximal element. Then given any element X_i in $\mathcal{T}$ there is an X_j in $\mathcal{T}$ such that $X_i \subset X_j$ but $X_i \neq X_j$. Hence we can construct by induction an ascending chain of distinct elements in $\mathcal{T}$ as follows. Let X_1 be an arbitrary element of $\mathcal{T}$. Suppose we have defined X_n. Since $\mathcal{T}$ has no maximal element, there are elements in $\mathcal{T}$ distinct from X_n but which contain X_n. Define X_{n+1} to be one of these elements in $\mathcal{T}$. In this way we obtain an ascending chain of distinct elements of $\mathcal{T}$

$$X_1 \subset X_2 \subset \cdots \subset X_n \subset \cdots$$

which contradicts the hypothesis of (a). Hence, if $\mathcal{S}$ satisfies condition (a), then each nonempty subset $\mathcal{T}$ of $\mathcal{S}$ must have a maximal element, which shows that (a) implies (b).

(**b**) implies (**a**). Suppose every nonempty subset $\mathcal{T}$ of $\mathcal{S}$ has a maximal element. Let $X_1 \subset X_2 \subset \cdots \subset X_n \subset \cdots$ be an ascending chain of elements in $\mathcal{S}$ and let $\mathcal{T}$ be the subset of $\mathcal{S}$ consisting of the elements $X_1, X_2, \ldots, X_n, \ldots$ of $\mathcal{S}$. Then $\mathcal{S}$ has a maximal element, say X_n. Obviously, $X_i = X_n$ for all $i \geq n$. This shows that (b) implies (a).

Because subsets $\mathcal{S}$ of 2^X satisfying either of the above equivalent conditions play an important role throughout all of algebra we make the following definition.

Definition

Let X be a set. A nonempty subset $\mathcal{S}$ of 2^X is said to be **noetherian**, or to satisfy the **ascending chain condition**, if every nonempty subset of $\mathcal{S}$ contains a maximal element or, equivalently, given any ascending chain

$$X_1 \subset X_2 \subset \cdots \subset X_n \subset \cdots$$

of elements in $\mathcal{S}$ there is an n such that $X_i = X_n$ for all $i \geq n$.

We now state and prove our final result of this section.

Theorem 4.11

A ring R is a unique factorization domain if and only if the set of principal ideals $PI(R)$ of R satisfies:

(**a**) $PI(R)$ is noetherian.
(**b**) If I_1 and I_2 are in $PI(R)$, then $I_1 \cap I_2$ is also in $PI(R)$.

PROOF: Suppose R is a unique factorization domain. Since we have already seen that for UFD's the intersection of two principal ideals is again a principal ideal, we only have to show that $PI(R)$ is noetherian. Suppose

$$(r_1) \subset (r_2) \subset \cdots \subset (r_k) \subset \cdots$$

is an ascending chain of principal ideals. We can assume without loss of generality that none of the r_k are zero. The fact that $(r_1) \subset (r_2) \subset \cdots \subset (r_k) \subset \cdots$ is equivalent to

$$n_i(r_1) \geq n_i(r_2) \geq \cdots \geq n_i(r_k) \geq \cdots$$

for all i where $\{n_i(r_k)\}_{i \in I}$ are the prime exponents for r_k and $k = 1, 2, \ldots$. Hence, for each i in I, there are integers m_i' such that $n_i(r_k) = n_i(r_{m_i'})$ for all $k \geq m_i'$ [remember

all the $n_i(r_k) \geq 0$]. For each i let m_i be the smallest integer satisfying this condition. Because $n_i(r_k) = 0$ for all but a finite number of i in I, only a finite number of the m_i are different from 1 and thus the set of m_i has a maximum which we denote by m. Hence, $n_i(r_k) = n_i(r_m)$ for all i in I and all $k \geq m$, which means that $(r_k) = (r_m)$ for all $k \geq m$ (see Basic Properties 4.4). This shows that if R is a unique factorization domain, then $PI(R)$ is noetherian which finishes the proof that R satisfies (a) and (b).

Suppose R is a ring which has the property that the intersection of any two principal ideals is principal and $PI(R)$ is noetherian. We want to show this implies R is a unique factorization domain. Because we have already shown that a ring R in which the intersection of two principal ideals in principal is a unique factorization domain provided every nonzero, nonunit element in R can be written as a finite product of irreducible elements, it suffices to show that if $PI(R)$ is noetherian, then every nonzero, nonunit in R can be written as a finite product of irreducible elements.

Let T be the subset of $PI(R)$ consisting of all principal ideals $(x) \neq R$ such that x is not a finite product of irreducible elements. Suppose $\mathcal{T}$ is not empty. Because $PI(R)$ is noetherian, we know that $\mathcal{T}$ has a maximal element (x). Now x is not irreducible because an irreducible element is obviously the finite product of irreducible elements, namely, of one element. Hence, $x = yz$ where neither y nor z is zero or a unit. Therefore, $(y) \supset (x)$ and $(z) \supset (x)$ and both are different from (x). This implies that neither (y) nor (z) is in $\mathcal{T}$ since (x) is a maximal element of $\mathcal{T}$. Therefore, y and z can both be written as a finite product of irreducible elements which implies that $x = yz$ can also be written as a finite product of irreducible elements. But this contradicts the fact that x could not be so written. Therefore, the set $\mathcal{T}$ is empty, which means that every nonzero element in R which is not a unit is a finite product of irreducible elements. This finishes the proof of the theorem.

5. PRINCIPAL IDEAL DOMAINS

In this section we give an introduction to the important type of unique factorization domains known as principal ideal domains. Here we shall be mainly concerned with the ideal theory of such rings. Much later on we shall examine the module theory for principal ideal domains.

Definition
A **principal ideal domain** is an integral domain R which has the property that every ideal in R is principal. We shall often use PID as an abbreviation for principal ideal domain.

Our first concern is to show that every PID is a unique factorization domain. In the last section we showed that an integral domain is a unique factorization domain if the intersection of two principal ideals is a principal ideal and the set of principal ideals is noetherian. Because in a PID all ideals are principal, PID's certainly have the property that the intersection of two principal ideals is a principal

ideal. Therefore, to show that PID's are unique factorization domains, it suffices to show that the set of principal ideals in a PID, namely, the set of all ideals in a PID, is noetherian. This will follow trivially from the following more general considerations which will play a large role in the rest of this book.

Proposition 5.1
Let R be an arbitrary, commutative ring. Then the set of ideals in R is noetherian if and only if every ideal I in R is finitely generated; that is, if I is an ideal in R, then there are a finite number of elements in I which generate I.

PROOF: Suppose every ideal in R is finitely generated. We want to show that this implies that the set of ideals in R is noetherian. Suppose

$$I_1 \subset I_2 \subset \cdots \subset I_n \subset \cdots$$

is an ascending chain of ideals in R. Then it is easily seen that $J = \bigcup_{n \in N} I_n$ is an ideal in R. Because every ideal in R is finitely generated, we know that J is finitely generated. Suppose $x_1, \ldots, x_t$ generate J. Because $J = \bigcup_{n \in \mathbf{N}} I_n$, it follows that each x_j is in $I_{n(x_j)}$ for some $n(x_j)$ in $\mathbf{N}$. Hence, the finite set of integers $n(x_1), \ldots, n(x_t)$ has a maximum m which has the property that each $x_1, \ldots, x_t$ is in I_m. Because $I_m \subset J$ and I_m contains a set of generators for J, it follows that $I_m = J$. This clearly implies that $I_n = I_m$ for all $n \geq m$. Therefore, if every ideal in R is finitely generated, then the set of all ideals in R is noetherian because we have shown that given any ascending chain of ideals in R

$$I_1 \subset I_2 \subset \cdots \subset I_n \subset \cdots$$

there is an integer m such that $I_n = I_m$ for all $n \geq m$.

Suppose now that the set of ideals in R is noetherian. We want to show that each ideal in R is finitely generated. Suppose that this is not the case. Then there is an ideal J in R which is not finitely generated. We define the sequence x_1, $x_2, \ldots, x_n$ of elements in R by induction as follows. Let x_1 be an arbitrary element in J. Suppose we have defined the sequence $x_1, \ldots, x_n$. Because J is not finitely generated, we know that the ideal $(x_1, \ldots, x_n)$ which is contained in J and generated by $x_1, \ldots, x_n$ is not all of J. Define x_{n+1} to be an arbitrary element of J not in $(x_1, \ldots, x_n)$. In this way we obtain a sequence $x_1, x_2, \ldots, x_n, \ldots$ of elements in R with the property that all the ideals in the ascending chain

$$(x_1) \subset (x_1, x_2) \subset \cdots \subset (x_1, \ldots, x_n) \subset \cdots$$

are distinct. Because this contradicts the fact that the set of ideals in the ring R is noetherian, we see that there are no ideals J in R which are not finitely generated. Hence, we have shown that if the set of ideals in R is noetherian, then every ideal in R is finitely generated.

For ease of reference we make the following definition.

Definition
A commutative ring R is said to be a **noetherian ring** if the set of ideals in R is noetherian.

Hence, our previous result can be rephrased as follows: A commutative ring R is noetherian if and only if every ideal in R is finitely generated. As an immediate consequence we have the following.

Corollary 5.2
Every PID is a noetherian ring.

Because this was precisely the missing step in proving that a PID is a unique factorization domain, we also have the following.

Theorem 5.3
Every PID is a unique factorization domain.

We pause now in our general development of PID's in order to give some examples of PID's. The first example we consider is that of the ring $\mathbf{Z}$ of all integers. The fact that $\mathbf{Z}$ is a PID is based on the following well-known proposition.

Proposition 5.4
Let a and b be integers with $b \neq 0$.

Then there exist integers q and r such that:

(a) $a = qb + r$.
(b) $0 \leq |r| < |b|$.

PROOF: We prove this result under the additional hypothesis that $a > 0$ and $b > 0$. The fact that this implies the general result is left as an exercise to the reader.

Since $a > 0$ and $b > 0$, it follows that $(a+1)b > a$. Therefore, the subset N' of $\mathbf{N}$ consisting of those n in $\mathbf{N}$ such that $nb > a$ is not empty. Hence, the fact that $\mathbf{N}$ is well ordered implies that N' has a first element q'. Because $q'b > a > 0$, it follows that $q' \geq 1$ or, equivalently, $q = q' - 1$ is in $\mathbf{N}$. Because $q' = q + 1$ is the smallest integer n in $\mathbf{N}$ such that $nb > a$, it follows that $(q+1)b > a > qb > 0$ or, equivalently, $b > a - qb \geq 0$. Therefore, the pair q and $r = a - qb$ satisfy our desired conditions: $a = qb + r$ and $0 \leq |r| < |b|$.

We now use this to prove the following.

Proposition 5.5
The ring $\mathbf{Z}$ of all integers is a PID.

PROOF: Let I be an ideal of $\mathbf{Z}$. If $I = 0$, then I is certainly principal, so we may assume that $I \neq 0$. Let N' be the subset of $\mathbf{N}$ consisting of all $|x|$ as x runs through all the nonzero elements in I. Because $I \neq 0$, we know that N' is a nonempty subset of $\mathbf{N}$ and hence has a first element of the form $|b| \neq 0$ with b in I. Suppose a is an arbitrary element of I. Then by our previous result we know that there are q and r in $\mathbf{Z}$ such that $a = qb + r$ with $0 \leq |r| < |b|$. Because a and b are in I, it follows that $r = a - bq$ is also in I. Hence, if $r \neq 0$, then $|r| < |b|$ is also in N' which contradicts the fact that $|b|$ is the first element of $\mathbf{N}'$. Thus, $r = 0$ or $a = bq$. Therefore, every element of I is divisible by b which implies that I is the principal ideal $\mathbf{Z}b$. Hence,

every ideal I of $\mathbf{Z}$ is a principal ideal which means that $\mathbf{Z}$ is a principal ideal domain.

A few minutes' thought should suffice to convince the reader that the proof that $\mathbf{Z}$ is a PID which we just gave depends in an essential way on the existence of the absolute value map $|\,|:\mathbf{Z}\to\mathbf{N}$. But the only property of the absolute value actually used in the proof is that if a and b are in $\mathbf{Z}$ with $b\neq 0$, then there are q and r in $\mathbf{Z}$ satisfying $a=qb+r$ where $0\leq|r|<|b|$. These observations suggest the following question: Is an integral domain R a PID if there is a map $f:R-\{0\}\to\mathbf{N}$ satisfying the following condition: Given any a and b in R with $b\neq 0$, there exist q and r in R such that $a=qb+r$ where either $r=0$ or $f(r)<f(b)$? The proof that this is indeed the case is essentially identical to the proof that $\mathbf{Z}$ is a PID. All one has to do is show that if I is a nonzero ideal of R, then (1) there is an element b in I different from zero such that $f(b)\leq f(x)$ for all nonzero x in I, and (2) the defining property of f implies that any such b in I is a generator for I. The details are left to the reader to carry out. We summarize this discussion in the following definition and proposition.

Definition
Let R be an integral domain. A map $f:R-\{0\}\to\mathbf{N}$ is called a **Euclidean function** on R if given a and b in R with $b\neq 0$, there are q and r in R such that $a=qb+r$ and either $r=0$ or $f(r)<f(b)$. An integral domain for which there exists a Euclidean function is called a **Euclidean domain**.

Proposition 5.6
R is a PID if R is a Euclidean domain.

It is obvious from our discussion that the absolute value is a Euclidean function on $\mathbf{Z}$ and so $\mathbf{Z}$ is a Euclidean domain. We now show that if R is a field, then the degree of a polynomial is a Euclidean function on $R[X]$, the ring of polynomials over R, and so $R[X]$ is a Euclidean domain and therefore a PID. This fact is an easy consequence of the following general lemma.

Lemma 5.7
Let R be an arbitrary ring and $b(X)$ a nonzero polynomial in $R[X]$ whose leading coefficient is a unit in R. Then given any polynomial $a(X)$ in $R[X]$, there exist polynomials $q(X)$ and $r(X)$ in $R[X]$ such that $a(X)=q(X)b(X)+r(X)$ where either $r(X)=0$ or $\deg(r(X))<\deg(b(X))$.

PROOF: The result is obvious if $a(X)=0$. So we may suppose $a(X)\neq 0$. If $\deg(a(X))<\deg(b(X))$, then $q(X)=0$ and $r(X)=a(X)$ have our required properties. Suppose now that $m=\deg(a(X))\geq\deg(b(X))=t$. Let $a(X)=\Sigma_{i\in\mathbf{N}}a_iX^i$ and $b(X)=\Sigma_{i\in\mathbf{N}}b_iX^i$. Then $a_m\neq 0$ and $a_i=0$ for all $i>m$ while b_t is a unit in R and $b_i=0$ for all $i>t$. Then it is easily checked that the degree of $a_1(X)$, where $a_1(X)=a(X)-b_t^{-1}a_mb(X)X^{m-t}$, is less than the degree of $a(X)$ if $a_1(X)\neq 0$. Thus, if $\deg(a(X))\geq\deg b(X)$, then there are $q_1(X)$ and $a_1(X)$ in $R[X]$ such that $a(X)=q_1(X)b(X)+a_1(X)$ where either $a_1(X)=0$ or $\deg(a_1(X))<\deg(a(X))$. We leave it

to the reader to show how this result may be used to prove the lemma by induction on $n = \deg(a(X)) - \deg(b(X))$.

This lemma immediately implies the following.

Proposition 5.8
Let R be a field. Then the map $\deg: R[X] - \{0\} \to \mathbf{N}$ is a Euclidean function on the integral domain $R[X]$ and so $R[X]$ is a Euclidean domain and consequently a PID.

With these examples of PID's in mind, we return to our general discussion of PID's. In this connection the following notion is useful.

Definition
An ideal I in an arbitrary commutative ring R is said to be a **maximal ideal** if R is the only ideal of R containing I properly, that is, R is the only ideal different from I containing R.

The following characterization of maximal ideals is very useful.

Basic Property 5.9
Let I be a proper ideal of R (that is, $I \neq R$). Then I is a maximal ideal of R if and only if the ring R/I is a field.

PROOF: We have already seen that a ring is a field if and only if the zero ideal is the only proper ideal in the ring. Hence, the ring R/I is a field if and only if (0) is the only proper ideal in R/I. The fact that $I \neq R$ is equivalent to the fact that (0) is a proper ideal of R. The bijective correspondence between the ideals of R/I and the ideals of R containing I (see Chapter 4, Proposition 4.4), shows that (0) is the only proper ideal of R/I if and only if R is the only ideal of R containing I properly. Therefore, R/I is a field if and only if I is a maximal ideal of R.

A maximal ideal is a special case of a prime ideal which we now define.

Definition
Let R be an arbitrary commutative ring. An ideal I of R is a **prime ideal** of R if R/I is an integral domain.

Basic Properties 5.10
Let R be an arbitrary commutative ring.

(a) An ideal $I \neq R$ is a prime ideal if and only if xy in I implies either x or y is in I.
(b) R is an integral domain if and only if (0) is a prime ideal in R.
(c) An ideal $I \neq R$ is a prime ideal in R if and only if $I_1 I_2 \subset I$ implies either $I_1 \subset I$ or $I_2 \subset I$ for all ideals I_1 and I_2 in I.
(d) If R is an integral domain, then x is a prime element if and only if Rx is a nonzero prime ideal in R.

We now point out the following important characterization of prime elements in a PID.

Proposition 5.11
For an element x in a ring R which is a PID, the following statements are equivalent:

(a) x is irreducible.
(b) x is a prime element.
(c) (x) is a prime ideal.
(d) (x) is a maximal ideal in R.
(e) $R/(x)$ is a field.

PROOF: The equivalence of (a), (b), and (c) as well as the equivalence of (d) and (e) have already been established. We finish the proof by showing that (a) and (d) are equivalent.

By definition, an element x in R is irreducible if and only if R is the only principal ideal of R which contains (x) properly. But all the ideals of R are principal ideals because R is a PID. Hence, x is irreducible if and only if R is the only ideal of R containing (x) properly. Therefore, x is irreducible if and only if (x) is maximal, which is our desired result.

This proposition is very useful in constructing fields of various types. For instance, for each prime p in $\mathbf{Z}$, we have that $\mathbf{Z}/p\mathbf{Z}$ is a field. Because $p = card(\mathbf{Z}/p\mathbf{Z})$, this shows that there are a great many fields with only a finite number of elements. Other examples of how this proposition can be used to construct fields are given in the exercises.

6. FACTOR RINGS OF PID'S

This section is devoted to studying the rings of the form R/I where I is a proper nonzero ideal in a ring R which is a PID.

Proposition 6.1
Let (x) be a proper nonzero ideal in the PID, R. Then:

(a) A maximal ideal (p) in R contains (x) if and only if p is a prime which divides x. Hence;
(b) (x) is contained in only a finite number of maximal ideals of R. A set $(p_1), \ldots, (p_t)$ of distinct maximal ideals of R is precisely the set of all maximal ideals of R containing (x) if and only if $x = u \prod_{i=1}^{t} p_i^{n_i}$ where u is a unit in R and all the $n_i > 0$.
(c) Suppose

$$(r_1) \supset (r_2) \supset \cdots \supset (r_n) \supset \cdots$$

is a descending chain of ideals in R each of which contains the ideal (x). Then there is an integer m such that $(r_k) = (r_m)$ for all $k \geq m$.

PROOF: (a) and (b) are left as exercises to the reader.

(c) Suppose $(r_1) \supset (r_2) \supset \cdots \supset (r_i) \supset \cdots$ is a descending chain of ideals in R all of which contain (x). Letting $\{n_i(r)\}_{i \in I}$ be the family of prime exponents of an

element r in R with respect to a representative family of primes $\{p_i\}_{i\in I}$, we have

$$n_i(x) \geq \cdots \geq n_i(r_{t+1}) \geq n_i(r_t) \geq \cdots \geq n_i(r_1)$$

for all i in I. Hence, for each i in I there is a nonnegative integer, and hence a smallest integer m_i, such that $n_i(r_k) = n_i(r_{m_i})$ for all $k \geq m_i$. Because $n_i(x) = 0$ for all but a finite number of i in I, all but a finite number of $m_i = 0$. Hence, the family $\{m_i\}_{i\in I}$ has a maximum, m, which has the property that $n_i(r_k) = n_i(r_m)$ for all $k \geq m$ and all i in I. This obviously implies our desired result that $(r_k) = (r_m)$ for all $k \geq m$.

Because of the importance in all of algebra of the type of phenomenon described in the last part of this proposition, we make the following definition.

Definition
A set $\mathcal{S}$ of subsets of a set X is said to be **artinian** or to satisfy the **descending chain condition** if given any descending chain

$$X_1 \supset X_2 \supset \cdots \supset X_n \supset \cdots$$

of subsets of X in $\mathcal{S}$ there is an integer m such that $X_i = X_m$ for all $i \geq m$.

A commutative ring R is said to be **artinian** or satisfy the **descending chain condition** if the set of all ideals in R is artinian.

We leave it to the reader to verify the following.

Basic Properties 6.2

(a) Let $\mathcal{S}$ be a set of subsets of a set X. Then $\mathcal{S}$ is artinian if and only if every nonempty subset $\mathcal{T}$ of $\mathcal{S}$ has a minimal element, that is, there is an X_0 in $\mathcal{T}$ such that if X is in $\mathcal{T}$ and $X \subset X_0$, then $X = X_0$.
(b) A ring R is artinian if every nonempty set of ideals has a minimal element.

Using the relationship between the ideals in a ring R containing a fixed ideal I of R and the ideals of R/I, we obtain the following restatement of Proposition 6.1.

Proposition 6.3
Let R be a PID and I a nonzero proper ideal of R. Then the ring R/I has the following properties:

(a) Every ideal in R/I is principal.
(b) R/I has a finite number of maximal ideals.
(c) R/I is an artinian ring,

We now develop two results concerning arbitrary commutative rings which, when applied to factor rings of a PID, will give us the final result of this section.

Lemma 6.4
Let $I_1, \ldots, I_n$ be ideals in an arbitrary commutative ring such that for all $j > 1$ the ideal generated by I_1 and I_j is all of R. Then R is also generated by I_1 and the product $I_2 \cdots I_n$ of the set of ideals $\{I_2, \ldots, I_n\}$.

PROOF: Because R is generated by I_1 and I_j for all $j>1$ we know that for each $j>1$ there are elements r_j in I_1 and s_j in I_j such that $r_j+s_j=1$. Hence, $1=\prod_{j=2}^{n}(r_j+s_j)$. After carrying out the indicated multiplication, we see that all the terms except one in the resulting sum are in I_1 while the remaining one is in $I_2\cdots I_n$. Therefore, 1 is in the ideal generated by I_1 and $I_2\cdots I_n$ from which it follows that I_1 and $I_2\cdots I_n$ generate R.

As our first application of this lemma, we prove the following.

Proposition 6.5
Let $I_1,\ldots,I_n$ be a finite set of ideals in a ring R such that the ideal generated by I_i and I_j is R whenever $i\neq j$. Then $\bigcap_{i=1}^{n} I_i=I_1\cdots I_n$.

PROOF: Because it is obvious that $\bigcap_{i=1}^{n} I_i\supset\prod_{i=1}^{n} I_i$, we only have to show that $\prod I_i\supset\bigcap_{i\in I} I_i$, which we do by induction on n. Suppose $n=2$. Then $1=r_1+r_2$ with r_1 in I_1 and r_2 in I_2. Suppose r is in $I_1\cap I_2$. Then $r=rr_1+rr_2$ where both rr_1 and rr_2 are in the product I_1I_2. Hence, $I_1\cap I_2\subset I_1I_2$ or, equivalently, $I_1\cap I_2=I_1I_2$ if I_1 and I_2 generate R.

Suppose $I_1,\ldots,I_n$ have the property that I_i and I_j generate R whenever $i\neq j$. By the inductive hypothesis we have $I_2\cap\ldots\cap I_n=I_2\ldots I_n$, and so $I_1\cap I_2\cap\ldots\cap I_n=I_1\cap(I_2\ldots I_N)$. Because I_1 and I_j generate R for $j=2,\ldots,n$, we have by our lemma that I_1 and $I_2\ldots I_n$ generate R. Hence, it follows from our inductive hypothesis that $I_1\cap(I_2\ldots I_n)=I_1(I_2\ldots I_n)$. Therefore, $I_1\cap\ldots\cap I_n=I_1\ldots I_n$, which finishes our inductive proof.

Suppose $\{I_k\}_{k\in K}$ is a family of ideals in R and $f_k:R\to R/I_k$ is the canonical surjective morphism of rings for each k. Then define the morphism $f:R\to\prod_{k\in K} R/I_k$ by $f(x)=(f_k(x))_{k\in K}$ where $\prod_{k\in K} R/I_k$ is the product of the family of rings $\{R/I_k\}_{k\in K}$. Then it is obvious that $\operatorname{Ker} f=\bigcap_{k\in K} I_k$. Although it is not so clear how to describe $\operatorname{Im} f$ in general, there is a special case in which this can easily be done.

Proposition 6.6
Let $I_1,\ldots,I_n$ be a finite set of ideals in the ring R such that R is generated by I_i and I_j whenever $i\neq j$. If for every k in $[1,\ldots,n]$ we denote by $f_k:R\to R/I_k$ the canonical surjective morphism, then the morphism $f:R\to\prod_{k=1}^{n} R/I_k$ given by $f(r)=(f_k(r))_{k\in[1,\ldots,n]}$ for all r in R is a surjective morphism with $\operatorname{Ker} f=\bigcap_{k=1}^{n} I_k=I_1\cdots I_n$. Hence, f induces an isomorphism

$$\frac{R}{\bigcap_{k=1}^{n} I_k}\cong\prod_{i=1}^{n}\frac{R}{I_k}$$

PROOF: We leave it to the reader to show that the proposition in question is

equivalent to the statement that given any set of elements $a_1, \ldots, a_n$ in R there is a single element r in R satisfying $r - a_k$ is in I_k for each k in $[1, \ldots, n]$. We now prove this second assertion by induction on n.

Suppose $n = 2$. We want to show that given any pair of elements a_1 and a_2 in R there is an r in R such that $r - a_1$ is in I_1 and $r - a_2$ is in I_2. Because R is generated by I_1 and I_2 we know there are elements r_1 in I_1 and r_2 in I_2 such that $1 = r_1 + r_2$. Thus, $a_1 = a_1 r_1 + a_1 r_2$ while $a_2 = a_2 r_1 + a_2 r_2$. Then it is easily seen that $r = a_2 r_1 + a_1 r_2$ has the desired properties: $r - a_1$ is in I_1 and $r - a_2$ is in I_2.

Assume now that $n > 2$. We want to show that given any sequence $a_1, \ldots, a_n$ of elements in R, we can find an r in R such that $r - a_k$ is in I_k for all $k = 1, \ldots, n$. By our inductive hypothesis, we know there is an element r' in R such that $r' - a_i$ is in I_i for $i = 1, \ldots, n-1$. But by Lemma 6.4, we know that $I_1 \cdots I_{n-1}$ and I_n generate R so that there is an r in R with $r - r'$ in $I_1 \cdots I_{n-1}$ and $r - a_n$ is in I_n. However, $r - a_1 = r - r' + r' - a_i$ and, since $r - r'$ is in $I_1 \cdots I_{n-1} \subset I_i$ and $r' - a_i$ is in I_i, it follows that $r - a_i$ is in I_i for $i = 1, \ldots, n-1$.

Therefore, we see that the morphism $f: R \to \prod_{k=1}^{n} R/I_k$ is a surjective morphism. Since we have already observed that the kernel of f is $\bigcap_{k=1}^{n} I_k$, we see that the morphism f induces an isomorphism $R/\bigcap_{k=1}^{n} I_k \approx \prod_{k=1}^{n} R/I_k$.

This proposition is known as the Chinese Remainder Theorem.

If R is a PID and I is a nonzero proper ideal of R, we have $I = (a)$ for some a in R. Letting $a = p_1^{n_1} \cdots p_t^{n_t}$ be a factorization of a into powers of distinct primes, we see that $I = I_1, \ldots, I_t$ where the ideal $I_j = (p_j^{n_j})$ for $j = 1, \ldots, t$. Because I_j and I_k together generate R when $j \neq k$, our last result tells us that $R/I \approx R/I_1 \times \cdots \times R/I_t$. Thus, we have the following.

Proposition 6.7
Let R be a PID and I a nonzero proper ideal of R. Then R/I is isomorphic to the product of rings $R/I_1 \times \cdots \times R/I_t$ where each ideal I_j is generated by a power of a prime element in R.

7. DIVISORS

In the preceding sections, the formulation of concepts and proofs have been primarily in terms of elements in a commutative ring. In this section, we indicate how these ideas may be presented in ideal-theoretic terms. Because the proofs of the propositions stated in this section can be obtained from those already given in preceding sections, no proofs are included here. The reader is urged, however, to familiarize himself with the contents of this section because this language will be extensively used from now on.

Definition
A **principal divisor** of an integral domain R is any nonzero principal ideal of R.

The reader should have no difficulty in establishing the following.

Basic Properties 7.1
Let R be an integral domain.

(a) If I_1 and I_2 are principal divisors of R, then I_1I_2 is a principal divisor. Hence, the set of all principal divisors of R is submonoid of $PI(R)$, the monoid of all principal ideals of R (see Section 1).
(b) If I, I_1, and I_2 are principal divisors of R such that $II_1 = II_2$, then $I_1 = I_2$.
(c) For two principal divisors $I_1 = (r_1)$ and $I_2 = (r_2)$ of R, the following are equivalent:
  (i) $I_1 \supset I_2$.
  (ii) $r_1 | r_2$.
  (iii) There is a principal divisor I such that $II_1 = I_2$.

We summarize some of these observations in the following.

Definitions
Let R be an integral domain. We denote by $PD(R)$ the submonoid of $PI(R)$ consisting of all principal divisors.

If I_1 and I_2 are two principal divisors of R, we say that I_1 **divides** I_2, which we denote by $I_1 | I_2$, if there is a principal divisor I such that $II_1 = I_2$.

We now point out, for principal divisors, the following obvious analogs of the notions of irreducible and prime elements.

Definition
Let R be an integral domain and I a principal divisor of R different from R.

(a) I is said to be an **irreducible divisor** of R if R and I are the only principal divisors containing I.
(b) I is said to be a **principal prime divisor** of R if I is a prime ideal of R.

Basic Properties 7.2
Let r be a nonzero element of R. Then:

(a) r is irreducible if and only if (r) is an irreducible divisor.
(b) r is a prime element if and only if (r) is a principal prime divisor.

For convenience of reference we make the following.

Definition
The set of all principal prime divisors for an integral domain R is denoted by $PPD(R)$.

We now list some properties of principal divisors of R which can either be easily derived from or proven in essentially the same way as their analogs for nonzero elements of R.

Basic Properties 7.3
Let R be an integral domain.

(a) A principal divisor I is a principal prime divisor if and only if given two principal divisors I_1 and I_2 such that $I_1I_2 \subset I$, then either I_1 or I_2 is contained in I.

(b) Every principal prime divisor is irreducible.
(c) If I_1 and I_2 are principal prime divisors and $I_1 \subset I_2$, then $I_1 = I_2$.
(d) Suppose $\{n_{\mathfrak{P}}\}_{\mathfrak{P} \in PPD(R)}$ and $\{n'_{\mathfrak{P}}\}_{\mathfrak{P} \in PPD(R)}$ are two almost zero families of elements in **N**. Then

$$\prod_{\mathfrak{P} \in PPD(R)} \mathfrak{P}^{n_{\mathfrak{P}}} = \prod_{\mathfrak{P} \in PPD(R)} \mathfrak{P}^{n_{\mathfrak{P}}} \quad \text{if and only if} \quad n_{\mathfrak{P}} = n'_{\mathfrak{P}}$$

for all $\mathfrak{P} \in PPD(R)$.

The reader should have no difficulty establishing the following characterization of UFD's in terms of the monoid $PPD(R)$.

Proposition 7.4
An integral domain R is a unique factorization domain if and only if for each principal divisor I of R there is an almost zero family $\{n_{\mathfrak{P}}\}_{\mathfrak{P} \in PPD(R)}$ of elements in **N** such that $I = \prod_{\mathfrak{P} \in PPD(R)} \mathfrak{P}^{n_{\mathfrak{P}}}$.

Combining this last proposition with the previous basic properties we have the following.

Proposition 7.5
Suppose R is a unique factorization domain. Then, given any principal divisor I of R, there is one and only one almost zero family $\{n_{\mathfrak{P}}\}_{\mathfrak{P} \in PPD(R)}$ of elements in **N** such that $I = \prod_{\mathfrak{P} \in PPD(R)} \mathfrak{P}^{n_{\mathfrak{P}}}$.

This result serves as the basis for the following.

Definition
Suppose R is a unique factorization domain. If I is a principal divisor of R, then the unique representation $I = \prod_{\mathfrak{P} \in PPD(R)} \mathfrak{P}^{n_{\mathfrak{P}}}$, with $\{n_{\mathfrak{P}}\}_{\mathfrak{P} \in PPD(R)}$ an almost zero family of elements in **N**, is called the **primary decomposition** of I.

The uniquely determined integers $n_{\mathfrak{P}}$ for each $\mathfrak{P}$ in $PPD(R)$ which appear in the primary decomposition of I is denoted by $n_{\mathfrak{P}}(I)$ for each $\mathfrak{P}$ in $PPD(R)$.

Finally, if x is a nonzero element of R, then for each $\mathfrak{P}$ in $PPD(R)$, we denote by $n_{\mathfrak{P}}(x)$ the integer $n_{\mathfrak{P}}((x))$.

It is important, at this point, to compare the almost zero family $\{n_{\mathfrak{P}}(x)\}_{\mathfrak{P} \in PPD(R)}$ with the family of prime exponents $\{n_i(x)\}_{i \in I}$ associated with a representative family of prime elements, introduced in Section 4.

The following basic facts concerning primary decompositions of principal divisors of unique factorization domains should be verified.

Basic Properties 7.6
Let R be a unique factorization domain. Then we have:

(a) A principal divisor I is R if and only if $n_{\mathfrak{P}}(I) = 0$ for all $\mathfrak{P}$ in $PPD(R)$.
(b) If I_1 and I_2 are principal divisors of R, then $n_{\mathfrak{P}}(I_1 I_2) = n_{\mathfrak{P}}(I_1) + n_{\mathfrak{P}}(I_2)$ for all $\mathfrak{P}$ in $PPD(R)$.

(c) For two principal divisors I_1 and I_2, the following are equivalent:
  (i) $I_1 \supset I_2$.
  (ii) $I_1 | I_2$, that is, $I_1 I = I_2$ for some principal divisor I of R.
  (iii) For each $\mathfrak{P}$ in $PPD(R)$ we have

$$n_{\mathfrak{P}}(I_2) \geq n_{\mathfrak{P}}(I_1).$$

We underscore, at the risk of being redundant, the following analogs for the nonzero elements of a unique factorization domain.

Basic Properties 7.7
Let R be a unique factorization domain.

(a) An element r in R is a unit in R if and only if $n_{\mathfrak{P}}(r) = 0$ for all $\mathfrak{P}$ in $PPD(R)$.
(b) Suppose that r_1 and r_2 are two nonzero elements of R. Then $n_{\mathfrak{P}}(r_1) = n_{\mathfrak{P}}(r_2)$ for all $\mathfrak{P}$ in $PPD(R)$ if and only if there is a unit u such that $ur_1 = r_2$.
(c) If r_1 and r_2 are nonzero elements of R, then $n_{\mathfrak{P}}(r_1 r_2) = n_{\mathfrak{P}}(r_1) + n_{\mathfrak{P}}(r_2)$ for all $\mathfrak{P}$ in $PPD(R)$.
(d) For two nonzero elements r_1 and r_2 in R, the following are equivalent:
  (i) $r_1 | r_2$.
  (ii) $n_{\mathfrak{P}}(r_2) \geq n_{\mathfrak{P}}(r_1)$ for all $\mathfrak{P}$ in $PPD(R)$.

As for greatest common divisors and least common multiples, the basic properties already cited for these notions for elements fully justify the following analogs for principal divisors.

Definition
Let $I_1, \ldots, I_n$ be a finite family of principal divisors of an integral domain R.

(a) A principal divisor I is said to be the **greatest common divisor** of $I_1, \ldots, I_n$ which we denote by $I = \gcd[I_1, \ldots, I_n]$ if:
  (i) $I \supset (I_1, \ldots, I_n)$, the ideal generated by $\bigcup_{i=1}^{n} I_i$.
  (ii) If J is a principal divisor of R containing $(I_1, \ldots, I_n)$, then $J \supset I$.
(b) A principal divisor I is said to be the **least common multiple** of $I_1, \ldots, I_n$, which we denote by $I = \operatorname{lcm}[I_1, \ldots, I_n]$ if $I = I_1 \cap \cdots \cap I_n$.

We now state the analog, for divisors, of Proposition 4.5.

Proposition 7.8
Let $I_1, \ldots, I_n$ be a finite family of principal divisors of a unique factorization domain R.

(a) For each $\mathfrak{P}$ in $PPD(R)$, let $n_{\mathfrak{P}} = \min(n_{\mathfrak{P}}(I_1), \ldots, n_{\mathfrak{P}}(I_n))$. Then $\{n_{\mathfrak{P}}\}_{\mathfrak{P} \in PPD(R)}$ is an almost zero family of elements in $\mathbf{N}$ with the property that $I = \prod_{\mathfrak{P} \in PPD(R)} \mathfrak{P}^{n_{\mathfrak{P}}}$ is $\gcd[I_1, \ldots, I_n]$.
(b) For each $\mathfrak{P}$ in $PPD(R)$, let $n_{\mathfrak{P}} = \max(n_{\mathfrak{P}}(I_1), \ldots, n_{\mathfrak{P}}(I_n))$. Then $\{n_{\mathfrak{P}}\}_{\mathfrak{P} \in PPD(R)}$ is an almost zero family of elements in $\mathbf{N}$ with the property that $I = \prod_{\mathfrak{P} \in PPD(R)} \mathfrak{P}^{n_{\mathfrak{P}}}$ is the $\operatorname{lcm}[I_1, \ldots, I_n]$.
(c) If I_1 and I_2 are two principal divisors of R, then

$$I_1 I_2 = (\operatorname{lcm}[I_1, I_2]) \cdot (\gcd[I_1, I_2])$$

Note that this last statement asserts that if x and y are nonzero elements of a UFD, R, then if c is a lcm$[x, y]$ we have $xy = cd$ where d is a gcd$[x, y]$.

Corollary 7.9
Suppose R is a unique factorization domain. Let $\{p_{\mathfrak{P}}\}_{\mathfrak{P} \in PPD(R)}$ be a family of prime elements of R with the property $(p_{\mathfrak{P}}) = \mathfrak{P}$ for all $\mathfrak{P}$ in $PPD(R)$ (that is, a representative family of prime elements of R).

(a) If x is a nonzero element of R, then there is a uniquely determined unit u in R such that $x = u \prod_{\mathfrak{P} \in PPD(R)} p_{\mathfrak{P}}^{n_{\mathfrak{P}}(x)}$.
(b) If $x_1, \ldots, x_n$ is a finite family of nonzero elements in R and $n_{\mathfrak{P}} = \min(n_{\mathfrak{P}}(x_1), \ldots, n_{\mathfrak{P}}(x_n))$ for each $\mathfrak{P}$ in $PPD(R)$, then $x = \prod_{\mathfrak{P} \in PPD(R)} p_{\mathfrak{P}}^{n_{\mathfrak{P}}}$ is a gcd$[x_1, \ldots, x_n]$.
(c) If $x_1, \ldots, x_n$ is a finite family of nonzero elements in R and $n_{\mathfrak{P}} = \max(n_{\mathfrak{P}}(x_1), \ldots, n_{\mathfrak{P}}(x_n))$, then $x = \prod p_{\mathfrak{P}}^{n_{\mathfrak{P}}}$ is a lcm$[x_1, \ldots, x_n]$.
(d) If x_1 and x_2 are nonzero elements of R and $x = \text{lcm}[x_1, x_2]$ and $y = \text{gcd}[x_1, x_2]$, then $xy = ux_1x_2$ with u a unit in R.

8. LOCALIZATION IN INTEGRAL DOMAINS

In this section we apply the notion of localization for integral domains to obtain some new examples of unique factorization domains as well as principal ideal domains. We begin by saying what we mean by localization for integral domains.

Let R be an integral domain with field of quotients $Q(R)$. Suppose S is a submonoid of the multiplicative monoid of nonzero elements in R. Let R_S be the subset of $Q(R)$ consisting of all quotients r/s with s in S. Then it is easily checked that R_S is a subring of $Q(R)$ containing R as a subring. Clearly, if $S = R - \{0\}$, then $R_S = Q(R)$. Because the rings of the form R_S play an important role in studying integral domains, they are given a special name.

Definitions
Let R be an integral domain. A subset S of R is called a multiplicative subset of R if it is a submonoid of the multiplicative monoid of nonzero elements of R. If S is a multiplicative subset of R, the subring of $Q(R)$ consisting of all quotients r/s with s in S is called the **localization of R with respect to S** and is denoted by R_S.

In order to underscore the connection between localization and unique factorization domains, we show how to express, in terms of localization, that an integral domain is a unique factorization domain.

Proposition 8.1
For an integral domain R the following are equivalent:

(a) R is a unique factorization domain.
(b) $R_S = Q(R)$, where S is the multiplicative set consisting of all elements of R which can be written as a unit in R times a finite product of prime elements in R.

(c) There is a multiplicative set S consisting of elements which can be written as a unit in R times a finite product of prime elements in R such that $R_S = Q(R)$.

PROOF: (a) implies (b) and (b) implies (c) are trivial.

(c) implies (a). In order to show that R is a UFD, we must show that every nonzero element r in R is a unit times a finite product of prime elements. Since we are assuming that $R_S = Q(R)$, we know that if r is a nonzero element in R, then $1/r = r'/s$ where r' is in R and s is in S. From this it follows that $rr' = s = u \prod_{i=1}^{n} p_i$ where u is a unit in R and the p_i are prime elements in R. The fact that this implies that both r and r' are of the form a unit in R times a finite product of prime elements in R follows easily by induction on n, the number of prime elements in the expression $rr' = u \prod_{i=1}^{n} p_i$. This proof is left as an exercise for the reader.

The following general properties of localization play an important role in our applications of this technique to the study of unique factorization domains.

Proposition 8.2
Let S and T be multiplicative subsets of the integral domain R.

(a) Because $R \subset R_S \subset Q(R)$, we have $Q(R) = Q(R_S)$.
(b) If $S \subset T$, then $R \subset R_S \subset R_T \subset Q(R)$. Hence, if $R_S = Q(R) = Q(R_S)$, then $R_T = Q(R_T) = Q(R)$.
(c) T is a multiplicative subset of R_S and S is a multiplicative subset of R_T which are related by:
  (i) $(R_S)_T = R_{ST} = (R_T)_S$ where ST is the multiplicative subset of R consisting of all products st where s is in S and t is in T. Clearly, ST contains $S \cup T$.
  (ii) R_{ST} contains R_S and R_T.
(d) If x is a nonzero element in R, then x is a unit in R_S if and only if rx is in S for some r in R or, what is the same thing, $Rx \cap S \neq \emptyset$.
(e) If x is a prime element in R, then x is either a prime or a unit in R_S according to whether $Rx \cap S$ is empty or not empty.

PROOF: Left as an exercise for the reader.

As an easy consequence of this proposition we have the following.

Corollary 8.3
Suppose S is a multiplicative subset of the unique factorization domain R. Then R_S is also a unique factorization domain.

PROOF: Let T be the multiplicative subset $R - \{0\}$ of R. Because R is a UFD, we know that all the elements of T can be written as a unit in R times a finite product of prime elements in R. Because units in R remain units in R_S and prime elements in R are either units or primes in R_S, we know that T, viewed as a multiplicative subset of R_S, consists of elements which are units times a finite product of primes in R_S. Therefore, by our characterizations of unique factorization domains by means of localization (see Proposition 8.1), it will follow that R_S is a unique factorization domain if we show that $(R_S)_T = Q(R_S)$. But the fact that

$Q(R_S) = Q(R) \supset (R_S)_T \supset R_T = Q(R) = Q(R_S)$ implies $(R_S)_T = Q(R_S)$, which gives us our desired result, namely, that R_S is a UFD.

The remaining results concerning unique factorization domains which we develop in this section depend on the relationship between $PPD(R)$ and $PPD(R_S)$ where S is a multiplicative subset of the integral domain R. Because the elements of $PPD(R)$ and $PPD(R_S)$ are principal prime ideals of R and R_S, respectively, it is appropriate to begin this discussion by pointing out some of the general connections between the ideals of R and those of R_S. We first make the useful definition.

Definition
Suppose S is a multiplicative subset of the integral domain R. If I is an ideal of R we denote by IR_S or I_S the ideal of R_S generated by the subset I of R_S.

Basic Properties 8.4
Let S be a multiplicative subset of the integral domain R. If I is an ideal of R, then:

(a) The ideal I_S of R_S consists of all elements of R_S which can be written in the form x/s with x in I and s in S.
(b) If $\{x_k\}_{k\in K}$ generates I as an ideal in R, then $\{x_k\}_{k\in K}$ generates I_S as an ideal in R_S.
(c) Hence, if I is a finitely generated ideal of R, then I_S is a finitely generated ideal of R_S.
(d) In particular, if I is a principal ideal of R, then I_S is a principal ideal of R_S.
(e) $I_S = R_S$ if and only if $I \cap S \neq \emptyset$.

PROOF: (a) Because I_S is the ideal of R_S generated by the subset I of R_S, we know that the elements of I_S are the finite sums $(r_1/s_1)x_1 + \cdots + (r_n/s_n)x_n$ with the x_j in I and r_j and s_j in R and S, respectively. But

$$\left(\frac{r_1}{s_1}\right)x_1 + \cdots + \left(\frac{r_n}{s_n}\right)x_n = \frac{\sum_{j=1}^{n} r_j t_j x_j}{\prod_{j=1}^{n} s_j}$$

where $t_j = \prod_{k\neq j} s_k$. Because $\Sigma_{j=1}^{n} r_j t_j x_j$ is in I and $\prod_{j=1}^{n} s_j$ is in S, it follows that the elements of I_S can all be written in the form x/s with x in I and s in S. Because each element x/s with x in I and s in S can be written as $(1/s)\cdot x$ and each $1/s$ is in R_S, it follows that each such x/s is in I_S. Thus, I_S consists precisely of the elements x/s with x in I and s in S.

(b) Suppose $\{x_k\}_{k\in K}$ generates the ideal I of R. We want to show that $\{x_k\}_{k\in K}$ also generates the ideal I_S of R_S. Suppose x/s is in I_S. Then there is a finite subset K' of K and a family $\{r_k\}_{k\in K'}$ of elements of R such that $x = \sum_{k\in K'} r_k x_k$. Therefore, $x/s = (\Sigma_{k\in K'} r_k x_k)/s = \Sigma_{k\in K'} (r_k/s)\cdot x_k$. Because each r_k/s is in R_S, we see that the family $\{x_k\}_{k\in K}$ does indeed generate the ideal I_S of R_S.

(c), (d), and (e) are left as exercises.

If J is an ideal of R_S, then it is not difficult to check that $J \cap R$ is an ideal of R. Thus, associated with the ideal J of R_S is the ideal $J \cap R$ of R. On the other hand,

associated with each ideal I of R is the ideal I_S of R_S. We now investigate the connections between these two operations.

Proposition 8.5
Suppose S is a multiplicative subset of the integral domain R. Then for each ideal J of R_S, the ideal $J \cap R$ of R has the following properties:

(a) For an element x in R the following statements are equivalent:
  (i) x is in $J \cap R$.
  (ii) There is an s in S such that sx is in J.
  (iii) sx is in J for all s in S.
  (iv) x/s is in J for all s in S.
  (v) x/s is in J for some s in S.
(b) $(J \cap R)_S = J$.
(c) If $J \neq R_S$, then $(J \cap R) \cap S = \emptyset$.
(d) If J is a prime ideal of R_S, then $J \cap R$ is a prime ideal of R.
(e) If J_1 and J_2 are ideals of R_S such that $J_1 \cap R = J_2 \cap R$, then $J_1 = J_2$.

PROOF: We leave everything except part (d) as an exercise. Part (d) is an obvious special case of the following.

Lemma 8.6
Suppose $f: R \to T$ is a morphism of arbitrary commutative rings. If I is a prime ideal of T, then $f^{-1}(I)$ is a prime ideal of R.

PROOF: The morphism $f: R \to T$ induces an injective morphism of rings $R/f^{-1}(I) \to T/I$. Because I is a prime ideal of T, we know that T/I is an integral domain. Hence, $R/f^{-1}(I)$ is an integral domain because it is isomorphic to a subring of the integral domain T/I. Therefore, $f^{-1}(I)$ is a prime ideal of R.

As an immediate consequence of Proposition 8.5 and Basic Properties 8.4 we have the following.

Corollary 8.7
Suppose S is a multiplicative subset of the integral domain R.

(a) R_S is noetherian if R is noetherian.
(b) R_S is a PID if R is a PID.

PROOF: (a) We show that R_S is noetherian by showing that each ideal J of R_S is finitely generated. Because R is noetherian every ideal of R is finitely generated. In particular, $J \cap R$ is a finitely generated ideal of R for each ideal J of R_S. Hence, by our previous basic properties, $(J \cap R)_S$ is a finitely generated R_S ideal. But we have just shown that $J = (J \cap R)_S$, which means that J is a finitely generated ideal of R_S because a finite set of generators for the ideal $J \cap R$ of R is also a set of generators for the R_S ideal $(J \cap R)_S = J$. Hence, R_S is noetherian, if R is noetherian.
(b) Proven similarly.

We now want to investigate the connections between an ideal I of R and the ideal $I_S \cap R$ of R. It is obvious that $I_S \cap R \supset I$. In the exercises, an example is given

to show that $I_S \cap R$ need not be I. The reason for this is that the ideal $I_S \cap R$ must, as we have already seen, have the property that if x is in R and sx is in $I_S \cap R$ for some s in S, then x is in $I_s \cap R$ [see Proposition 8.5(a)]. However, the ideal I need not have this property. In fact, this observation completely accounts for the difference between I and $I_S \cap R$ as we shall presently see.

To this end, we notice that if I is any ideal in an integral domain R and S is a multiplicative set in R, then the set of all x in R such that sx is in I for some s in S is an ideal of R containing I. The importance of this ideal is that it is precisely $I_S \cap R$, a fact which we shall verify soon. For convenience of reference we make the following definition.

Definition
Let I be an ideal in an integral domain R and let S be a multiplicative subset of R. We call the ideal consisting of all x in R such that sx is in I for some s in S, the **S-closure of I**. We denote the S-closure of I by $Cl_S(I)$. We say that I is **S-closed** if $I = Cl_S(I)$.

Basic Properties 8.8
Suppose S is a multiplicative set in the integral domain R. If I is an ideal of R, then:

(a) $Cl_S(I) = I_S \cap R$.
(b) $Cl_S(I)$ is S-closed because if J is any ideal of R_S, then $J \cap R$ is S-closed.
(c) Hence, $I_S \cap R = I$ if and only if I is S-closed.
(d) If I is a prime ideal of R, then I is S-closed if and only if $I \cap S = \emptyset$.
(e) If I' is another ideal in R, then $I_S = (I')_S$ if and only if $Cl_S(I) = Cl_S(I')$.
(f) If $I \neq R$ is an S-closed ideal, then $I \cap S = \emptyset$.

PROOF: (a) We have already seen that since I_S is an ideal in R_S, an element x in R is in $I_S \cap R$ if and only if there is an s in S such that sx is in I_S, that is, such that $sx = r/s'$ with r in I and s' in S. Therefore, if x is in $I_S \cap R$, then $ss'x = r$ in I. From this it follows that $I_S \cap R \subset Cl_S(I)$. On the other hand, if x is in $Cl_S(I)$, then $sx = r$ for some r in I and some s in S which implies that $x = r/s$ or, equivalently, x is in $I_S \cap R$. This shows that $Cl_S(I) \subset I_S \cap R$ and thus $Cl_S(I) = I_S \cap R$.

(b) and (c) are left as exercises.

(d) Suppose I is a prime ideal in R. Then $I \neq R$. Now if $S \cap I \neq \emptyset$, then $I_s = R_s$ and so $I_s \cap R = R \neq I$. Thus, under these circumstances, I is not S-closed. Hence, if I is S-closed, then $S \cap I = \emptyset$.

Assume $S \cap I = \emptyset$. Let x be an element in R having the property that there is an s in S such that sx is in I. This implies x is in I because I is a prime ideal and sx is in I although s is not in I.

(e) Clearly, in order to show that $I_S = I'_S$ if $Cl_S(I) = Cl_S(I')$, it suffices to show that $I_S = (Cl_S(I))_S$. Because $Cl_S(I) \supset I$, we have that $(Cl_S(I))_S \supset I_S$. We now show $I_S \supset (Cl_S(I))_S$. Each element of $(Cl_S(I))_S$ can be written in the form y/s with y in $Cl_S(I)$ and s in S. The fact that y is in $Cl_S(I)$ means that $s'y = x$ for some x in I and some s' in S. Therefore, $y = x/s'$ which implies $y/s = x/ss'$. But x/ss' is in I_S. This shows that $(Cl_S(I))_S \subset I_S$ which finishes the proof that $I_S = (Cl_S(I))_S$. The rest of part (e) is left as an exercise, as is the proof of (f).

As a partial summary of these results concerning the relationship between the ideals of R and those of R_S we have the following.

Proposition 8.9
Let S be a multiplicative subset of an integral domain R. Then the map

$$f:\{S\text{-closed ideals of } R\}\to\{\text{ideals of } R_S\}$$

given by $f(I)=I_S$ for all S-closed ideals I of R has the following properties:

(a) f is a bijective map whose inverse is given by $f^{-1}(J)=J\cap R$ for all ideals J of R_S.
(b) If the S-closed ideal I can be generated by n elements, then $f(I)=I_S$ can also be generated by n elements.
(c) An S-closed ideal I of R is a prime ideal of R if and only if $f(I)=I_S$ is a prime ideal of R_S.
(d) Hence, f induces a bijective map between the prime ideals of R which do not meet S and all the prime ideals of R_S.

9. A CRITERION FOR UNIQUE FACTORIZATION

We now apply these results to unique factorization domains. We begin with the following.

Definition
Suppose S is a multiplicative subset of the integral domain R. Then we denote by $PPD_S(R)$ the subset of $PPD(R)$ consisting of all principal prime divisors I such that $I\cap S=\emptyset$.

By our previous results we know that if $\mathfrak{P}$ is in $PPD_S(R)$, then $\mathfrak{P}_S$ is a principal prime divisor of R_S. This suggests considering the map $PPD_S(R)\to PPD(R_S)$ given by $\mathfrak{P}\to\mathfrak{P}_S$ for all $\mathfrak{P}$ in $PPD_S(R)$. Because the elements of $PPD_S(R)$ are S-closed, we know that the map $PPD_S(R)\to PPD(R_S)$ is always injective. Hence, it is natural to ask when the map $PPD_S(R)\to PPD(R_S)$ is surjective or, what is the same thing, bijective. In the exercises we will give an example of an integral domain R which has a multiplicative set S such that the map $PPD_S(R)\to PPD(R_S)$ is not bijective. In the meantime, we point out some cases where this map is an isomorphism and give some applications of this fact. We begin with the simplest case.

Proposition 9.1
Suppose R is a PID. If S is any multiplicative subset of R, then the map

$$PPD_S(R)\to PPD(R_S)$$

is bijective.

PROOF: We have already shown that R_S is a PID because R is a PID. Hence, $PPD(R_S)$ is nothing more than the set of nonzero prime ideals of R_S. Suppose $\mathfrak{P}$ is a nonzero prime ideal of R_S. Then we known that $\mathfrak{P}\cap R$ is a prime ideal of R with

the property $(\mathfrak{P} \cap R)_S = \mathfrak{P}$. Therefore, $\mathfrak{P} \cap R$ is a nonzero prime ideal of R and hence an element of $PPD(R)$ since R is a PID. Because $\mathfrak{P} \cap R$ is also S-closed, it follows that $\mathfrak{P} \cap R$ is an element of $PPD_S(R)$ which goes to $\mathfrak{P}$ under the map $PPD_S(R) \to PPD(R_S)$. This shows that if R is a PID, then the map $PPD_S(R) \to PPD(R_S)$ is surjective and hence bijective.

In connection with this result, it is interesting to observe the following easily verified proposition.

Proposition 9.2

Suppose R is a PID and X is an arbitrary subset of $PPD(R)$. If S consists of all s in R not divisible by any prime element p in R such that (p) is in X, then:

(a) S is a multiplicative subset of R.
(b) $X = PPD_S(R)$.
(c) The map $X \to PPD(R_S)$ given by $\mathfrak{P} \to \mathfrak{P}_S$ for all $\mathfrak{P}$ in X is bijective.

This last observation can be used to show that given any integer $n \geq 0$ there is a PID R with Card $(PPD(R)) = n$. To accomplish this, all one has to do is show that there is a PID R with $PPD(R)$ an infinite set. For suppose R is a PID with $PPD(R)$ an infinite set. Then given any integer $n \geq 0$, there is a subset X of $PPD(R)$ with n-elements. But by our previous result, there is a multiplicative subset S of R such that $X = PPD_S(R)$. Therefore, R_S is a PID such that $PPD(R_S)$ has n-elements because the map $X \to PPD(R_S)$ given by $\mathfrak{P} \to \mathfrak{P}_S$ for all $\mathfrak{P}$ in X is an isomorphism of sets.

We now show that there are PID's R with $PPD(R)$ an infinite set. In particular, we show that $\mathbf{Z}$ and $R[X]$, the ring of polynomials over a field, all have an infinite number of principal prime divisors.

We first observe that the rings $\mathbf{Z}$ and $R[X]$, with R a field, all have the property that if a is any nonzero noninvertible element in any such ring, then either $1 + a$ or $1 - a$ is not a unit. That this is true for $\mathbf{Z}$ is left as an exercise. If R is a field, then we know that a nonzero polynomial in $R[X]$ is a unit if and only if its degree is zero. Because $\deg(1 + a) = \deg(a)$ for any nonzero element a of degree greater than zero, then $1 + a$ is not a unit. The fact that $\mathbf{Z}$ and $R[X]$, with R a field, have infinitely many principal prime divisors, now follows from the more general proposition.

Proposition
Let R be a unique factorization domain with the property that either $1 + r$ or $1 - r$ is not a unit for all elements r in R which are neither zero nor a unit. Then $PPD(R)$ is an infinite set.

PROOF: Suppose $PPD(R)$ is finite with $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ its distinct elements. Let p_i be a generator of $\mathfrak{P}_i$ for each $i = 1, \ldots, n$. Then $\prod_{i=1}^{n} p_i$ is not a unit in R and hence $1 + \prod_{i=1}^{n} p_i$ or $1 - \prod_{i=1}^{n} p_i$ is not a unit in R. Therefore, $p_j | (1 + \prod_{i=1}^{n} p_i)$ or $p_j | (1 - \prod_{i=1}^{n} p_i)$ for some $j = 1, \ldots, n$. In either event $p_j | 1$, which is impossible. This contradiction shows that the set $PPD(R)$ is infinite.

We now return to our general question as to when the injective map $PPD_S(R) \to PPD(R_S)$ is bijective for a multiplicative subset S of an integral domain R. We have already seen that $PPD_S(R) \to PPD(R_S)$ is an isomorphism of sets for any multiplicative set S in a PID, R. As a result of some slightly more general considerations which are of considerable interest in their own right, we will extend this result to arbitrary multiplicative sets of arbitrary UFD's.

We begin with the following useful lemma.

Lemma 9.4

Let S be a multiplicative subset of the integral domain R. Suppose the set of principal ideals in R is noetherian. If x is a nonzero element of R, then $x = st$ where s is in S and t is not divisible by any nonunit in S.

PROOF: Let $\mathcal{T}$ be the set of all principal ideals (y) with the property that there is an s in S such that $sy = x$. $\mathcal{T}$ is not empty because (x) is in $\mathcal{T}$. Therefore, the set $\mathcal{T}$ has a maximal element (t), because the set of principal ideals in R is noetherian. We claim that only the units in S divide t. For suppose $t = t_1 s_1$ with s_1 a nonunit in S. Because $x = ts$ with s in S, then $x = t_1 s_1 s$. So (t_1) is in $\mathcal{T}$ and contains (t) but is not equal to t because $t = t_1 s_1$ with s_1 not a unit. This contradicts the fact that (t) is a maximal element of $\mathcal{T}$, Therefore $x = ts$ where s is in S and t is not divisible by any nonunit in S, completing the proof of the lemma.

Definition

Let S be a multiplicative subset of the integral domain R. We say S is **generated by primes** if for each x in S there are elements $u, p_1, \ldots, p_t$ in S such that $x = up_1, \ldots, p_t$ where u is a unit of R and $p_1, \ldots, p_t$ are prime elements of R.

Lemma 9.5

Let R be an integral domain such that the set of principal ideals in R is noetherian. Suppose S is a multiplicative subset of R generated by primes.

(a) Let x be an element in R with the property that s in S is a unit in R if $s|x$ in R. Suppose r is an element in R such that $x|r$ in R_S. Then $x|r$ in R.
(b) Each principal divisor of R_S can be generated by an element x in R satisfying the hypothesis of (a).
(c) If $\mathfrak{P}$ is a principal prime divisor of R_S, then there is a principal prime divisor $\mathfrak{P}'$ of R such that $\mathfrak{P}'_S = \mathfrak{P}$.

PROOF: (a) Suppose x has the property that if s in S divides x in R, then s is a unit in R. Further, suppose r in R is divisible by x in R_S, that is, $r = (t/s)x$ with t in R and s in S. By hypothesis, $s = u \prod_{i=1}^{n} p_i$ where u is a unit in S and $p_1, \ldots, p_n$ are prime elements of R contained in S. Because none of the p_i divide x, the fact that $tx = rs = ru \prod_{i=1}^{n} p_i$ implies (by induction on n) that $s|t$ in R. Hence, $x|r$ in R, which completes the proof of (a).

(b) Suppose (r/s) is a principal divisor of R_S. Because the set of principal

ideals of R is noetherian, we know by Lemma 9.4 that $r = s'x$ with s' in S and x an element of R satisfying the hypothesis of (a). Hence, $x = (r/s)(s/s')$, which means that $xR_S = (r/s)R_S$ because s/s' is a unit in R_S. This completes the proof of (b).

(c) Suppose $\mathfrak{P}$ is a principal prime divisor of R_S. Then by (a) and (b) we know that $\mathfrak{P} = xR_S$ where x is an element of R having the property that if r is in R and $x|r$ in R_S, then $x|r$ in R. We now show that this property, combined with the fact that x is a prime element in R_S (remember xR_S is a principal prime divisor in R_S), implies that x is a prime element in R.

For suppose r_1 and r_2 are in R and $x|r_1r_2$ in R. Then $x|r_1r_2$ in R_S and hence $x|r_1$ or $x|r_2$ in R_S. Therefore, it follows that $x|r_1$ or $x|r_2$ in R which finishes the proof that x is a prime element in R. Because the principal prime divisor (x) of R obviously has the property $(x)_S = xR_S = \mathfrak{P}$, the proof of (c) is complete.

As an immediate consequence of this lemma we have the following.

Proposition 9.6
Suppose R is an integral domain with the property that its set of principal ideals is noetherian. Further, suppose S is a multiplicative subset of R generated by primes. Then the map

$$PPD_S(R) \to PPD(R_S)$$

given by $P \to P_S$ is an isomorphism of sets.

We now give several applications of this result. The first is to show that if S is a multiplicative set in a unique factorization domain R, then the map $PPD_S(R) \to PPD(R_S)$ is an isomorphism. This will follow from our previous proposition and the following general observation.

Proposition 9.7
Suppose S is a multiplicative subset of an integral domain R. Let T be the subset of R consisting of all r in R such that r divides s for some s in S. Then:

(a) T is a multiplicative subset of R containing S.
(b) $R_T = R_S$.
(c) $I_T = I_S$ for all ideals I of R.
(d) $Cl_S(I) = Cl_T(I)$ for all ideals I or R.
(e) $PPD_S(R) = PPD_T(R)$.
(f) The maps $PPD_S(R) \to PPD(R_S)$ and $PPD_T(R) \to PPD(R_T)$ are the same.

PROOF: (a) Left as an exercise.

(b) Because $T \supset S$ we know that $R_T \supset R_S$. Therefore, we must show that $R_T \subset R_S$. To do this, it suffices to show that $1/t$ is in R_S for all t in T. But if t is in T, then rt is in S for some r in R. Because $1/t = r/rt$ and r/rt is in R_S, it follows that $1/t$ is in R_S, which is our desired result.

(c), (d), and (f) are left as exercises.

We are now in position to prove the following.

Proposition 9.8
If S is a multiplicative subset of the unique factorization domain R, then the map

$$PPD_S(R) \to PPD(R_S)$$

is an isomorphism.

PROOF: Suppose S is a multiplicative subset of R and T is the multiplicative set of R consisting of all t in R such that $t|s$ for some s in S. By Proposition 9.7, because we know that $PPD_S(R) = PPD_T(R)$, that $R_S = R_T$, and the maps $PPD_S(R) \to PPD(R_S)$ and $PPD_T(R) \to PPD(R_T)$ are the same, it suffices to show that $PPD_T(R) \to PPD(R_T)$ is an isomorphism. Clearly, T has the property that if r in R divides something in T, then r is in T. Hence, the fact that R is a UFD implies that T is generated by primes. Also, the set of principal ideals of R is noetherian because R is a UFD. Hence, by Proposition 9.6, the map $PPD_T(R) \to PPD(R_T)$ is an isomorphism.

Earlier we showed that if R is a unique factorization domain and S is a multiplicative set in R, then R_S is a unique factorization domain. Under these circumstances it is tempting to ask if R is a unique factorization domain just because there is a multiplicative set S such that R_S is a unique factorization domain. Although the answer for arbitrary S is no (see the exercises for examples), there are special rings and special sorts of multiplicative sets for which the answer is yes, as we now show in the following.

Theorem 9.9
Suppose R is an integral domain whose set of principal ideals is noetherian. Also, suppose S is a multiplicative subset of R generated by primes. If R_S is a unique factorization domain, then R is a unique factorization domain.

PROOF: Let T be the multiplicative set of R consisting of all r in R which divide elements in S. Since every element of S is a finite product of units and prime elements in R, it is obvious that T has the property that T is generated by primes. Therefore, because the set of principal ideals in R is noetherian, we know that $PPD_T(R) \to PPD(R_T)$ is an isomorphism. Because $R_T = R_S$, we also know that R_T is a unique factorization domain. We now show that these facts together imply that R is a unique factorization domain.

Let V be the elements in R which can be written as a finite product of units and primes in R. Then clearly V is a multiplicative set of R containing T. By Proposition 8.1, we know that in order to show that R is a unique factorization domain, it suffices to show that $Q(R) = R_V$.

To do this, we first observe that since a prime element in R is either a unit or a prime element in R_T, the multiplicative subset V of R_T has the property that every element of V is a finite product of units and prime elements in R_T. Further, the fact that $PPD_T(R) \to PPD(R_T)$ is surjective implies that given any prime element y in R_T, there is a prime element r in R (and hence in V) such that $yR_T = rR_T$ or, equivalently, there is a unit z in R_T such that $y = zr$. Hence, every nonzero element of R_T can be written as zv with z a unit in R_T and v in V. This implies that $(R_T)_V = Q(R_T) = Q(R)$. Because $V \supset T$, we know that $VT = V$. Hence, $Q(R) = (R_T)_V = R_{TV} = R_V$, which gives us our desired conclusion that $R_V = Q(R)$ or, equivalently, R is a unique factorization domain.

10. WHEN $R[X]$ IS A UFD

In this section we show how the criterion established in Theorem 9.9 for when an integral domain is a unique factorization domain can be used to show that if R is a UFD, then $R[X]$ is a UFD. The proof will proceed in several steps. First of all we show that because R is a unique factorization domain, the set of principal ideals in $R[X]$ is noetherian, something which must be true if $R[X]$ is to be a unique factorization domain. What we actually show is a little more general, namely, the following.

Lemma 10.1
Suppose R is an integral domain whose set of principal ideals is noetherian. Then the set of principal ideals of $R[X]$ is also noetherian.

PROOF: Suppose

$$(a_1(X)) \subset (a_2(X)) \subset \cdots \subset (a_n(X)) \subset \cdots$$

is an ascending chain of principal ideals in $R[X]$. We want to show that for some integer $m \geq 0$ we have $(a_n(X)) = (a_m(X))$ for all $n \geq m$. Clearly, we can assume without loss of generality that all the $a_n(X) \neq 0$.

Because

$$a_2(X) | a_1(X),\ a_3(X) | a_2(X), \ldots, \text{etc.}$$

we have that $\deg(a_1(X)) \geq \deg(a_2(X)) \geq \cdots \geq \deg(a_n(X)) \geq \cdots$. Hence, there is an integer $h \geq 1$ such that $\deg(a_i(X)) = \deg(a_n(X))$ for all $i \geq h$. So again without loss of generality we can assume that the degrees of the $a_i(X)$ are all the same. The fact that $\deg(a_{n+1}(X)) = \deg(a_n(X))$ and $a_{n+1}(X) | a_n(X)$ implies that there is a nonzero r_{n+1} in R such that $r_{n+1}a_{n+1}(X) = a_n(X)$. In particular, if we let b_n be the leading coefficient of $a_n(X)$ for each n, we have the ascending chain of principal ideals in R

$$b_1R \subset b_2R \subset \cdots \subset b_nR \subset \cdots$$

because $r_{n+1}b_{n+1} = b_n$ for all n. Because the set of principal ideals of R is noetherian, we know there is an integer $m \geq 1$ such that $b_nR = b_mR$ for all $n \geq m$. Hence, $r_{m+1}b_{m+1}R = b_{m+1}R$, which implies r_{m+1} is a unit in R. A similar argument shows that r_{m+j} is a unit in R for all $j \geq 1$. Therefore, the ideals $(a_{m+j}(X))$ are all the same for integers $j \geq 0$, which shows that $(a_n(X)) = (a_m(X))$ for all $n \geq m$. Hence, the set of principal ideals in $R[X]$ is noetherian if the set of principal ideals in R is noetherian.

Next we want to show that if p is a prime element in R, then p is also a prime element in $R[X]$. This follows from the following general lemma.

Lemma 10.2
Let I be a prime ideal in the arbitrary commutative ring R. Then the ideal $IR[X]$ of $R[X]$ generated by the set I is a prime ideal of $R[X]$. Actually $R[X]/IR[X]$ is isomorphic to the integral domain $(R/I)[X]$.

PROOF: Let $g: R \to R/I$ be the canonical surjective morphism. Then $(R/I)[X]$ is an integral domain and we know that the morphism of rings $g': R \to (R/I)[X]$

which is the composition $R \xrightarrow{f} R/I \xrightarrow{\text{inc}} (R/I)[X]$ can be extended to a unique morphism $f: R[X] \to (R/I)[X]$ with the property that $f(X) = X$ (see Chapter 4, Proposition 2.2). More precisely, the map $f: R[X] \to (R/I)[X]$ given by $f(\Sigma_{i \in N} a_i X^i) = \Sigma_{i \in N} g(a_i) X^i$ is a morphism of rings. Clearly, f is a surjective morphism of rings. Because $f(\Sigma a_i X^i) = 0$ if and only if each a_i is in I, it follows that $\text{Ker } f = IR[X]$. Therefore, f induces an isomorphism $R[X]/IR[X] \to (R/I)[X]$, which shows that $R[X]/IR[X]$ is an integral domain because $(R/I)[X]$ is an integral domain. Thus, $IR[X]$ is a prime ideal in $R[X]$.

Before proving our main theorem, we need one more preliminary result. Suppose R is an integral domain. Then so is $R[X]$. Because $R \subset Q(R)$, we know that $R[X] \subset Q(R)[X]$. Clearly, $Q(R)[X] \subset Q(R[X])$, because $R - \{0\} \subset R[X] - \{0\}$, and so $Q(R[X]) = Q(Q(R)[X])$. Now let S be a multiplicative set in R. Because $R_S \subset Q(R)$, we have that $R_S[X] \subset Q(R)[X] \subset Q(R[X])$. On the other hand, S is also a multiplicative set in $R[X]$ because $R \subset R[X]$. Thus, the ring $R[X]_S$ is also contained in $Q(R[X])$. We claim that $R_S[X]$ and $R[X]_S$ are the same subring of $Q(R[X])$ as we now show.

An element of $R[X]_S$ can be written as $(\Sigma_{i \in N} a_i X^i)/s$ for some $\Sigma_{i \in N} a_i X^i$ in $R[X]$ and s in S. But this is clearly the same thing: $\Sigma_{i \in N} a_i/sX^i$ in $Q(R[X])$. Because $\Sigma_{i \in N} a_i/sX^i$ is in $R_S[X]$, it follows that $R[X]_S \subset R_S[X]$.

An element of $R_S[X]$ can be written as $\Sigma_{i \in N} a_i/s_i X^i$. Suppose $\deg(\Sigma_{i \in N} a_i/s_i X^i) = n$. Then $a_i = 0$ for all $i > n$. Let $s = \prod_{i=0}^{n} s_i$ and let $t_j = s/s_j$ for all $j = 0, \ldots, n$. If we set $b_i = a_i t_i$ for $i = 0, \ldots, n$ and $b_i = 0$ for all $i > n$, then $\Sigma_{i \in N} b_i/sX^i = \Sigma_{i \in N} (a_i/s_i) X^i$ in $Q(R[X])$. Because $\Sigma_{i \in N} b_i/sX^i$ also equals $(\Sigma_{i \in N} b_i X^i)/s$ which is in $R[X]_S$, we have that $R_S[X] \subset R[X]_S$. This finishes the proof of the lemma.

Lemma 10.3
Let S be a multiplicative subset of the integral domain R. Then the subrings $R_S[X]$ and $R[X]_S$ of $Q(R[X])$ are the same.

Putting together these preliminary results with our previous results about unique factorization domains, we obtain the following.

Theorem 10.4
If R is a unique factorization domain, then so is $R[X]$.

PROOF: Because R is a UFD, we know that its set of principal ideals is noetherian. Therefore, we know by Lemma 10.1 that the set of principal ideals in $R[X]$ is noetherian. Hence, if we can find a multiplicative subset S of $R[X]$ such that (a) S is generated by primes and (b) $R[X]_S$ is a unique factorization domain, then it follows from Theorem 9.9 that $R[X]$ is a unique factorization domain.

Let S be the multiplicative set of R consisting of all nonzero elements in R. Because R is a unique factorization domain, S is generated by primes. Because R is a subring of $R[X]$, we know that the units in R are also units in $R[X]$. But we have also shown that prime elements in R are also prime elements in $R[X]$. Therefore, viewing S as a multiplicative set in $R[X]$, it follows that the multiplicative

subset S of $R[X]$ is generated by primes. Hence, if we show that $R[X]_S$ is a UFD, then we will have finished the proof that $R[X]$ is a UFD.

But we have already seen that $R[X]_S = R_S[X]$. Because S is the set of all nonzero elements of R, we have that $R_S = Q(R)$ and so $R[X]_S = Q(R)[X]$. But $Q(R)[X]$ is a principal ideal domain, and hence a unique factorization domain, since $Q(R)$ is a field. Therefore, $R[X]_S$ is a unique factorization domain, which finishes the proof that $R[X]$ is a unique factorization domain.

EXERCISES

(1) Let R be an integral domain.

(a) Show by induction on n that $R[X_1, \ldots, X_n]$ is an integral domain for all positive integers n.

(b) Show that if I is any set, then $R[X_i]_{i \in I}$ is an integral domain. [*Hint*: Show that if $f_1, \ldots, f_m$ is any finite set of elements of $R[X_i]_{i \in I}$, then there is a finite subset J of I such that the image of $R[X_j]_{j \in J}$ in $R[X_i]_{i \in I}$ under the usual injection morphism $R[X_j]_{j \in J} \to R[X_i]_{i \in I}$ contains $f_1, \ldots, f_m$.

(c) Show that if J is a subset of I and $f(X)$ is a prime element in $R[X_j]_{j \in J}$, then the image of $f(X)$ in $R[X_i]_{i \in I}$ is also a prime element in $R[X_i]_{i \in I}$.

(d) Prove by induction on n that if R is a unique factorization domain, then so is $R[X_1, \ldots, X_n]$ for all positive integers n.

(e) Prove that if R is a unique factorization domain, then so is $R[X_i]_{i \in I}$ for any set I.

(f) Prove that if $R[X_i]_{i \in I}$ is a unique factorization domain for some nonempty set I, then R is a unique factorization domain.

(2) Let R be an arbitrary nonzero commutative ring.

(a) Show that $R[X]$ is not an artinian ring.

(b) Show that $R[X_i]_{i \in I}$ is not a noetherian ring if I is an infinite set.

(3) Let K be a field and $f(X)$ an element of $K[X]$. An element α in K is said to be a root of the polynomial $f(X)$ if $f(\alpha) = 0$.

(a) Show that α in K is a root of the polynomial $f(X)$ if and only if $(X - \alpha) | f(X)$ in $K[X]$. [*Hint*: Using the Euclidean algorithm write $f(X) = q(X)(X - \alpha) + r(X)$ where either $r(X) = 0$ or $\deg r(X) < \deg (X - \alpha)$.]

(b) Show that if $\deg f(X) = n$, then $f(X)$ has no more than n roots in K.

(4) Suppose K is a field and G is a finite subgroup of the multiplicative group of nonzero elements of K. Prove that G is a cyclic group. [*Hint*: Use the preceding exercise together with Exercise 29 of Chapter 2.]

(5) Show that if K is a finite field, then the group of units in K is a cyclic group. In particular, if p is a prime integer, then $U(\mathbf{Z}/p\mathbf{Z})$ is a cyclic group of order $p - 1$.

If R is a commutative ring and I is an ideal of R, we shall write $x \equiv y(I)$ to mean $x - y$ is in I. If I is a principal ideal generated by an element m, we shall write $x \equiv y'(m)'$ instead of $x \equiv y((m))$.

(6) Let p be a positive prime element in the ring of integers, $\mathbf{Z}$.

(a) Prove that $\binom{p}{k} \equiv 0(p)$ for $1 < k < p$ where $\binom{p}{k}$ is the binomial coefficient.

[*Hint*: Recall that

$$\binom{p}{k} = \frac{p!}{k!(p-k)!}$$

and that $\mathbf{Z}$ is a UFD.

(b) If x and y are integers, prove that $(x+y)^{p^n} \equiv x^{p^n} + y^{p^n}(p)$ for any n in $\mathbf{N}$.

(c) Prove that if x and y are integers such that $x \equiv y(p^n)$, then $x^p \equiv y^p(p^{n+1})$ for any n in $\mathbf{N}$.

(d) Prove that if p is an odd prime, then $(1+p)^{p^{n-2}} \equiv 1 + p^{n-1}(p^n)$ for all $n \geq 2$. [*Hint*: Use induction on n.]

The purpose of the next exercise is to outline a proof of the fact that the group of units in the ring $\mathbf{Z}/p^n\mathbf{Z}$ is a cyclic group of order $(p-1)p^{n-1}$ for all odd prime numbers p in $\mathbf{Z}$ and all positive integers n.

(7) Let p be a prime number in $\mathbf{Z}$ and n a positive integer.

(a) Show that every integer z in $[0, p^n)$ can be written in one and only one way as the sum $a_0 + a_1 p + \cdots + a_{n-1}p^{n-1}$ where the a_i are in $[0, p)$.

(b) Let $k: \mathbf{Z} \to \mathbf{Z}/p^n\mathbf{Z}$ be the canonical surjective morphism of rings. Show that $k|[0, p^n): [0, p^n) \to \mathbf{Z}/p^n\mathbf{Z}$ is a bijective morphism of sets.

(c) Show that $k(a_0 + a_1 p + \cdots + a_{n-1}p^{n-1})$ is a unit in $\mathbf{Z}/p^n\mathbf{Z}$ if and only if $a_0 \neq 0$. Hence, card $(U(\mathbf{Z}/p^n\mathbf{Z})) = (p-1)(p^{n-1})$.

(d) Let $k': \mathbf{Z}/p^n\mathbf{Z} \to \mathbf{Z}/p\mathbf{Z}$ be the canonical surjective map of rings. Show that if x is in $U(\mathbf{Z}/p^n\mathbf{Z})$, then $k'(x)$ is in $U(\mathbf{Z}/p\mathbf{Z})$. Also show that the induced map $f: U(\mathbf{Z}/p^n\mathbf{Z}) \to U(\mathbf{Z}/p\mathbf{Z})$ given by $f(x) = k'(x)$ is a surjective morphism of groups.

(e) Show that Ker f consists precisely of all the elements $k(a_0 + a_1 p + \cdots + a_{n-1}p^{n-1})$ with $a_0 = 1$. Thus, card (Ker f) $= p^{n-1}$.

(f) Assume, now, that p is an odd prime. Show that Ker f is a cyclic group by showing that the order of $k(1+p)$ in $U(\mathbf{Z}/p^n\mathbf{Z})$ is p^{n-1}. [*Hint*: Use the fact that $(1+p)^{p^{n-2}} \equiv 1 + p^{n-1}(p^n)$.]

(g) Prove that $U(\mathbf{Z}/p^n\mathbf{Z})$ is cyclic of order $(p-1)p^n$. [*Hint*: Use the fact that Ker f is a cyclic group of order p^{n-1} and $U(\mathbf{Z}/p^n\mathbf{Z})/\text{Ker } f \simeq U(\mathbf{Z}/p\mathbf{Z})$ is a cyclic group of order $p-1$.]

(h) Is $U(\mathbf{Z}/8\mathbf{Z})$ cyclic? More generally, is $U(\mathbf{Z}/2^n\mathbf{Z})$ cyclic if $n \geq 3$?

(8) Let R be a commutative ring, $\mathfrak{P}$ a prime ideal of R, and $f(X) = X^n + a_1 X^{n-1} + \cdots + a_0$ a polynomial in $R[X]$ such that all the a_i are in $\mathfrak{P}$ but a_0 is not in $\mathfrak{P}^2$. Show that $f(X)$ is an irreducible element of $R[X]$. [*Hint*: Suppose $f(X)$ is not irreducible. Then $f(X) = g(X)h(X)$ with neither $g(X)$ nor $h(X)$ units in $R[X]$. Then show that either $g(0)$ or $h(0)$ is in $\mathfrak{P}$. Finally, show that if $g(0)$ is in $\mathfrak{P}$, then every coefficient of $g(X)$ is in $\mathfrak{P}$, which is impossible. This result is known as **Eisenstein's irreducibility criterion.**

(9) Let K be a field and $f: K \to R$ a K-algebra with R not the zero ring. Show:

(a) $f: K \to R$ is an injective morphism of rings.

(b) The additive group of R can be viewed as a K-vector space by defining $kr = f(k)\,r$ for all k in K and r in R. This is the only way we consider R as a K-vector space.

(c) Show that for each r in R, the map $l_r: R \to R$ given by $l_r(x) = rx$ is a linear transformation of the K-vector space R. Further, show that $l_{r_1} = l_{r_2}$ if and only if $r_1 = r_2$.

(d) Let V be a vector space over K. Show that the following data define a K-algebra called the **K-endomorphism ring of** V and is denoted by $\mathrm{End}_K(V)$.
  (i) As a set $\mathrm{End}_K(V)$ consists of all the linear transformations $f: V \to V$.
  (ii) The addition in $\mathrm{End}_K(V)$ is given by $(f+g)(v) = f(v)+g(v)$ for all f, g in $\mathrm{End}_K(V)$.
  (iii) The multiplication in $\mathrm{End}_K(V)$ which is written as $f \cdot g$ is the composition of the linear transformations g followed by f.
  (iv) The ring morphism $K \to \mathrm{End}_K(V)$ which makes $\mathrm{End}_K(V)$ a K-algebra is given by $k \mapsto f_k$ where $f_k: V \to V$ is the linear transformation $f_k(v) = kv$ for all v in V.

(e) Show that the map $g: R \to \mathrm{End}_K(\Lambda)$ given by $g(r) = l_r$ for each r in R is an injective K-algebra morphism.

(f) Show that if V is a finite-dimensional vector space over K of dimension n, then the K-algebras $\mathrm{End}_K(V)$ and $M_n(K)$ are isomorphic K-algebras which as vector spaces over K are of dimension n^2. [*Hint*: Choose a basis $v_1, \ldots, v_n$. Define the map $\alpha: \mathrm{End}_K(V) \to M_n(K)$ by $\alpha(f)$ as the matrix corresponding to f with respect to the basis $v_1, \ldots, v_n$. Show that α is an isomorphism of K-algebras.]

(g) Show that if R is a K-algebra whose dimension as a vector space over K is finite, say n, then R is isomorphic as a K-algebra to a K-subalgebra of $M_n(K)$.

(h) Give an example of a K-algebra R whose dimension as a vector space over K is n and which is isomorphic as a K-algebra to a K-subalgebra of $M_m(K)$ with $m < n$.

(10) Let K be a field and consider $K[X]$ a K-algebra in the usual way, that is, by means of the ring morphism $K \to K[X]$ defined by $a \mapsto \Sigma a_i X^i$ where $a_0 = a$ and $a_i = 0$ for $i > 0$. If $f(X)$ is a polynomial in $K[X]$, we consider $K[X]/f(X) \times K[X]$ a K-algebra by means of the composite ring morphism $K \to K[X] \xrightarrow{k} F[X]/f(X)K[X]$ where k is the usual canonical surjective ring morphism.

(a) Show that $K \to K[X]/f(X)K[X]$ is an injective morphism unless $f(X)$ is a nonzero constant, that is, $f(X) = \Sigma a_i X^i$ with $a_0 \neq 0$ and $a_i = 0$ for $i > 0$.

(b) Suppose $\deg f(X) = n$. Show that the K-algebra $K[X]/f(X)K[X]$ is an n-dimensional vector space over K. [*Hint*: Show that $k(X^0), \ldots, k(X^{n-1})$ are a basis for $K[X]/f(X)K(X)$ as a vector space over K.]

(c) Let $f(X)$ be an irreducible polynomial over $K[X]$. Recall that the K-algebra $L = K[X]/f(X)K[X]$ is a field. If we identify K with its image in L, the polynomial $f(X)$ can then be considered as an element of $L[X]$. Show that the polynomial $f(X)$ has a root in L. [*Hint*: Show that the element $k(X)$ in L is a root of $f(X)$.]

(d) Show that if $f(X)$ is an arbitrary polynomial, then there is a field L containing K as a subfield such that L is a finite-dimensional vector space over K and $f(X)$ has a root in L.

(11) Let K be a finite field; for instance, $K = \mathbf{Z}/p\mathbf{Z}$ where p is a prime in $\mathbf{Z}$.

(a) For each integer $n \geq 0$, show that there is an irreducible polynomial $f(X)$ in $K[X]$ with $\deg f(X) > n$.

(b) Show that if n is any positive integer, there is a finite field $L \supset K$ with $\mathrm{card}(L) > n$.

(c) Let K be a finite field. Show that the unique ring morphism $f:\mathbf{Z}\to K$ given by $f(z)=z\cdot 1$ for each z in $\mathbf{Z}$ has $\operatorname{Ker} f=p\mathbf{Z}$ for some positive prime element p. This prime number p is called **the characteristic of the field** K. Hence, the field K is a $\mathbf{Z}/p\mathbf{Z}$-algebra where p is the characteristic of the field K.
(d) Show that $\operatorname{card}(K)=p^n$ for some positive integer n.
(e) Show that $pk=0$ for all k in K.
(f) Show that if x and y are arbitrary elements of K, then $(x+y)^p=x^p+y^p$. More generally, $(x+y)^{p^n}=x^{p^n}+y^{p^n}$ for all integers n.
(g) Show that if K is of characteristic p, then there is an irreducible element $f(X)$ in $\mathbf{Z}/p\mathbf{Z}[X]$ such that K is isomorphic to $(\mathbf{Z}/p\mathbf{Z}[X])/f(X)(\mathbf{Z}/p\mathbf{Z}[X])$. [*Hint*: Use the fact that $U(K)$ is a cyclic group.]
(h) Show that if K is a finite field of characteristic p, then for each integer n and each k in K, the polynomial $X^{p^n}-k$ has at least one root in K. [*Hint*: Use the fact that for each integer n, the map $f:K\to K$ given by $f(k)=k^{p^n}$ is a morphism of rings.]

(12) Let K be a field and consider the unique ring morphism $f:\mathbf{Z}\to K$ given by $f(z)=z\cdot 1$.
(a) Show that $\operatorname{Ker} f=(n)$ where n is either 0 or a positive prime number. The number n is called the **characteristic** of the field K.
(b) Suppose the characteristic of the field K is zero. Show that there is a unique morphism of rings $\mathbf{Q}\to K$ where $\mathbf{Q}$ is the field of rational numbers. This unique morphism of rings is injective and one usually identifies $\mathbf{Q}$ with its image in K by means of this unique morphism.
(c) If the characteristic of K is p, then show that there is a unique morphism of rings $\mathbf{Z}/p\mathbf{Z}\to K$. Show that this unique morphism is an injective morphism. One usually identifies $\mathbf{Z}/p\mathbf{Z}$ with its image in K by means of this unique morphism.
(d) Show that if the characteristic of K is p, then $(x+y)^{p^n}=x^{p^n}+y^{p^n}$ for all integers n.

(13) Let K be a field. Then the field of quotients of $K[X]$ is called **the field of rational functions in one variable over** K, or more simply, the rational function field over K, and is usually denoted by $K(X)$. $K(X)$ is considered a K-algebra by means of the composition of ring morphisms $K\to K[X]\to K(X)$ where $K[X]\to K(X)$ is the usual inclusion of an integral domain into its field of quotients. K is usually identified with its image in $K(X)$ by means of this injective ring morphism.
(a) Show that the characteristic of $K(X)$ is the same as the characteristic of K.
(b) Show that $K(X)$ is always an infinite-dimensional vector space over K.
(c) Show that there is always an injective morphism $f:K(X)\to K(X)$ of K-algebras which is not surjective. [*Hint*: Let $f:K[X]\to K(X)$ be the uniquely determined K-algebra morphism with the property $f(X)=X^2$. Show that there is a unique morphism $g:K(X)\to K(X)$ of K-algebras such that $g|K[X]=f$. Prove that the morphism g is injective but not surjective.]
(d) Suppose K is a field of characteristic $p\neq 0$. Show that the polynomial $t^{p^n}-X$ in $K(X)[t]$ has no solution for any $n\geq 1$. Hence, the ring morphisms $K(X)\to K(X)$ given by $x\mapsto x^{p^n}$ for all x in $K(X)$ are injective but not surjective ring mophisms for all $n\geq 1$.

(14) Prove Basic Properties 1.1 and 1.2.

(15) Prove that the product of two ideals I and J in a commutative ring R, as defined in Section 1, is an ideal in R.

(16) Let K be a field, let $S = K[X, Y, U, V]$, and let I be the ideal in S generated by the elements $X - UY$ and $Y - VX$. Define R to be S/I.

(a) Prove that the principal ideals $(k(X))$ and $(k(Y))$ are equal in R where $k: S \to R$ is the canonical ring surjection.

(b) Let $f: R \to PI(R)$ be the monoid morphism described in Section 1. Use **(a)** to show that $xU(R) \neq [x]$ where $x = k(X)$, $U(R)$ is the group of units of R, and $[x]$ is the unique element of Coim f containing x.

(17) Give a detailed proof of Basic Properties 3.2.

(18) Let R be an integral domain, x an irreducible element of R, and y an element of R not divisible by x. Prove that if x and y have a least common multiple, then it must be (up to a unit factor) the element xy.

(19) Let K be a field and let $R = K[X, Y, Z]/I$ where I is the principal ideal generated by the polynomial $X^2 - YZ$.

(a) Prove that $X^2 - YZ$ is a prime element in $K[X, Y, Z]$ and hence, I is a prime ideal.

(b) Let $k: K[X, Y, Z] \to R$ be the canonical surjective morphism. Prove that $k(X)$ is irreducible in R, but show that $k(X)$ is not a prime element in R. [*Hint*: Observe that $k(X^2) = k(Z)k(Y)$.]

(c) Prove that $k(X)$ and $k(Y)$ do not have a least common multiple in R.

(d) Let $S = R - \{0\}$. Show that R_S is a UFD even though R is not a UFD.

(20) Prove Basic Properties 4.3.

(21) Prove Basic Properties 4.4.

(22) Prove Basic Properties 5.10.

(23) Let $\mathbf{Z}$ be the ring of integers and let S be the set of all odd integers. If I is the ideal in $\mathbf{Z}$ consisting of all multiples of 6, that is, $I = (6)$, show that $I_S \cap \mathbf{Z} \neq I$. In fact, prove that $I_S \cap \mathbf{Z} = (2)$.

(24) Let $R = K[X, Y, Z]/(X^2 - YZ)$ be the ring of Exercise 19, with $k: K[X, Y, Z] \to R$ the canonical surjective ring morphism.

(a) Prove that (X, Y) is a prime ideal in $K[X, Y, Z]$ containing $(X^2 - YZ)$.

(b) Prove that $\mathfrak{P} = (k(X), k(Y))$ is a prime ideal in R.

(c) Can $\mathfrak{P}$ be generated by a single element?

(d) Let $S = R - \mathfrak{P}$. Prove that S is a multiplicative subset of R.

(e) Show that the ideal $\mathfrak{P}R_S$ is principal in R_S. [*Hint*: Show that $\mathfrak{P}R_S = k(X)R_S$.]

(f) Show that the natural map $PPD_S(R) \to PPD(R_S)$ is not surjective. [*Hint*: Consider the principal prime divisor $k(X)R_S$ in $PPD(R_S)$.]

Chapter 6 GENERAL MODULE THEORY

Earlier, in discussing the group $\mathbf{Z}$ of integers under addition, we showed that for each element a in the abelian group A, there is one and only one group morphism $f_a: \mathbf{Z} \to A$ such that $f_a(1) = a$. Also, for each a in A and n in $\mathbf{Z}$ we defined na by $na = f_a(n)$. We then showed that viewing $\mathbf{Z}$ as the ring of integers, the map $\mathbf{Z} \times A \to A$ given by $(n, a) \mapsto na$ for all n in $\mathbf{Z}$ and a in A has the following properties:

(1) $(n_1 + n_2)a = n_1 a + n_2 a$.
(2) $n(a_1 + a_2) = na_1 + na_2$.
(3) $(n_1 n_2)a = n_1(n_2 a)$.
(4) $1a = a$.

The reader should have no difficulty in seeing that properties **(1)** and **(4)** alone guarantee that the map $\mathbf{Z} \times A \to A$ we just described is the only map from $\mathbf{Z} \times A$ to A satisfying properties **(1)** through **(4)**.

Vector spaces over a field K give another example of a similar structure. We recall that a vector space V over a field K consists of an abelian group V together with a map $K \times V \to V$, usually described by $(k, v) \mapsto kv$, satisfying the following conditions:

(1) $(k_1 + k_2)v = k_1 v + k_2 v$,
(2) $k(v_1 + v_2) = kv_1 + kv_2$,
(3) $(k_1 k_2)v = k_1(k_2 v)$,
(4) $1v = v$,

for all k, k_1, k_2 in K and v, v_1, v_2 in V.

The striking similarity of these structures suggests that they are simply examples of a single general notion. The following definitions show that this is indeed the case.

Definitions
Let R be a ring.

(a) By an ***R*-module structure** on an abelian group M we mean a map $R \times M \to M$, which we denote by $(r, m) \mapsto rm$ for all r in R and m in M, satisfying:
  (i) $(r_1 + r_2)m = r_1m + r_2m$,
  (ii) $r(m_1 + m_2) = rm_1 + rm_2$,
  (iii) $(r_1r_2)m = r_1(r_2m)$,
  (iv) $1m = m$,
  for all m, m_1, m_2, in M and r, r_1, r_2 in R.

(b) An ***R*-module** consists of an abelian group M together with an R-module structure $R \times M \to M$ on M.

Our previous remarks show that each abelian group has a unique $\mathbf{Z}$-module structure over the ring $\mathbf{Z}$ of integers. Because each $\mathbf{Z}$-module is also an abelian group, we see that $\mathbf{Z}$-modules and abelian groups are essentially the same thing.

Our previous remarks also show that vector spaces are nothing more than R-modules where R is a field. However, unlike the situation for $\mathbf{Z}$, it is perfectly possible for a given abelian group M to have more than one R-module structure if R is not the integers. For example, suppose $R = \mathbf{C}$, the field of complex numbers. Then it is easily checked that the two maps $f : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$ and $g : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$ defined by $f(z_1, z_2) = z_1z_2$ and $g(z_1, z_2) = \bar{z}_1z_2$, where $\bar{z}$ is the complex conjugate of z, are different $\mathbf{C}$-module structures on the abelian group consisting of the additive group $\mathbf{C}$ of the field $\mathbf{C}$.

Before pointing out other types of modules with which the reader has some familiarity, we make the following notational convention. If M stands for an R-module, then we will use the same letter M to denote the underlying abelian group of the R-module M.

Example 1 Let R be a ring. Denoting the additive group of R by R, it is easily checked that the map $R \times R \to R$ given by $(r_1, r_2) \mapsto r_1r_2$ for all r_1 and r_2 in R, is an R-module structure on R. Hence, associated with a ring R is the R-module structure $R \times R \to R$ given by $(r_1, r_2) \mapsto r_1r_2$. This R-module is usually denoted simply by R.

Example 2 Let R be a ring. We recall that an ideal I of R is a subgroup of R with the property that ri and ir are in I for all r in R and i in I. If I is an ideal in R, it is easily checked that the map $R \times I \to I$ given by $(r, i) \mapsto ri$ for all r in R and i in I is an R-module structure on the additive group of I. Hence, associated with each ideal I of R is the R-module consisting of the additive group of I together with the R-module structure $R \times I \to I$ given by $(r, i) \mapsto ri$. This R-module will usually be denoted also by I.

Example 3 Suppose $f : R \to R'$ is a morphism of rings. Then it is easily checked that the map $R \times R'$ given by $(r, r') \mapsto f(r)r'$ for all r in R and r' in R' is an

R-module structure on the additive group R' of the ring R'. Hence, associated with each ring morphism $f: R \to R'$ is the R-module consisting of the additive group R' of the ring R' together with the R-module structure $R \times R' \to R'$ given by $(r, r') \mapsto f(r)r'$. In particular, if I is an ideal of R, the canonical morphism of rings $k_{R/I}: R \to R/I$ gives an R-module structure on the additive group R/I. This R-module will be denoted simply by R/I.

Thus, we see that a great many of the mathematical objects familiar to the reader are either modules or have modules associated with them in a fairly obvious and simple-minded way. This list by no means exhausts the types of mathematical objects that can be viewed as modules.

Because of this wide prevalence of modules in much of algebra, the theory of modules occupies a large place in this book. This chapter is devoted for the most part to generalities concerning arbitrary modules over arbitrary rings together with illustrative examples and applications. In succeeding chapters specific situations will be studied, such as modules over semisimple rings, principal ideal domains, and Dedekind domains.

1. CATEGORY OF MODULES OVER A RING

Our main concern in this section is to study the elementary properties of the collection of R-modules for a fixed ring R. After discussing the morphisms of R-modules, we show that the collection of R-modules together with these morphisms form a category. The rest of the section is then devoted to studying the basic properties of the category of R-modules.

As with the other mathematical objects we have considered, we begin our study of modules by deciding how to compare them. Because an R-module M is an abelian group M together with an R-module structure on M, it is clear that a morphism $f: M_1 \to M_2$ from the R-module M_1 to the R-module M_2 should be a morphism of the underlying groups of M_1 and M_2 which is compatible with the R-module structures on M_1 and M_2. Stated more precisely, we have the following.

Definition
Let M_1 and M_2 be R-modules. By an ***R*-module morphism** from M_1 to M_2 we mean a morphism of groups $f: M_1 \to M_2$ which satisfies $f(rm) = rf(m)$ for all r in R and all m in M_1. An R-module morphism will often be called simply an ***R*-morphism**.

The reader should have no difficulty seeing that if $R = \mathbf{Z}$, then a morphism $f: M_1 \to M_2$ of $\mathbf{Z}$-modules is nothing more than a morphism of groups. Also, it is obvious that if R is a field, then a morphism $f: M_1 \to M_2$ of R-modules is the same thing as a linear map of vector spaces.

We have the following easily verified analogs of results already obtained for sets, groups, and rings.

Basic Properties 1.1
Let R be a ring.

(a) For each R-module M, the identity map $\mathrm{id}_M: M \to M$ is a morphism of R-modules.
(b) If $f: M_1 \to M_2$ are $g: M_2 \to M_3$ are R-morphisms, then the composition of maps $gf: M_1 \to M_3$ is a morphism of R-modules.

These results suggest the following.

Definitions
Let R be a ring.

(a) For each R-module M, the morphism of R-modules $\mathrm{id}_M: M \to M$ is called the **identity morphism** of M.
(b) If $f: M_1 \to M_2$ and $g: M_2 \to M_3$ are morphisms of R-modules, then the **composition** $gf: M_1 \to M_3$ of f and g is the morphism from M_1 to M_3 given by the ordinary composition of f and g viewed as maps from M_1 to M_2 and M_2 to M_3, respectively.

As an immediate consequence of these definitions we have the following.

Basic Properties 1.2
Let R be a ring.

(a) The composition of morphisms of R-modules is associative, that is, if $f: M_1 \to M_2$, $g: M_2 \to M_3$, and $h: M_3 \to M_4$ are R-module morphisms, then $h(gf) = (hg)f$.
(b) For each R-module M the identity morphism $\mathrm{id}_M: M \to M$ has the following properties:
  (i) If $f: M \to M'$ is an arbitrary morphism of R-modules, then $f\,\mathrm{id}_M = f$.
  (ii) If $g: M' \to M$ is an arbitrary morphism of R-modules, then $\mathrm{id}_M g = g$.

Our discussion so far amounts to nothing more or less than the fact that the following data define a category. This category is called the category of R-modules and is usually denoted by $\mathrm{Mod}(R)$.

The objects of $\mathrm{Mod}(R)$ are the R-modules. For every pair of objects M_1 and M_2 in $\mathrm{Mod}(R)$ we define the set (M_1, M_2) of morphisms in $\mathrm{Mod}(R)$ from M_1 to M_2 to be $\mathrm{Hom}_R(M_1, M_2)$, the set of all R-module morphisms from the R-module M_1 to the R-module M_2. Next, we define the composition of morphisms in $\mathrm{Mod}(R)$. For all triples M_1, M_2, and M_3 of objects in $\mathrm{Mod}(R)$ define $(M_1, M_2) \times (M_2, M_3) \to (M_1, M_2)$ to be the map $\mathrm{Hom}_R(M_1, M_2) \times \mathrm{Hom}_R(M_2, M_3) \to \mathrm{Hom}_R(M_1, M_2)$ given by $(f, g) \mapsto gf$ where gf is the composition of the R-module morphisms $f: M_1 \to M_2$ and $g: M_2 \to M_3$. It is an immediate consequence of our previous observations that these data satisfy the axioms of a category, a fact we leave to the reader to verify.

One of the things that distinguishes the category $\mathrm{Mod}(R)$ from most of the categories we have considered previously is the fact that for each pair of objects M_1 and M_2 in $\mathrm{Mod}(R)$, the collection (M_1, M_2) of morphisms from M_1 to M_2 is not just a set but is an abelian group in a natural way. For it is not difficult to check that if $f, g: M_1 \to M_2$ are R-module morphisms, then the map $f+g: M_1 \to M_2$ defined by $(f+g)(m) = f(m) + g(m)$ for all m in M_1 is also an R-module morphism. Obviously, if f, g, and h are in $\mathrm{Hom}_R(M_1, M_2)$, then $f+(g+h) = (f+g)+h$ and $f+g = g+f$. Also, it is easily checked that the map $0: M_1 \to M_2$ given by $0(m) = 0$ for all m

in M_1 is an R-module morphism called the zero morphism from M_1 to M_2. Obviously, $0+f=f+0=f$ for all R-module morphisms $f: M_1 \to M_2$. Hence, the map $\mathrm{Hom}_R(M_1, M_2) \times \mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(M_1, M_2)$ given by $(f, g) \mapsto f+g$ is a law of composition on $\mathrm{Hom}_R(M_1, M_2)$ which makes $\mathrm{Hom}_R(M_1, M_2)$ a commutative monoid with the zero morphism as the identity element.

To see that this commutative monoid is actually a group, we observe that if $f: M_1 \to M_2$ is a morphism of R-modules, then the map $(-f): M_1 \to M_2$ defined by $(-f)(m) = -(f(m))$ for all m in M_1 is also an R-module morphism with the property $f+(-f)=(-f)+f=0$. Thus, $(-f)$ is the inverse of f for each element f in $\mathrm{Hom}_R(M_1, M_2)$, which finishes the proof that $\mathrm{Hom}_R(M_1, M_2)$ is an abelian group under the law of composition $(f, g) \mapsto f+g$ for all f and g in $\mathrm{Hom}_R(M_1, M_2)$. Since this law of composition plays an important role in studying modules, we make the following definition.

Definition

Let $f, g: M_1 \to M_2$ be R-module morphisms. We define their **sum** $f+g: M_1 \to M_2$ to be the R-module morphism given by $(f+g)(m) = f(m)+g(m)$ for all m in M_1. The abelian group consisting of $\mathrm{Hom}_R(M_1, M_2)$ together with the law of composition given by the sum of R-module morphisms will be called the **group of R-module morphisms** from M_1 to M_2.

Because the sets of morphisms $\mathrm{Hom}_R(M_1, M_2)$ in the category $\mathrm{Mod}(R)$ are abelian groups, it is natural to ask if the composition maps $\mathrm{Hom}_R(M_1, M_2) \times \mathrm{Hom}_R(M_2, M_3) \to \mathrm{Hom}_R(M_1, M_3)$ in $\mathrm{Mod}(R)$ are somehow compatible with the group structure in $\mathrm{Hom}_R(M_1, M_2)$, $\mathrm{Hom}_R(M_2, M_3)$, and $\mathrm{Hom}_R(M_1, M_3)$. That this is indeed the case is easily seen. Suppose that f and g are in $\mathrm{Hom}_R(M_1, M_2)$ and h and g are in $\mathrm{Hom}_R(M_2, M_3)$. A simple calculation shows that the morphism $h(f+g): M_1 \to M_3$ is the same as the morphism $hf+hg: M_1 \to M_3$. For

$$h(f+g)(m) = h((f+g)(m)) = h(f(m)+g(m)) = hf(m)+hg(m) = (hf+hg)(m)$$

for all m in M, which means that $h(f+g) = hf+hg$. Similarly, one can show that $(h+g)f = hf+gf$. Thus, we have established the following.

Basic Property 1.3

Let M_1, M_2, and M_3 be objects in $\mathrm{Mod}(R)$. The composition map $\psi: \mathrm{Hom}_R(M_1, M_2) \times \mathrm{Hom}_R(M_2, M_3) \to \mathrm{Hom}_R(M_1, M_3)$ has the following properties for all f_1, f_2 in $\mathrm{Hom}_R(M_1, M_2)$ and all g_1, g_2 in $\mathrm{Hom}_R(M_2, M_3)$:

(a) $\psi(f_1+f_2, g_1) = \psi(f_1, g_1)+\psi(f_2, g_1)$.
(b) $\psi(f_1, g_1+g_2) = \psi(f_1, g_1)+\psi(f_1, g_2)$.
(c) For a fixed g in $\mathrm{Hom}_R(M_1, M_2)$, the map $\alpha_g: \mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(M_1, M_3)$ given by $\alpha_g(f) = \psi(f, g)$ for all f in $\mathrm{Hom}_R(M_1, M_2)$ is a morphism of abelian groups.
(d) For a fixed f in $\mathrm{Hom}_R(M_1, M_2)$ the map $\beta_f: \mathrm{Hom}_R(M_2, M_3) \to \mathrm{Hom}_R(M_1, M_3)$ given by $\beta_f(g) = \psi(f, g)$ for all g in $\mathrm{Hom}_R(M_2, M_3)$ is a morphism of abelian groups.

The properties of the composition maps in $\mathrm{Mod}(R)$ just described are a special case of a general notion of considerable importance in algebra.

Definition
Let A, B, and C be abelian groups. A map $\psi: A \times B \to C$ is said to be a **bilinear map** from A and B to C if for all a_1, a_2 in A and b_1, b_2 in B we have:

(i) $\psi(a_1 + a_2, b_1) = \psi(a_1, b_1) + \psi(a_2, b_1)$.
(ii) $\psi(a_1, b_1 + b_2) = \psi(a_1, b_1) + \psi(a_1, b_2)$.

While we postpone a systematic development of the notion of bilinear maps until later, the reader should be able to get some preliminary feel for this subject by working out the following easily established properties.

Basic Properties 1.4
Let A, B, and C be abelian groups.

(a) A map $\psi: A \times B \to C$ is bilinear if and only if it satisfies both of the following conditions:
  (i) For each a in A, the map $\alpha_a: B \to C$ given by $\alpha_a(b) = \psi(a, b)$ for all b in B is a morphism of abelian groups.
  (ii) For each b in B, the map $\beta_b: A \to C$ given by $\beta_b(a) = \psi(a, b)$ for all a in A is a morphism of abelian groups.
(b) If $\psi: A \times B \to C$ is a bilinear map, then $\psi(na, b) = n\psi(a, b) = \psi(a, nb)$ for all n in $\mathbf{Z}$, a in A, and b in B. In particular, $\psi(0, b) = 0 = \psi(a, 0)$ for all a in A and b in B.
(c) The map $0: A \times B \to C$ given by $0(a, b) = 0$ for all a in A and b in B is bilinear and is called the **zero bilinear map**.
(d) If ψ_1, $\psi_2: A \times B \to C$ are bilinear maps, then the map $\psi_1 + \psi_2: A \times B \to C$ defined by $(\psi_1 + \psi_2)(a, b) = \psi_1(a, b) + \psi_2(a, b)$ is bilinear.
(e) If $\psi: A \times B \to C$ is bilinear and $f: C \to D$ is a morphism of abelian groups, then the composition $f\psi: A \times B \to D$ is bilinear.
(f) If $\psi: A \times B \to C$ is bilinear, then the map $(-\psi): A \times B \to C$ defined by $(-\psi)(a, b) = -(\psi(a, b))$ is bilinear.

As a consequence of these results, it is not difficult to see that the set $B(A \times B, C)$ of all bilinear maps from $A \times B$ to C is an abelian group, where the addition $\psi_1 + \psi_2$ is the bilinear map defined by $(\psi_1 + \psi_2)(a, b) = \psi_1(a, b) + \psi_2(a, b)$ for all a in A and b in B.

This leads to the following.

Definition
Let A, B, and C be abelian groups. The **sum** of two bilinear maps ψ_1, $\psi_2: A \times B \to C$ is the bilinear map $\psi_1 + \psi_2$ given by $(\psi_1 + \psi_2)(a, b) = \psi_1(a, b) + \psi_2(a, b)$ for all a in A and b in B. The abelian group $B(A \times B, C)$ consisting of the set of all bilinear maps from $A \times B \to C$ together with the addition given by the sum of bilinear maps is called the **group of bilinear maps** from $A \times B$ to C.

We end this preliminary discussion of bilinear maps of groups with the following examples.

Example 1.5 For each abelian group C the map $f: B(\mathbf{Z} \times \mathbf{Z}, C) \to C$ given by $f(\psi) = \psi(1, 1)$ for all ψ in $B(\mathbf{Z} \times \mathbf{Z}, C)$ is an isomorphism of abelian groups.

PROOF: It is left as an exercise to the reader to show that $f: B(\mathbf{Z}\times\mathbf{Z}, C)\to C$ is a morphism of abelian groups. Having this result, we show that f is injective by showing that Ker $f = 0$.

Suppose ψ is in Ker f or, what is the same thing, $\psi(1, 1) = 0$. We want to show that this implies that ψ is the zero element of $B(\mathbf{Z}\times\mathbf{Z}, C)$. By Basic Properties 1.4, because $\psi(n, 1) = n\psi(1, 1) = \psi(1, n)$ for all n in $\mathbf{Z}$, the fact that $\psi(1, 1) = 0$ implies $\psi(n, 1) = \psi(1, n) = 0$ for all n in $\mathbf{Z}$. But again by Basic Properties 1.4, we know that for each n in $\mathbf{Z}$ the map $\alpha_n: \mathbf{Z}\to C$ given by $\alpha_n(m) = \psi(n, m)$ is a morphism of groups. Because $\alpha_n(1) = \psi(n, 1) = 0$, it follows that $\alpha_n = 0$, because a morphism of groups from $\mathbf{Z}$ to C is completely determined by where it sends 1. Hence, for each n in $\mathbf{Z}$ we have that $\alpha_n(m) = \psi(n, m) = 0$ for all m in $\mathbf{Z}$. Therefore, $\psi(n, m) = 0$ for all n and m in $\mathbf{Z}$, which means that ψ is the zero element of $B(\mathbf{Z}\times\mathbf{Z}, C)$. Hence, the morphism of groups $f: B(\mathbf{Z}\times\mathbf{Z}, C)\to C$ has a zero kernel and is therefore injective.

We now show that f is surjective. This is based on the easily verified fact that the map $\psi_0: \mathbf{Z}\times\mathbf{Z}\to\mathbf{Z}$ given by $\psi_0(m, n) = mn$ is bilinear and $\psi_0(1, 1) = 1$. Now suppose c is an element of C. Then we know that there is a unique morphism of groups $h: \mathbf{Z}\to C$ such that $h(1) = c$. By Basic Properties 1.4, the composition $\mathbf{Z}\times\mathbf{Z}\xrightarrow{\psi_0}\mathbf{Z}\xrightarrow{h} C$ is a bilinear map. Because $h\psi(1, 1) = h(1) = c$, the bilinear map $h\psi_0: \mathbf{Z}\times\mathbf{Z}\to C$ has the property $(h\psi_0)(1, 1) = c$. Hence, $f(h\psi_0) = c$. This means that the morphism of groups $f: B(\mathbf{Z}\times\mathbf{Z}, C)\to C$ is surjective. Therefore, we have shown that f is both injective and surjective and hence is an isomorphism.

Example 1.6 Let m_1 and m_2 be two relatively prime integers. Then $B(\mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}, C) = 0$ for all abelian groups C.

PROOF: Let $k_i: \mathbf{Z}\to\mathbf{Z}/m_i\mathbf{Z}$ be the canonical surjective morphisms of groups for $i = 1$ and 2, and define $k_1\times k_2: \mathbf{Z}\times\mathbf{Z}\to\mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}$ by $(k_1\times k_2)(n_1, n_2) = (k_1(n_1), k_2(n_2))$ for all n_1 and n_2 in $\mathbf{Z}\times\mathbf{Z}$. Clearly, $k_1\times k_2: \mathbf{Z}\times\mathbf{Z}\to\mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}$ is a surjective map.

Now suppose C is an abelian group and $\psi: \mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}\to C$ is a bilinear map. Then it is easily shown that the composition $\mathbf{Z}\times\mathbf{Z}\xrightarrow{k_1\times k_2}\mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}\xrightarrow{\psi} C$ is also a bilinear map. Because $k_1\times k_2: \mathbf{Z}\times\mathbf{Z}\to\mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}$ is a surjective map, we will have shown that the bilinear map $\psi: \mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}\to C$ is zero if we show that the composition $\psi(k_1\times k_2): \mathbf{Z}\times\mathbf{Z}\to C$ is zero. To do this, it suffices by Example 1.5 to show that $\psi(k_1\times k_2)(1, 1) = 0$.

The fact that m_1 and m_2 are relatively prime integers means that there are integers z_1 and z_2 such that $z_1m_1 + z_2m_2 = 1$. Because $k_1: \mathbf{Z}\to\mathbf{Z}/m_1\mathbf{Z}$ is the canonical morphism of groups, it follows that $k_1(1) = k_1(z_1m_1 + z_2m_2) = k_1(z_2m_2) = m_2k_1(z_2)$ because $k_1(z_1m_1) = 0$. Hence, $[\psi(k_1\times k_2)](1, 1) = \psi(k_1(1), k_2(1)) = \psi(m_2k_1(z_2), k_2(1)) = m_2\psi(k_1(z_2), k_2(1)) = \psi(k_1(z_2), m_2k_2(1))$ (see Basic Properties 1.4). Because $k_2(1)$ is in $\mathbf{Z}/m_2\mathbf{Z}$ it follows that $m_2k_2(1) = 0$, so that $\psi(k_1\times k_2)(1, 1) = \psi(k_1(z_2), 0) = 0$ (see Basic Properties 1.4). Therefore, the bilinear map $\psi(k_1\times k_2): \mathbf{Z}\times\mathbf{Z}\to C$ is zero, which means that $\psi: \mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}\to C$ is zero, due to the fact that the map $k_1\times k_2: \mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}$ is surjective. Because this is true for each ψ in $B(\mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}, C)$, it follows that $B(\mathbf{Z}/m_1\mathbf{Z}\times\mathbf{Z}/m_2\mathbf{Z}, C) = 0$ for all abelian groups C.

2. THE COMPOSITION MAPS IN Mod(R)

Returning to the category Mod(R) of modules over the ring R, it turns out that the sets of morphisms $\mathrm{Hom}_R(M_1, M_2)$ are not only abelian groups but are also modules over the center of the ring R. Recall that if R is a ring, then the center $C(R)$ of R is the set of all x in R such that $rx = xr$ for all r in R. Recall also that if R is a ring, then the center of R is a commutative ring and $C(R) = R$ if and only if R is a commutative ring.

We now describe a natural $C(R)$-module structure on the abelian group $\mathrm{Hom}_R(M_1, M_2)$ for all R-modules M_1 and M_2. Suppose $f: M_1 \to M_2$ is an R-module morphism and c is an element in $C(R)$. Define the map $cf: M_1 \to M_2$ by $(cf)(m) = c(f(m))$ for all m in M_1. We claim that $cf: M_1 \to M_2$ is an R-module morphism. Clearly, $cf: M_1 \to M_2$ is a morphism of groups, that is, $cf(m_1 + m_2) = cf(m_1) + cf(m_2)$. Also, if r is in R, then $cf(rm) = c(f(rm)) = c(rf(m)) = (cr)f(m) = (rc)f(m) = r(cf(m)) = r((cf)(m))$ for all m in M. Therefore, $cf: M_1 \to M_2$ is indeed an R-module morphism. We leave it to the reader to check that the map $C(R) \times \mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(M_1, M_2)$ given by $(c, f) \mapsto cf$ is a $C(R)$-module structure on $\mathrm{Hom}_R(M_1, M_2)$.

Having seen that the sets of morphisms $\mathrm{Hom}_R(M_1, M_2)$ in Mod(R) are modules over the ring $C(R)$, the center of R, it is natural to ask if the composition of maps in Mod(R) is at all related to the $C(R)$-module structure on the groups of morphisms in Mod(R).

Suppose $f: M_1 \to M_2$ and $g: M_2 \to M_3$ are R-module morphisms and c is in $C(R)$. Then for each m in M, we have $g((cf)(m)) = g(c(f(m)) = c(g(f(m)) = (cg)(f(m))$. Hence, $g(cf) = (cg)f$. But $c(g(f(m))$ also equals $c[(gf)(m)] = (c(gf))(m)$ for all m in M_1. Thus, $g(cf) = (cg)f = c(gf)$. Therefore, in addition to being bilinear maps of abelian groups, the composition maps $\psi: \mathrm{Hom}_R(M_1, M_2) \times \mathrm{Hom}_R(M_2, M_3) \to \mathrm{Hom}_R(M_1, M_3)$ in Mod(R) have the property that $\psi((cf, g)) = \psi(f, cg) = c\psi(f, g)$ for all c in $C(R)$, f in $\mathrm{Hom}_R(M_1, M_2)$, and g in $\mathrm{Hom}_R(M_2, M_3)$. These properties of the composition maps in Mod(R) are a special case of the following general notion.

Definition
Let A, B, and C be modules over a commutative ring R. A map $\psi: A \times B \to C$ is said to be a **bilinear map of R-modules** if:

(a) $\psi: A \times B \to C$ is a bilinear map of the underlying abelian groups of A, B, and C.
(b) $\psi(ra, b) = \psi(a, rb) = r\psi(a, b)$ for all r in R, a in A, and b in B.

This terminology enables us to summarize our previous discussion as follows:

Proposition 2.1
Let $C(R)$ be the center of the ring R. Then the category Mod(R) has the following properties:

(a) For all R-modules M_1 and M_2, the sets $\mathrm{Hom}_R(M_1, M_2)$ are abelian groups under

the addition $f_1 + f_2$ where $f_1 + f_2$ is the morphism of R-modules $f_1 + f_2 : M_1 \to M_2$ defined by $(f_1 + f_2)(m) = f_1(m) + f_2(m)$ for all f_1 and f_2 in $\mathrm{Hom}_R(M_1, M_2)$.

(b) Each $\mathrm{Hom}_R(M_1, M_2)$ is a $C(R)$-module where for each c in $C(R)$ and f in $\mathrm{Hom}_R(M_1, M_2)$, the morphism of R-modules $cf : M_1 \to M_2$ is defined by $cf(m) = c(f(m))$ for all m in M_1.

(c) The composition maps $\mathrm{Hom}_R(M_1, M_2) \times \mathrm{Hom}_R(M_2, M_3) \to \mathrm{Hom}_R(M_1, M_3)$ in $\mathrm{Mod}(R)$ are bilinear maps of $C(R)$-modules.

It should be noted that if R is commutative, then $C(R) = R$, and so each of the groups $\mathrm{Hom}_R(M_1, M_2)$ is an R-module, and the composition maps $\mathrm{Hom}_R(M_1, M_2) \times \mathrm{Hom}_R(M_2, M_3) \to \mathrm{Hom}_R(M_1, M_3)$ are bilinear maps of R-modules.

Returning to the category of modules over an arbitrary ring R, we point out certain extremely useful facts concerning the $C(R)$-modules $\mathrm{Hom}_R(M_1, M_2)$ which follow readily from the bilinearity of the composition of maps in $\mathrm{Mod}(R)$.

Definitions

Suppose M_1, M_2, and M_3 are R-modules.

(a) For each R-module morphism $g : M_2 \to M_3$ define the map $\mathrm{Hom}_R(M_1, g)$: $\mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(M_1, M_3)$ by $\mathrm{Hom}_R(M_1, g)(f) = gf$ for all f in $\mathrm{Hom}_R(M_1, M_2)$.

(b) For each R-module morphism $h : M_3 \to M_1$, define the map $\mathrm{Hom}_R(h, M_2)$: $\mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(M_3, M_2)$ by $\mathrm{Hom}_R(h, M_2)(f) = fh$.

Basic Properties 2.2

Suppose M_1, M_2, X, and Y are R-modules where R is an arbitrary ring.

(a) For each R-module morphism $g : M_2 \to X$, the map $\mathrm{Hom}_R(M_1, g)$: $\mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(M_1, X)$ is a morphism of $C(R)$-modules.

(b) If $g : M_2 \to X$ and $h : X \to Y$ are R-module morphisms, then $\mathrm{Hom}_R(M_1, hg)$: $\mathrm{Hom}_R(M_1, M_2 \to \mathrm{Hom}_R(M_1, Y)$ is the composition $\mathrm{Hom}_R(M_1, h)\mathrm{Hom}_R(M_1, g)$, that is, $\mathrm{Hom}_R(M_1, hg) = \mathrm{Hom}_R(M_1, h)\mathrm{Hom}_R(M_1, g)$.

(c) If $g_1, g_2 : M_2 \to X$ are R-module morphisms, then $\mathrm{Hom}_R(M_1, g_1 + g_2) = \mathrm{Hom}_R(M_1, g_1) + \mathrm{Hom}_R(M_1, g_2)$.

(d) The map $\mathrm{id}_{M_2} : M_2 \to M_2$ has the property that $\mathrm{Hom}_R(M_1, \mathrm{id}_{M_2})$: $\mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(M_1, M_2)$ is the identity on $\mathrm{Hom}_R(M_1, M_2)$.

(e) If $0 : M_2 \to X$ is the zero morphism, then $\mathrm{Hom}(M_1, 0) = 0$.

For morphisms from X to M_1, we have a similar list of properties.

(a′) For each R-module morphism $g : X \to M_1$, the map $\mathrm{Hom}_R(g, M_2)$: $\mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(X, M_2)$ is a morphism of $C(R)$-modules.

(b′) If $h : Y \to X$ and $g : X \to M_1$ are R-module morphisms, then $\mathrm{Hom}_R(gh, M_2)$: $\mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(Y, M_2)$ is the composition $\mathrm{Hom}_R(h, M_2)\mathrm{Hom}_R(g, M_2)$, that is, $\mathrm{Hom}_R(gh, M_2) = \mathrm{Hom}_R(h, M_2)\mathrm{Hom}_R(g, M_2)$.

(c′) If $g_1, g_2 : X \to M_1$ are R-module morphisms, then $\mathrm{Hom}_R(g_1 + g_2, M_2) = \mathrm{Hom}_R(g_1, M_2) + \mathrm{Hom}_R(g_2, M_2)$.

(d′) The map $\mathrm{id}_{M_1} : M_1 \to M_1$ has the property that $\mathrm{Hom}_R(\mathrm{id}_{M_1}, M_2) : \mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_R(M_1, M_2)$ is the identity morphism.

(e′) If $0 : X \to M_1$ is the zero morphism, then $\mathrm{Hom}_R(0, M_2) = 0$.

As with bilinear maps of groups, we are delaying a systematic development of the notion of bilinear maps of modules over commutative rings until later on. In the meantime the reader will gain some familiarity with this notion by working out the following facts concerning maps of modules over arbitrary commutative rings. As is readily seen, they are simply a generalization of the basic properties already given for bilinear maps of abelian groups.

Basic Properties 2.3

Let A, B, and C be modules over the commutative ring R.

(a) A map $\psi: A \times B \to C$ is a bilinear map of R-modules if and only if it satisfies both of the following conditions:
  (i) For each a in A, the map $\alpha_a: B \to C$ given by $\alpha_a(b) = \psi(a, b)$ for all b in B is a morphism of R-modules.
  (ii) For each b in B, the map $\beta_b: A \to C$ given by $\beta_b(a) = \psi(a, b)$ for all a in A is a morphism of R-modules.

(b) $\psi(0, b) = 0 = \psi(a, 0)$ for all a in A and b in B.

(c) The map $0: A \times B \to C$ given by $0(a, b) = 0$ for all a in A and b in B is a bilinear map of R-modules which is called the **zero bilinear map**.

(d) If $\psi_1, \psi_2: A \times B \to C$ are bilinear maps of R-modules, then the map $\psi_1 + \psi_2: A \times B \to C$ defined by $(\psi_1 + \psi_2)(a, b) = \psi_1(a, b) + \psi_2(a, b)$ is bilinear.

(e) If $\psi: A \times B \to C$ is a bilinear map of R-modules and $f: C \to D$ is a morphism of R-modules, then $f\psi: A \times B \to D$ is a bilinear map of R-modules.

(f) If $\psi: A \times B \to C$ is a bilinear map of R-modules and r is in R, then the map $r\psi: A \times B \to C$ given by $(r\psi)(a, b) = r(\psi(a, b))$ is a bilinear map of R-modules. In particular, $(-1)\psi$ is a bilinear map of R-modules.

As a consequence of these results it is not difficult to see that the set $B(A \times B, C)$ of all bilinear R-module maps from $A \times B$ to C is an R-module under the following operations: (1) for ψ_1 and ψ_2 in $B(A \times B, C)$ the sum $\psi_1 + \psi_2$ is the bilinear map of R-modules defined by $(\psi_1 + \psi_2)(a, b) = \psi_1(a, b) + \psi_2(a, b)$ for all a in A and b in B, and (2) for r in R and ψ in $B(A \times B, C)$, the product $r\psi$ is defined to be the bilinear map $r\psi: A \times B \to C$ given by $r\psi(a, b) = r(\psi(a, b))$ for all a in A and b in B.

This leads to the following.

Definition

Let A, B, and C be modules over the commutative ring R. The R-module consisting of all bilinear R-module maps from $A \times B$ to C is called the **module of all bilinear R-module maps from $A \times B$ to C** and is denoted by $B(A \times B, C)$.

3. ANALYSES OF R-MODULE MORPHISMS

This section is devoted to developing the analogs for the category of modules of the notions of surjective and injective morphisms, analyses of morphisms, etc., that we have already discussed in other contexts. The only essentially new ideas introduced are those of exact sequences and the fact that various properties of a morphism $f: M_1 \to M_2$ of R-modules can be expressed in terms of the morphisms

of $C(R)$-modules $\mathrm{Hom}_R(X, f): \mathrm{Hom}_R(X, M_1) \to \mathrm{Hom}_R(X, M_2)$ for all R-modules X as well as the $C(R)$-module morphisms $\mathrm{Hom}_R(f, X): \mathrm{Hom}_R(M_2, X) \to \mathrm{Hom}_R(M_1, X)$ for all R-modules X.

Because $\mathrm{Mod}(R)$ is a category, there is no need to define the notions of isomorphism, epimorphism, or monomorphism for R-modules since we have already defined these notions for arbitrary categories. However, we have never defined the notions of surjective, injective, and bijective morphisms for arbitrary categories. Nonetheless, the reader's previous experience with these notions should immediately suggest their definitions for morphisms of R-modules. We simply record them to avoid any possible doubt.

Definitions
Let $f: M_1 \to M_2$ be a morphism of R-modules. Then the morphism f is **surjective**, **injective**, or **bijective** if as a map on the underlying sets of M_1 and M_2 it is respectively surjective, injective, or bijective.

We leave it to the reader to verify the following useful criteria for when a map between modules is actually a morphism of modules. The reader who has difficulty carrying out these demonstrations should consult the analogous results for monoids, groups, and rings.

Basic Properties 3.1
Let $f: M_1 \to M_2$ be a morphism of R-modules and X an R-module.

(a) Suppose $f: M_1 \to M_2$ is a surjective morphism of modules. If $g: M_2 \to X$ is a map of sets such that the composition $gf: M_1 \to X$ of maps of sets is a morphism of R-modules, then $g: M_2 \to X$ is a morphism of R-modules.
(b) Suppose $f: M_1 \to M_2$ is an injective morphism of R-modules. If $g: X \to M_1$ is a map of sets such that the composition $fg: X \to M_2$ of maps of sets is an R-module morphism, then $g: X \to M_1$ is a morphism of R-modules.

As in all the other situations we have studied so far, we have the following easily verified connections between isomorphisms and bijective morphisms, epimorphisms and surjective morphisms, etc.

Basic Properties 3.2
Let $f: M_1 \to M_2$ be a morphism of R-modules.

(a) The morphism f is an isomorphism if and only if it is bijective.
(b) If f is an injective morphism, then it is a monomorphism.
(c) If f is a surjective morphism, then it is an epimorphism.

PROOF: As an illustration of how our previous basic properties of surjective morphisms can be used, we prove that if the morphism $f: M_1 \to M_2$ is bijective, it is an isomorphism. The rest of the proofs are left to the reader.

Suppose $f: M_1 \to M_2$ is a bijective morphism. Then the inverse map $f^{-1}: M_2 \to M_1$ has the property $f^{-1}f = \mathrm{id}_{M_1}$. Because f is surjective and id_{M_1} is a morphism of modules, it follows from Basic Properties 3.1 that $f^{-1}: M_2 \to M_1$ is a morphism of

R-modules. Combining this with the fact that $f^{-1}f = \mathrm{id}_{M_1}$ and $ff^{-1} = \mathrm{id}_{M_2}$, it follows that $f: M_1 \to M_2$ is an isomorphism if it is bijective.

Actually, as we shall see shortly, a morphism of R-modules $f: M_1 \to M_2$ is injective (surjective) if and only if f is a monomorphism (epimorphism). However, before introducing the notions of Kernel and Cokernel of morphisms of R-modules, which will accomplish this, we point out the following useful interpretation of the notions of isomorphism, epimorphism, and monomorphism.

Basic Properties 3.3
Let $f: M_1 \to M_2$ be a morphism of R-modules.

(a) The following statements are equivalent:
  (i) $f: M_1 \to M_2$ is an isomorphism.
  (ii) For each R-module X, the morphism of $C(R)$-modules $\mathrm{Hom}_R(X, f)$: $\mathrm{Hom}_R(X, M_1) \to \mathrm{Hom}_R(X, M_2)$ is an isomorphism.
  (iii) For each R-module X, the morphism of R-modules $\mathrm{Hom}_R(f, X)$: $\mathrm{Hom}_R(M_2, X) \to \mathrm{Hom}_R(M_1, X)$ is an isomorphism.

(b) The morphism f is an epimorphism if and only if for each R-module X, the morphism of $C(R)$-modules $\mathrm{Hom}(f, X): \mathrm{Hom}_R(M_2, X) \to \mathrm{Hom}_R(M_1, X)$ is injective.

(c) The morphism $f: M_1 \to M_2$ is a monomorphism if and only if for each R-module X, the morphism of $C(R)$-modules $\mathrm{Hom}_R(X, f): \mathrm{Hom}_R(X, M_1) \to \mathrm{Hom}_R(X, M_2)$ is injective.

(d) The following statements are equivalent:
  (i) $f: M_1 \to M_2$ is the zero morphism.
  (ii) For each R-module X, the morphism of $C(R)$-modules $\mathrm{Hom}(X, f)$: $\mathrm{Hom}_R(X, M_1) \to \mathrm{Hom}_R(X, M_2)$ is the zero morphism.
  (iii) For each R-module X, the morphism of $C(R)$-modules $\mathrm{Hom}(f, X)$: $\mathrm{Hom}(M_2, X) \to \mathrm{Hom}(M_1, X)$ is the zero morphism.

PROOF: (a) We show that the morphism $f: M_1 \to M_2$ is an isomorphism if and only if $\mathrm{Hom}_R(X, f): \mathrm{Hom}_R(X, M_1) \to \mathrm{Hom}_R(X, M_2)$ is an isomorphism of $C(R)$-modules for all R-modules X.

Suppose $f: M_1 \to M_2$ is an isomorphism. Then the inverse $f^{-1}: M_2 \to M_1$ has the property that $f^{-1}f = \mathrm{id}_{M_1}$ and $ff^{-1} = \mathrm{id}_{M_2}$. Suppose X is an R-module. Then we have $\mathrm{Hom}_R(X, \mathrm{id}_{M_1}) = \mathrm{Hom}_R(X, f^{-1}f)$. But, by what we saw in the last section (see Basic Properties 2.2), $\mathrm{Hom}_R(X, f^{-1}f) = \mathrm{Hom}_R(X, f^{-1})\ \mathrm{Hom}_R(X, f)$ while $\mathrm{Hom}_R(X, \mathrm{id}_{M_1})$ is the identity on $\mathrm{Hom}_R(X, M_1)$. Hence, $\mathrm{Hom}_R(X, f^{-1})\ \mathrm{Hom}_R(X, f) = \mathrm{id}_{\mathrm{Hom}_R(X, M_1)}$. A similar argument shows that because $ff^{-1} = \mathrm{id}_{M_2}$, then $\mathrm{Hom}_R(X, f)\ \mathrm{Hom}_R(X, f^{-1}) = \mathrm{id}_{\mathrm{Hom}_R(X, M_2)}$. Therefore, $\mathrm{Hom}_R(X, f): \mathrm{Hom}_R(X, M_1) \to \mathrm{Hom}_R(X, M_2)$ is an isomorphism with $\mathrm{Hom}_R(X, f^{-1})$ as the inverse.

Suppose now that the morphism $f: M_1 \to M_2$ has the property that for each R-module X, the morphism of $C(R)$-modules $\mathrm{Hom}_R(X, f): \mathrm{Hom}_R(X, M_1) \to \mathrm{Hom}_R(X, M_2)$ is an isomorphism and hence a bijection. In particular, the morphism $\mathrm{Hom}_R(M_2, f): \mathrm{Hom}_R(M_2, M_1) \to \mathrm{Hom}_R(M_2, M_2)$ is a bijection and so there is a unique morphism of R-modules $g: M_2 \to M_1$ such that $\mathrm{Hom}_R(M_2, f)(g) = \mathrm{id}_{M_2}$. By definition, $\mathrm{Hom}_R(M_2, f)(g)$ is the composition $M_2 \xrightarrow{g} M_1 \xrightarrow{f} M_2$, so we have $fg = \mathrm{id}_{M_2}$. We now finish the proof that f is an isomorphism by showing that $gf = \mathrm{id}_{M_1}$.

By hypothesis we know that the morphism of $C(R)$-modules $\mathrm{Hom}_R(M_1, f): \mathrm{Hom}_R(M_1, M_1) \to \mathrm{Hom}_R(M_1, M_2)$ is an isomorphism and is therefore injective. Now gf is in $\mathrm{Hom}_R(M_1, M_1)$ and by definition $\mathrm{Hom}_R(M_1, f)(gf) = f(gf)$. Because $f(gf) = (fg)f$ and $fg = \mathrm{id}_{M_2}$, it follows that $\mathrm{Hom}_R(M, f)(gf) = f$. On the other hand, $\mathrm{Hom}_R(M_1, f)(\mathrm{id}_{M_1}) = f\mathrm{id}_{M_1} = f$. Because $\mathrm{Hom}_R(M_1, f)$ is injective, the fact that $\mathrm{Hom}_R(M, f)(gf) = f = \mathrm{Hom}_R(M_1, f)(\mathrm{id}_{M_1})$ implies that $gf = \mathrm{id}_{M_1}$, which is our desired result. This finishes the proof of the equivalence of part (a), (i) and (ii).

The proof of the equivalence of parts (i) and (iii) is very similar to the proof already given, and is left as an exercise for the reader. (b) and (c) are simply restatements of the definitions involved and are therefore left to the reader to verify.

(d) We saw in the last section that (i) implies (ii) and (iii). We will prove that (iii) implies (i) and leave the proof that (ii) implies (i) to the reader.

(iii) implies (i). Suppose the morphism of R-modules $f: M_1 \to M_2$ has the property that $\mathrm{Hom}_R(f, X): \mathrm{Hom}_R(M_2, X) \to \mathrm{Hom}_R(M_1, X)$ is the zero morphism of $C(R)$-modules for each R-module X. In particular, the morphism $\mathrm{Hom}_R(f, M_2): \mathrm{Hom}_R(M_2, M_2) \to \mathrm{Hom}_R(M_1, M_2)$ is the zero morphism. Hence, $0 = \mathrm{Hom}_R(f, M_2)\mathrm{id}_{M_2}) = \mathrm{id}_{M_2} f = f$, which finishes the proof.

As we have already seen, the notion of a subset, subgroup, subring, etc., plays an important role in studying sets, groups, and rings. Similarly, the notion of a submodule plays an important role in studying modules.

Definition

An R-module M' is a **submodule** of an R-module M if:

(a) The underlying set of M' is a subset of the underlying set of M.
(b) The inclusion map $M' \to M$ is a morphism of R-modules.

We now give an alternate description of the submodules of a module as well as various elementary properties of submodules.

Basic Properties 3.4

Let M be an R-module.

(a) A submodule of M is nothing more than a subset M' of the underlying set of M satisfying:
  (i) M' is a subgroup of M.
  (ii) If m is in M', then rm is in M' for all r in R. In particular, the subsets (0) and M of M are submodules of M.
(b) If $\{M_i\}_{i \in I}$ is a family of submodules of M, then $\bigcap_{i \in I} M_i$ is a submodule of M.
(c) If $\{M_i\}_{i \in I}$ is a family of submodules of M which is totally ordered by inclusion, then $\bigcup_{i \in I} M_i$ is a submodule of M.
(d) If $f: M \to N$ is a morphism of R-modules and M' is a submodule of M, then the subset $f(M')$ of N is a submodule of N.
(e) If $f: M \to N$ is a morphism of modules and N' is a submodule of N, then the subset $f^{-1}(N')$ of M is a submodule of M.

These results suggest the following.

Definitions
Let $f: M \to N$ be a morphism of R-modules.

(a) If N' is a submodule of N, then the submodule $f^{-1}(N')$ is called the **preimage of N' under f.**
(b) If M' is a submodule of M, then the submodule $f(M')$ of N is called the **image of M' under f.**
(c) The submodule $f(M)$ of N is called the **image of f** and is also denoted by $\operatorname{Im} f$.

Before going on to develop more general theory, we pause to consider some examples.

Example 3.5 Suppose A is a $\mathbf{Z}$-module. Then a subset A' of A is a submodule of A if and only if it is a subgroup of A.

Example 3.6 We have already seen that for a field K the notions of a vector space over K and a module over K are the same. Similarly, the notions of subvector spaces and submodules coincide.

Example 3.7 Let R be a commutative ring. Then the submodules of the R-module R are nothing more or less than the ideals of the ring R. That this is not necessarily the case when R is not commutative is shown later on.

Suppose we are given a morphism of R-modules $f: M \to N$. Then we know that as a map of sets the map $f: M \to N$ is the composition $M \xrightarrow{f_0} \operatorname{Im} f \xrightarrow{\text{inc}} N$. Because $\operatorname{Im} f$ is a submodule of N, the inclusion map, $\text{inc}: \operatorname{Im} f \to N$, is an injective morphism of R-modules. Hence, $f_0: M \to \operatorname{Im} f$ is a morphism of R-modules because $f = \text{inc}\, f_0$ is a morphism of R-modules with inc an injective morphism of R-modules. Clearly, the morphism $f_0: M \to \operatorname{Im} f$ is a surjective morphism of R-modules. Therefore, we have shown that the composition $M \xrightarrow{f_0} \operatorname{Im} f \xrightarrow{\text{inc}} N$ is a factorization of the morphism f into a surjective morphism followed by an injective morphism. This leads to the following.

Definition
Let $f: M \to N$ be a morphism of R-modules. The factorization $M \xrightarrow{f_0} \operatorname{Im} f \xrightarrow{\text{inc}} N$ of f into the surjective morphism f_0 followed by the injective morphism $\text{inc}: \operatorname{Im} f \to N$ is called the **image analysis of f.**

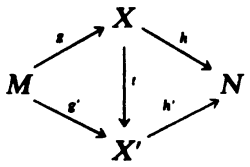
More generally, we have the following.

Definition
Suppose $f: M \to N$ is a morphism of R-modules. Any factorization $M \xrightarrow{g} X \xrightarrow{h} N$ of f with g a surjective morphism of R-modules and h an injective morphism of R-modules is called an **analysis of f.**

As with sets, monoids, groups, etc., we have the following uniqueness property for analyses of morphisms of R-modules.

Basic Property 3.8
Suppose $M \xrightarrow{g} X \xrightarrow{h} N$ and $M \xrightarrow{g'} X \xrightarrow{h'} N$ are two analyses of the same morphism $f: M \to N$ of R-modules. Then there is one and only one morphism

$t: X \to X'$ of R-modules such that the diagram



commutes, and this uniquely determined morphism of R-modules $t: X \to X'$ is an isomorphism.

PROOF: See the proofs of the analogous result for monoids and groups.

Having described the image analysis of a morphism of modules, it is natural to ask what the coimage analysis is for a morphism of modules. Past experience indicates that this question is intimately related to the problem of determining which partitions of the underlying set of a module M occur as the partitions associated with R-module morphisms $f: M \to N$.

Suppose $f: M \to N$ is a morphism of R-modules. Then it is also a morphism of the underlying abelian groups of M and N. Thus, we know that $f^{-1}(0) = K$ is a normal subgroup of f. Also, the coimage analysis of f as a morphism of abelian groups is the factorization $M \xrightarrow{k_{M/K}} M/K \xrightarrow{j_{M/K}} N$ of f where $k_{M/K}: M \to M/K$ is the canonical surjective morphism of abelian groups given by $k_{M/K}(m) = m + K$ for all m in M, and $j_{M/K}: M/K \to N$ is the injective morphism of abelian groups $j_{M/K}(m + K) = f(m)$ for all m in M.

Now $K = f^{-1}(0)$ is not just a subgroup of M. Because $f: M \to N$ is a morphism of modules, K is a submodule of M because $K = f^{-1}(0)$ and (0) is a submodule of N (see Basic Properties 3.4 of submodules). We will now use this fact to show that the abelian group M/K has a unique R-module structure such that the abelian group morphisms $k_{M/K}: M \to M/K$ and $j_{M/K}: M/K \to N$ are morphisms of R-modules. In this way, the composition $M \xrightarrow{k_{M/K}} M/K \xrightarrow{j_{M/K}} N$ is not only an analysis of f as a morphism of abelian groups but also as a morphism of R-modules. It is this analysis of f as an R-module which we will call, for obvious reasons, the coimage analysis of f.

Instead of just showing how to define an R-module structure on M/K having our desired properties, we deal with the following situation which, while it has the appearance of being more general, really is not, as we shall see later. Namely, suppose M' is an arbitrary submodule of the R-module M. Then M' is also a subgroup of the underlying abelian group of M and hence a normal subgroup of M. Therefore, we can form the factor group M/M' which is also an abelian group. We now use the fact that M' is a submodule, not just a subgroup of M, to show that there is an R-module structure on M/M' with the property that the canonical morphism of groups $k_{M/M'}: M \to M/M'$ is a morphism of R-modules.

We first show that if r is in R and X is a subset of M which is an element of M/M', then there is a unique subset Y of M consisting of all the elements rx with x in X. Because the elements of M/M' are of the form $m + M'$ for some m in M, we know that $X = m + M'$ for some m in M. Hence $rX = rm + rM'$. Because M' is a submodule of M we know that $rM' \subset M'$. Therefore, $rm + rM'$ is contained in

$rm + M'$. Hence, rX is contained in the element $rm + M$ of M/M'. Because $rX \neq \emptyset$ and M/M' is a partition of M, we know that $rm + M'$ is the only element of M/M' containing rX where $X = m + M'$. This means that we obtain a map $R \times M/M' \to M/M'$ by defining $(r, m + M') \mapsto rm + M'$ for all r in R and m in M. We claim that this map is our desired R-module structure on the abelian group M/M' because it has the property that the morphism of groups $k_{M/M'}: M \to M/M'$ is also a morphism of R-modules. This can be shown directly using straightforward calculations, or more indirectly as follows.

It is not difficult to check that the surjective morphism of abelian groups $k_{M/M'}: M \to M/M'$ has the property that $k_{M/M'}(rm) = rm + M'$ for all r in R and m in M. Combining this with the following easily verified general observation we obtain not only that the map $R \times M/M' \to M/M'$ given by $(r, m + M) \mapsto rm + M$ for all r in R and m in M is an R-module structure such that the surjective morphism of groups $k_{M/M'}: M \to M/M'$ is a morphism of R-modules, but that it is the only R-module structure on M/M' with this property.

Proposition 3.9
Let R be a ring. Suppose N is an abelian group and $R \times N \to N$ is a map which we denote by $(r, n) \mapsto rn$ for all r in R and n in N. Further, suppose that M is an R-module and that there is a surjective morphism of groups $f: M \to N$ satisfying $f(rm) = rf(m)$ for all r in R and m in M. Then:

(a) The map $R \times N \to N$ is an R-module structure on N.
(b) The R-module consisting of N together with the given R-module structure $R \times N \to N$ has the property that $f: M \to N$ is a morphism of R-modules.
(c) The given R-module structure $R \times N \to N$ is the only R-module structure on N such that the morphism of groups $f: M \to N$ is a morphism of R-modules.

We summarize this discussion in the following.

Definitions
Let M' be a submodule of the R-module M. We denote by M/M' the R-module consisting of the abelian group M/M' together with the unique R-module structure having the property that the surjective morphism of abelian groups $k_{M/M'}: M \to M/M'$ is a morphism of R-modules. This R-module structure on M/M' is given by $r(m + M') = rm + M'$ for all r in R and m in M.

The R-module M/M' is called the **factor module of M by M'** and the surjective morphism of R-modules $k_{M/M'}: M \to M/M'$ is called the **canonical morphism** from M to M/M'.

We now use the notion of a factor module to finish our discussion of the coimage analysis of a morphism of R-modules.

Suppose $f: M \to N$ is a morphism of R-modules. Then associated with the morphism f is the submodule $K = f^{-1}(0)$ of M together with the coimage analysis $M \xrightarrow{k_{M/K}} M/K \xrightarrow{i_{M/K}} N$ of f viewed as a morphism of the underlying abelian groups of M and N. Because K is a submodule of M, we know by our previous discussion that the R-module structure of the factor module M/K is the unique R-module structure on the abelian group M/K with the property that the mor-

phism $k_{M/K}: M \to M/K$ of abelian groups is a morphism of R-modules. Thus, viewing M/K as the factor module of M by K, we have the R-module morphism $f = j_{M/K}k_{M/K}$ where $k_{M/K}$ is a surjective morphism of R-modules. From this it follows that $j_{M/K}: M/K \to N$ is an injective morphism of R-modules. Therefore, the composition of R-module morphisms $M \xrightarrow{k_{M/K}} M/K \xrightarrow{j_{M/K}} N$ is an analysis of the R-module morphism f. This suggests the following.

Definitions
Let $f: M \to N$ be a morphism of R-modules.

(a) The submodule $f^{-1}(0)$ of M, which we will often denote by $\operatorname{Ker} f$, is called the **kernel of f.**
(b) The analysis $M \xrightarrow{k_{M/\operatorname{Ker} f}} M/\operatorname{Ker} f \xrightarrow{j_{M/\operatorname{Ker} f}} N$ of f is called the **coimage analysis of f** where $k_{M/\operatorname{Ker} f}: M \to M/\operatorname{Ker} f$ is the canonical surjective morphism of R-modules from M to the factor module $M/\operatorname{Ker} f$ given by $k_{M/\operatorname{Ker} f}(m) = m + K$ for all m in M, and $j_{M/\operatorname{Ker} f}: M/\operatorname{Ker} f \to N$ is the injective morphism of R-modules given by $j_{M/\operatorname{Ker} f}(m + K) = f(m)$ for all m in M.

The reader will recall that early in this section he was asked to show that an injective (surjective) morphism of R-modules is a monomorphism (epimorphism) of R-modules. At the time it was claimed that the converses of these statements are also true. This is now shown in the following.

Proposition 3.10
Let $f: M \to N$ be a morphism of R-modules.

(a) The following statements are equivalent:
  (i) $\operatorname{Ker} f = 0$.
  (ii) f is injective.
  (iii) f is a monomorphism.
(b) f is an epimorphism if and only if f is surjective.
(c) The following statements are equivalent:
  (i) f is an isomorphism.
  (ii) f is both surjective and injective.
  (iii) f is both an epimorphism and a monomorphism.

PROOF: (a) (i) $\Rightarrow$ (ii). The kernel of f and injectivity of f are the same regardless of whether f is viewed as a morphism of modules or abelian groups. Because it has already been shown that a morphism of abelian groups is injective if its kernel is zero, we know that $\operatorname{Ker} f = 0$ implies f is injective as a morphism of R-modules.

(ii) $\Rightarrow$ (iii). Left as an exercise.

(iii) $\Rightarrow$ (i). Suppose $f: M \to N$ is a monomorphism and let $K = \operatorname{Ker} f$. Then the inclusion morphism $\operatorname{inc}: K \to M$ has the property that $f \operatorname{inc} = 0$, while the zero morphism $0: K \to M$ also has the property $f0 = 0$. Hence, because f is a monomorphism, it follows that $\operatorname{inc} = 0$. This implies that $K = 0$, because the inclusion morphism is injective. Therefore, f being a monomorphism implies $\operatorname{Ker} f = 0$.

(b) Obviously the morphism of R-modules $f: M \to N$ is surjective if and only if $\operatorname{Im} f = N$. Because $\operatorname{Im} f$ is a submodule of N, it follows that $\operatorname{Im} f = N$ if and only

if $N/\text{Im}\, f = 0$, the zero R-module. Because the canonical morphism $k_{N/\text{Im}\, f}: N \to N/\text{Im}\, f$ is surjective, it follows that $N/\text{Im}\, f = 0$ if and only if $k_{N/\text{Im}\, f} = 0$, that is, is the zero morphism. Clearly, the morphism $f: M \to N$ has the property $k_{N/\text{Im}\, f} f = 0$ while $0 f$ also equals 0.

Assume now that f is an epimorphism. Then the fact that $k_{N/\text{Im}\, f} f = 0 = 0f$ implies that $k_{N/\text{Im}\, f} = 0$ which means that $N/\text{Im}\, f = 0$ or, what is the same thing, $\text{Im}\, f = N$. Hence, f being an epimorphism of R-modules implies that f is a surjective morphism of R-modules. Since the reader has already shown that surjective morphisms are epimorphisms, the proof of (b) is complete.

(c) This is an immediate consequence of previously established results.

So far, in dealing with a morphism $f: M \to N$ of R-modules, we have found it useful to introduce various other R-modules associated with f such as Ker f, Im f, and Coim f. In the course of the last proof, the R-module $N/\text{Im}\, f$ together with the canonical morphism $k_{N/\text{Im}\, f}: N \to N/\text{Im}\, f$ was also utilized. Because the morphism $k_{N/\text{Im}\, f}: N \to N/\text{Im}\, f$ is generally useful in studying the morphism $f: M \to N$, we make the following definition.

Definition
Let $f: M \to N$ be a morphism of R-modules. Then the canonical surjective morphism $k_{N/\text{Im}\, f}: N \to N/\text{Im}\, f$ is called the **cokernel of** f. We shall usually denote the R-module $N/\text{Im}\, f$ by Coker f and, unless stated to the contrary, whenever we write a morphism $N \to \text{Coker}\, f$ we mean the canonical surjective morphism $k_{N/\text{Im}\, f}: N \to N/\text{Im}\, f$.

With this definition of cokernel, part (b) of the preceding proposition may now be stated as follows. The following statements are equivalent:

(i) Coker $f = 0$.
(ii) f is surjective.
(iii) f is an epimorphism.

We end this section with the following easily verified set of properties.

Basic Properties 3.11
Let M be an R-module.

(a) If M' is a submodule of M, then M' is the kernel of the canonical surjective morphism $k_{M/M'}: M \to M/M'$.
(b) Suppose $f: M \to N$ is a morphism of R-modules. Then:
  (i) f is the zero morphism if and only if Ker $f = M$.
  (ii) Im f is the kernel of Coker f, that is, Im f is the kernel of the surjective morphism $N \to \text{Coker}\, f$.

4. EXACT SEQUENCES

We begin this section with the important notion of exact sequences of morphisms of R-modules. After developing some of the basic properties of exact sequences,

we apply this notion to obtain new descriptions of cokernels and kernels of morphisms of R-modules.

Definition

Let R be a ring. A sequence $M_1 \xrightarrow{f_1} M_2 \xrightarrow{f_2} M_3$ of morphisms of R-modules is said to be **exact** if $\operatorname{Im} f_1 = \operatorname{Ker} f_2$. Given an arbitrary subset I of consecutive integers, the sequence $\cdots \to M_{i-1} \xrightarrow{f_{i-1}} M_i \xrightarrow{f_i} M_{i+1} \xrightarrow{f_{i+1}} M_{i+2} \to \cdots$ is said to be exact if $\operatorname{Im} f_{i-1} = \operatorname{Ker} f_i$ for all i in I.

Before giving examples of exact sequences of R-modules, we make the following important general observation. A sequence of R-modules $\cdots \to M_{i-1} \xrightarrow{f_{i-1}} M_i \xrightarrow{f_i} M_{i+1} \to \cdots$ can also be viewed as a sequence of abelian groups because each R-module M_i has an underlying abelian group and each R-module morphism $f_i : M_i \to M_{i+1}$ is also a morphism from the underlying abelian group of M_i to that of M_{i+1}. Because $\operatorname{Im} f_{i-1}$ and $\operatorname{Ker} f_i$ are the same subsets of M_i whether we view f_{i-1} and f_i as morphisms of R-modules or morphisms of abelian groups, it follows that the sequence of R-modules $\cdots \to M_{i-1} \to M_i \to M_{i+1} \to \cdots$ is exact if and only if it is exact when viewed as a sequence of morphisms of the underlying abelian groups of the M_i.

We now give some examples to illustrate the utility of this terminology.

Example 4.1 Let $f: M_1 \to M_2$ be a morphism of R-modules. Then:

(a) f is a monomorphism if and only if the sequence $0 \to M_1 \xrightarrow{f} M_2$ is exact.
(b) f is an epimorphism if and only if the sequence $M_1 \xrightarrow{f} M_2 \to 0$ is exact.
(c) f is an isomorphism if and only if the sequence $0 \to M_1 \to M_2 \to 0$ is exact.

PROOF: (a) We know that a morphism $f: M_1 \to M_2$ is a monomorphism if and only if $\operatorname{Ker} f = 0$. On the other hand, by definition, the sequence $0 \to M_1 \xrightarrow{f} M_2$ is exact if and only if $\operatorname{Im}(0 \to M_1) = \operatorname{Ker} f$. Since $\operatorname{Im}(0 \to M_1) = (0)$, it follows that $0 \to M_1 \xrightarrow{f} M_2$ is exact if and only if $\operatorname{Ker} f = (0)$. Therefore, the sequence $0 \to M_1 \xrightarrow{f} M_2$ is exact if and only if f is a monomorphism.

(b) and (c). Because the proofs of (b) and (c) are similar to those given in part (a), these proofs are left to the reader to carry out.

Example 4.2 Suppose $f: M_1 \to M_2$ is a morphism of R-modules. Then the following sequences are exact:

(a) $0 \to \operatorname{Ker} f \xrightarrow{\text{inc}} M_1 \xrightarrow{f} M_2$.
(b) $M_1 \xrightarrow{f} M_2 \to \operatorname{Coker} f \to 0$.

PROOF: (a) The inclusion morphism $\text{inc}: \operatorname{Ker} f \to M_1$ is a monomorphism. Hence, we know by our previous example that $0 \to \operatorname{Ker} f \xrightarrow{\text{inc}} M_1$ is exact. The fact that $\operatorname{Ker} f \xrightarrow{\text{inc}} M_1 \xrightarrow{f} M_2$ is exact is obvious, because the image of the morphism $\operatorname{Ker} f \to M_1$ is $\operatorname{Ker} f$.

(b) Left as an exercise.

Example 4.3 Let $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ be a sequence of R-modules. Then:

(a) $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ is exact if and only if $\operatorname{Im} f = \operatorname{Ker} g$ and the morphism $f_0: M_1 \to \operatorname{Ker} g$ is an isomorphism.

(b) $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \to 0$ is exact if and only if $\operatorname{Im} f = \operatorname{Ker} g$ and the morphism $g: M_2 \to M_3$ is an epimorphism.

We end this preliminary discussion of exact sequences by pointing out how one compares sequences of morphisms of R-modules.

We first recall that a diagram of morphisms of R-modules

$$\begin{array}{ccc} M_1 & \xrightarrow{f_1} & M_2 \\ {\scriptstyle g_1}\downarrow & & \downarrow{\scriptstyle g_2} \\ N_1 & \xrightarrow{h_1} & N_2 \end{array}$$

is said to commute if $g_2 f_1 = h_1 g_1$. More generally, for an arbitrary subset I of consecutive integers, a diagram of morphisms of R-modules

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & M_{i-1} & \xrightarrow{f_{i-1}} & M_i & \xrightarrow{f_i} & M_{i+1} & \xrightarrow{f_{i+1}} & M_{i+2} \longrightarrow \cdots \\ & & \downarrow{\scriptstyle g_{i-1}} & & \downarrow{\scriptstyle g_i} & & \downarrow{\scriptstyle g_{i+1}} & & \downarrow{\scriptstyle g_{i+2}} \\ \cdots & \longrightarrow & N_{i-1} & \xrightarrow[h_{i-1}]{} & N_i & \xrightarrow[h_i]{} & N_{i+1} & \xrightarrow[h_{i+1}]{} & N_{i+2} \longrightarrow \cdots \end{array}$$

commutes if each square in the diagram commutes, that is, $g_{i+1} f_i = h_i g_i$ for all i in I.

As an almost immediate consequence of these definitions, we have the following.

Basic Property 4.4

Suppose the diagram of R-modules

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & M_{i-1} & \xrightarrow{f_{i-1}} & M_i & \xrightarrow{f_i} & M_{i+1} \longrightarrow \cdots \\ & & \downarrow{\scriptstyle g_{i-1}} & & \downarrow{\scriptstyle g_i} & & \downarrow{\scriptstyle g_{i+1}} \\ \cdots & \longrightarrow & N_{i-1} & \xrightarrow{h_{i-1}} & N_i & \xrightarrow{h_i} & N_{i+1} \longrightarrow \cdots \end{array}$$

commutes. Then for each R-module X, the diagrams of $C(R)$-modules

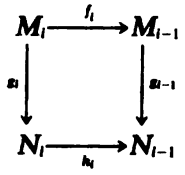
$$\begin{array}{ccccccccc} \cdots & \longrightarrow & \operatorname{Hom}_R(X, M_{i-1}) & \xrightarrow{\operatorname{Hom}_R(X, f_{i-1})} & \operatorname{Hom}_R(X, M_i) & \xrightarrow{\operatorname{Hom}_R(X, f_i)} & \operatorname{Hom}_R(X, M_{i+1}) \longrightarrow \cdots \\ & & \downarrow{\scriptstyle \operatorname{Hom}_R(X, g_{i-1})} & & \downarrow{\scriptstyle \operatorname{Hom}_R(X, g_i)} & & \downarrow{\scriptstyle \operatorname{Hom}_R(X, g_{i+1})} \\ \cdots & \longrightarrow & \operatorname{Hom}_R(X, N_{i-1}) & \xrightarrow[\operatorname{Hom}_R(X, h_{i-1})]{} & \operatorname{Hom}_R(X, N_i) & \xrightarrow[\operatorname{Hom}_R(X, h_i)]{} & \operatorname{Hom}_R(X, N_{i+1}) \longrightarrow \cdots \end{array}$$

and

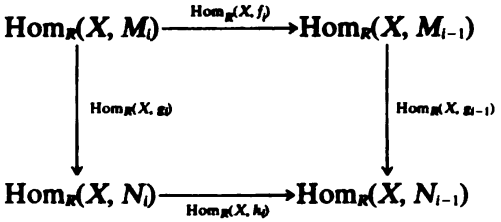
$$\begin{array}{ccccccccc} \cdots & \longrightarrow & \operatorname{Hom}_R(N_{i+1}, X) & \xrightarrow{\operatorname{Hom}_R(h_{i+1}, X)} & \operatorname{Hom}_R(N_i, X) & \xrightarrow{\operatorname{Hom}_R(h_i, X)} & \operatorname{Hom}_R(N_{i-1}, X) \longrightarrow \cdots \\ & & \downarrow{\scriptstyle \operatorname{Hom}_R(g_{i+1}, X)} & & \downarrow{\scriptstyle \operatorname{Hom}_R(g_i, X)} & & \downarrow{\scriptstyle \operatorname{Hom}_R(g_{i-1}, X)} \\ \cdots & \longrightarrow & \operatorname{Hom}_R(M_{i+1}, X) & \xrightarrow[\operatorname{Hom}_R(f_{i+1}, X)]{} & \operatorname{Hom}_R(M_i, X) & \xrightarrow[\operatorname{Hom}_R(f_i, X)]{} & \operatorname{Hom}_R(M_{i-1}, X) \longrightarrow \cdots \end{array}$$

commute.

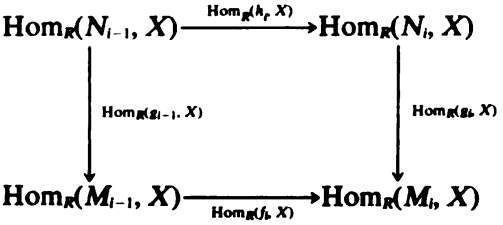
PROOF: All one has to do to prove this is to observe that if the diagram of R-modules



commutes, then for each R-module X the diagrams of $C(R)$-modules



and

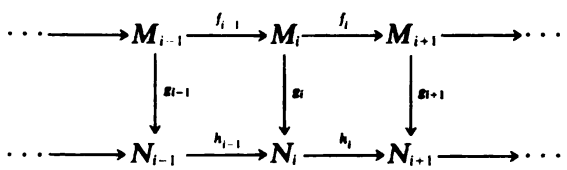


commute. To see this, we simply observe that because $g_{i-1}f_i = h_ig_i$, it follows that $\mathrm{Hom}_R(X, g_{i-1}f_i) = \mathrm{Hom}_R(X, h_ig_i)$ for each R-module X. But $\mathrm{Hom}_R(X, g_{i-1}f_i) = \mathrm{Hom}_R(X, g_{i-1})\mathrm{Hom}_R(X, f_i)$ while $\mathrm{Hom}_R(X, h_ig_i) = \mathrm{Hom}_R(X, h_i)\mathrm{Hom}_R(X, g_i)$. Hence, for each R-module X we have our desired result that $\mathrm{Hom}_R(X, g_{i-1})\mathrm{Hom}_R(X, f_i) = \mathrm{Hom}_R(X, h_i)\mathrm{Hom}_R(X, g_i)$. The rest of the proof goes in a similar way and is left to the reader to verify.

We now explain how to compare sequences of morphisms of R-modules.

Definition

By a **morphism from the sequence** $\cdots \longrightarrow M_{i-1} \xrightarrow{f_{i-1}} M_i \xrightarrow{f_i} M_{i+1} \longrightarrow \cdots$ of morphisms of R-modules to the sequence $\cdots \longrightarrow N_{i-1} \xrightarrow{h_{i-1}} N_i \xrightarrow{h_i} N_{i+1} \longrightarrow \cdots$ we mean a family $\{g_i\}_{i \in I}$ of R-module morphisms $g_i : M_i \to N_i$ such that the diagram of R-modules



commutes.

We say that the morphism $\{g_i\}_{i\in I}$ is respectively a **monomorphism, epimorphism,** or **isomorphism** if each of the morphisms g_i is either a monomorphism, epimorphism, or isomorphism.

We leave it to the reader to verify the following.

Basic Properties 4.5
Suppose we are given the morphism

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & M_{i-1} & \xrightarrow{f_{i-1}} & M_i & \xrightarrow{f_i} & M_{i+1} & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle g_{i-1}} & & \downarrow{\scriptstyle g_i} & & \downarrow{\scriptstyle g_{i+1}} & & \\
\cdots & \longrightarrow & N_{i-1} & \xrightarrow{h_{i-1}} & N_i & \xrightarrow{h_i} & N_{i+1} & \longrightarrow & \cdots
\end{array}$$

of sequences of R-modules.

(a) If $\{g_i\}_{i\in I}$ is an isomorphism, then the family $\{g_i^{-1}\}_{i\in I}$ of morphisms of R-modules is a morphism from the sequence $\cdots \to N_{i-1} \xrightarrow{h_{i-1}} N_i \xrightarrow{h_i} N_{i+i} \to \cdots$ to the sequence $\cdots \to M_{i-1} \xrightarrow{f_{i-1}} M_i \xrightarrow{f_i} M_{i+1} \to \cdots$ which is also an isomorphism.

(b) If $\{g_i\}_{i\in I}$ is an isomorphism, then the sequence $\cdots \to M_{i-1} \to M_i \to M_{i+1} \to \cdots$ is exact if and only if the sequence $\cdots \to N_{i-1} \to N_i \to N_{i+1} \to \cdots$ is exact.

(c) For the morphism of sequences $\{g_i\}_{i\in I}$, the following statements are equivalent:

(i) $\{g_i\}_{i\in I}$ is an isomorphism.

(ii) For each R-module X, the morphism of sequences of $C(R)$-modules

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & \mathrm{Hom}_R(X, M_{i-1}) & \xrightarrow{\mathrm{Hom}_R(X, f_{i-1})} & \mathrm{Hom}_R(X, M_i) & \xrightarrow{\mathrm{Hom}_R(X, f_i)} & \mathrm{Hom}_R(X, M_{i+1}) & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle \mathrm{Hom}_R(X, g_{i-1})} & & \downarrow{\scriptstyle \mathrm{Hom}_R(X, g_i)} & & \downarrow{\scriptstyle \mathrm{Hom}_R(X, g_{i+1})} & & \\
\cdots & \longrightarrow & \mathrm{Hom}_R(X, N_{i-1}) & \xrightarrow[\mathrm{Hom}_R(X, h_{i-1})]{} & \mathrm{Hom}_R(X, N_i) & \xrightarrow[\mathrm{Hom}_R(X, h_i)]{} & \mathrm{Hom}_R(X, N_{i+1}) & \longrightarrow & \cdots
\end{array}$$

is an isomorphism.

(iii) For each R-module X, the morphism of sequences of $C(R)$-modules

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & \mathrm{Hom}_R(N_{i+1}, X) & \xrightarrow{\mathrm{Hom}_R(h_{i+1}, X)} & \mathrm{Hom}_R(N_i, X) & \xrightarrow{\mathrm{Hom}_R(f_i, X)} & \mathrm{Hom}_R(N_{i-1}, X) & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\cdots & \longrightarrow & \mathrm{Hom}_R(M_{i+1}, X) & \xrightarrow[\mathrm{Hom}_R(f_{i+1}, X)]{} & \mathrm{Hom}_R(M_i, X) & \xrightarrow[\mathrm{Hom}_R(h_i, X)]{} & \mathrm{Hom}_R(M_{i-1}, X) & \longrightarrow & \cdots
\end{array}$$

is an isomorphism.

We now turn our attention to describing when a sequence of R-modules $0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ is exact in terms of the morphisms from R-modules X to the R-modules M_1, M_2, and M_3. We begin by making the following useful observation.

Suppose M is an arbitrary R-module. Considering the abelian group of R an R-module by means of the R-module structure $R \times R \to R$ given by $(r, x) \to rx$ where rx is the product in R of the elements r and x in R, we want to describe $\mathrm{Hom}_R(R, M)$. The first thing we do is show how we can consider $\mathrm{Hom}_R(R, M)$ an R-module.

We already know that $\mathrm{Hom}_R(R, M)$ is an abelian group. Therefore, it only remains to define an R-module structure on the abelian group $\mathrm{Hom}_R(R, M)$. Suppose $f: R \to M$ is a morphism of R-modules and suppose r is an element of R. Then define the map $(rf): R \to M$ by $(rf)(x) = f(xr)$ for all x in R. We claim that the map $rf: R \to M$ is a morphism of R-modules. For $(rf)(x_1 + x_2) = f(x_1 + x_2)r = f(x_1 r + x_2 r) = f(x_1 r) + f(x_2 r) = (rf)(x_1) + ((rf)(x_2)$ for all x_1 and x_2 in R. Therefore, rf is a morphism of abelian groups. Also, if s and x are in R, then $rf(sx) = f(sxr) = s(f(xr)) = s(rf(x))$. Hence, $rf: R \to M$ is a morphism of R-modules. It is not difficult to verify that we obtain an R-module structure on $\mathrm{Hom}_R(R, M)$ by means of the map $R \times \mathrm{Hom}_R(R, M) \to \mathrm{Hom}_R(R, M)$ defined for all r in R and f in $\mathrm{Hom}_R(R, M)$ by $(r, f) \to rf$ where $(rf)(x) = f(xr)$ for all x in R. Summarizing, we have the following.

Definition

Let M be an arbitrary R-module. We consider the abelian group $\mathrm{Hom}_R(R, M)$ an R-module by means of the R-module structure $R \times \mathrm{Hom}_R(R, M) \to \mathrm{Hom}_R(R, M)$ defined for all r in R and f in $\mathrm{Hom}_R(R, M)$ by $(r, f) \mapsto rf$ where rf is the R-module morphism from R to M given by $(rf)(x) = f(xr)$ for all x in R.

We now give our main result concerning the R-modules $\mathrm{Hom}_R(R, M)$.

Proposition 4.6

For each R-module M, the map $\alpha_M: \mathrm{Hom}_R(R, M) \to M$ given by $\alpha_M(f) = f(1)$ for all f in $\mathrm{Hom}_R(R, M)$ is an isomorphism of R-modules.

Further, if $g: M \to N$ is a morphism of R-modules, then the morphism of abelian groups $\mathrm{Hom}_R(R, g): \mathrm{Hom}_R(R, M) \to \mathrm{Hom}_R(R, N)$ is a morphism of R-modules with the property that the diagram

$$\begin{array}{ccc} \mathrm{Hom}_R(R, M) & \xrightarrow{\alpha_N} & M \\ {\scriptstyle \mathrm{Hom}_R(R, g)} \downarrow & & \downarrow {\scriptstyle g} \\ \mathrm{Hom}_R(R, N) & \xrightarrow{\alpha_N} & N \end{array}$$

commutes.

PROOF: We first show that for each R-module M, the map $\alpha_M: \mathrm{Hom}_R(R, M) \to M$ is a morphism of R-modules. For if f_1 and f_2 are in $\mathrm{Hom}_R(R, M)$, then $\alpha_M(f_1 + f_2) = (f_1 + f_2)(1) = f_1(1) + f_2(1) = \alpha_M(f_1) + \alpha_M(f_2)$. Thus, α_M is a morphism of abelian groups. Also, α_M is a morphism of R-modules because if r is in R and f is in $\mathrm{Hom}_R(R, M)$, then $\alpha_M(rf) = (rf)(1) = f(1r) = rf(1) = r(\alpha_M(f))$.

Next we show that α_M is an isomorphism of R-modules. Suppose $\alpha_M(f) = 0$. Then $f(1) = 0$ which implies $f(r) = 0$ for all r in R because $f(r) = rf(1) = 0$. Therefore, if $\alpha_M(f) = 0$, then $f = 0$ which means that Ker $\alpha_M = 0$. Hence, α_M is an injective morphism. We now show that α_M is also surjective. Suppose m is in M. Define the map $f: R \to M$ by $f(r) = rm$. It is easily checked that f is an R-module morphism with the property that $f(1) = m$ or, equivalently, $\alpha_M(f) = m$. Because m was an arbitrary element of M, we have that α_M is surjective as well as injective and hence an isomorphism of R-modules. This finishes the proof of the first part of the proposition.

We now show that if $f: M \to N$ is a morphism of R-modules, then the morphism of abelian groups $\mathrm{Hom}_R(R, g): \mathrm{Hom}_R(R, M) \to \mathrm{Hom}_R(R, N)$ is a morphism of R-modules. To do this we must show that if r is in R and f is in $\mathrm{Hom}_R(R, M)$, then $\mathrm{Hom}_R(R, g)(rf) = r(\mathrm{Hom}_R(R, g)(f))$. But $\mathrm{Hom}_R(R, g)(rf) = g(rf)$, the composition of the morphism (rf) and g. Because $g(rf)(x) = g(rf(x)) = g(f(xr))$ for each x in R while $r(\mathrm{Hom}_R(R, g)(f))(x) = ((r(gf))(x) = gf(xr)$ for all x in R, it then follows that $\mathrm{Hom}_R(R, g)(rf) = r(\mathrm{Hom}_R(R, g))(f)$. Hence, $\mathrm{Hom}_R(R, g)$: $\mathrm{Hom}_R(R, M) \to \mathrm{Hom}_R(R, N)$ is indeed a morphism of R-modules.

Finally, we show that the diagram of morphisms of R-modules

$$\begin{array}{ccc} \mathrm{Hom}_R(R, M) & \xrightarrow{\alpha_M} & M \\ {\scriptstyle \mathrm{Hom}_R(R, g)}\downarrow & & \downarrow{\scriptstyle g} \\ \mathrm{Hom}_R(R, N) & \xrightarrow{\alpha_N} & N \end{array}$$

commutes. For if f is in $\mathrm{Hom}_R(R, M)$, therefore $\alpha_M(f) = f(1)$, which implies that $(g\alpha_M)(f) = g(f(1))$. But on the other hand, $(\alpha_N \mathrm{Hom}_R(R, g))(f) = (\mathrm{Hom}_R(R, g)(f))(1) = (gf)(1) = g(f(1))$. Therefore we have our desired result that $g\alpha_M = \alpha_N \mathrm{Hom}_R(R, g)$.

Let us apply this result to show that a sequence $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ of R-modules is exact if for all R-modules X, the sequence of $C(R)$-modules $0 \to \mathrm{Hom}_R(X, M_1) \xrightarrow{\mathrm{Hom}_R(X, f)} \mathrm{Hom}_R(X, M_2) \xrightarrow{\mathrm{Hom}_R(X, g)} \mathrm{Hom}_R(X, M_3)$ is exact. In particular, we know that the sequence of $C(R)$-modules $0 \to \mathrm{Hom}_R(R, M_1) \xrightarrow{\mathrm{Hom}_R(R, f)} \mathrm{Hom}_R(R, M_2) \xrightarrow{\mathrm{Hom}_R(R, g)} \mathrm{Hom}_R(R, M_3)$ is exact, which implies that the sequences of the underlying abelian groups is also exact. But by Proposition 4.6, we know that this sequence of abelian groups can also be viewed as a sequence of R-modules. Therefore, viewed as a sequence of R-modules, it is also exact.

It also follows from our previous proposition that the R-module isomorphisms $\alpha_{M_i}: \mathrm{Hom}_R(R, M_i) \to M_i$ defined by $\alpha_{M_i}(f) = f(1)$ for all f in $\mathrm{Hom}_R(R, M_i)$ for $i = 1, 2, 3$ gives an isomorphism of the following sequences

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathrm{Hom}_R(R, M_1) & \xrightarrow{\mathrm{Hom}_R(R, f)} & \mathrm{Hom}_R(R, M_2) & \xrightarrow{\mathrm{Hom}_R(R, g)} & \mathrm{Hom}_R(R, M_3) \\ & & \downarrow{\scriptstyle \alpha_{M_1}} & & \downarrow{\scriptstyle \alpha_{M_2}} & & \downarrow{\scriptstyle \alpha_{M_3}} \\ 0 & \longrightarrow & M_1 & \xrightarrow{f} & M_2 & \xrightarrow{g} & M_3 \end{array}$$

of morphisms of R-modules. Because a sequence of R-modules is exact if it is isomorphic to an exact sequence of R-modules, it follows that the sequence of morphisms of R-modules $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ is also exact.

We now state the following general result, of which the preceding observation is a part.

Theorem 4.7

Let $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ be a sequence of R-modules. Then the following statements are equivalent:

(a) The sequence of R-modules $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ is exact.

(b) For each R-module X, the sequence of $C(R)$-modules $0 \to \mathrm{Hom}_R(X, M_1) \xrightarrow{\mathrm{Hom}_R(X,f)} \mathrm{Hom}_R(X, M_2) \xrightarrow{\mathrm{Hom}_R(X,g)} \mathrm{Hom}_R(X, M_3)$ is exact.

(c) The sequence of R-modules $0 \to \mathrm{Hom}_R(R, M_1) \xrightarrow{\mathrm{Hom}_R(R,f)} \mathrm{Hom}_R(R, M_2) \xrightarrow{\mathrm{Hom}_R(R,g)} \mathrm{Hom}_R(R, M_3)$ is exact.

(d) The composition gf is zero and for each morphism $h: X \to M_2$ such that $gh = 0$, there is a unique morphism $h': X \to M_1$ such that $fh' = h$.

PROOF: (a) implies (b). We suppose X is an arbitrary R-module and $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ is exact. From Basic Properties 3.3 we know that $0 \to \mathrm{Hom}_R(X, M_1) \xrightarrow{\mathrm{Hom}_R(X,f)} \mathrm{Hom}_R(X, M_2)$ is exact and, from Basic Properties 2.2, we know that $\mathrm{Hom}_R(X, gf) = 0$ because $gf = 0$. But $\mathrm{Hom}_R(X, gf) = \mathrm{Hom}_R(X, g)\mathrm{Hom}_R(X, f)$, so it follows that $\mathrm{Im}(\mathrm{Hom}_R(X, f)) \subset \mathrm{Ker}(\mathrm{Hom}_R(X, g))$. Thus, to show that $0 \to \mathrm{Hom}_R(X, M_1) \xrightarrow{\mathrm{Hom}_R(X,f)} \mathrm{Hom}_R(X, M_2) \xrightarrow{\mathrm{Hom}_R(X,g)} \mathrm{Hom}_R(X, M_3)$ is exact, it remains only to show that $\mathrm{Ker}(\mathrm{Hom}_R(X, g))$ is contained in $\mathrm{Im}(\mathrm{Hom}_R(X, f))$.

Suppose that u is in $\mathrm{Ker}(\mathrm{Hom}_R(X, g))$. Then $u: X \to M_2$ is such that $gu = 0$. Therefore, $\mathrm{Im}\, u$ is contained in $\mathrm{Ker}\, g$. But since $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ is exact, we know that $\mathrm{Ker}\, g = \mathrm{Im}\, f$. Because f is a monomorphism, the morphism $f_0: M_1 \to \mathrm{Im}\, f$ is an isomorphism with inverse f_0^{-1}. Define a map $u': X \to M_1$ by $u'(x) = f_0^{-1}u(x)$ for all x in X. It is easy to see that u' is a morphism of R-modules and that $u = \mathrm{Hom}_R(X, f)(u')$. Thus, u is in $\mathrm{Im}(\mathrm{Hom}_R(X, f))$ and we have shown that $0 \to \mathrm{Hom}_R(X, M_1) \xrightarrow{\mathrm{Hom}_R(X,f)} \mathrm{Hom}_R(X, M_2) \xrightarrow{\mathrm{Hom}_R(X,g)} \mathrm{Hom}_R(X, M_3)$ is exact.

(b) obviously implies (c) and that (c) implies (a) is precisely what we proved in the paragraphs immediately preceding the statement of this theorem.

(b) and (d) are clearly equivalent.

The rest of this section is devoted to developing criteria for when a sequence $M_1 \to M_2 \to M_3 \to 0$ of R-modules is exact in terms very similar to those already used to describe when a sequence of R-modules $0 \to M_1 \xrightarrow{f_1} M_2 \xrightarrow{f_2} M_3$ is exact. We have the following.

Theorem 4.8

Let $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ be a sequence of R-modules. Then the following statements are equivalent:

(a) $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \to 0$ is an exact sequence of R-modules.
(b) The composition gf is zero and, for each morphism of R-modules $h: M_2 \to X$ such that $hf = 0$, there is a unique morphism $h': M_3 \to X$ such that $h'g = h$.
(c) For each R-module X, the sequence of $C(R)$-modules $0 \to \mathrm{Hom}_R(M_3, X) \xrightarrow{\mathrm{Hom}_R(g, X)} \mathrm{Hom}_R(M_2, X) \xrightarrow{\mathrm{Hom}_R(f, X)} \mathrm{Hom}_R(M_1, X)$ is exact.

PROOF: (a) implies (b). Because $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$ is exact, the composition gf is zero. Suppose, now, that $h: M_2 \to X$ is a morphism such that $hf = 0$. If x is an element of M_3, there is some element y in M_2 such that $g(y) = x$ because g is an epimorphism. If y' is another element of M_2 such that $g(y') = x$, then $g(y - y') = 0$ and thus $y - y' = f(z)$ for some z in M_1 since $\mathrm{Im}\, f = \mathrm{Ker}\, g$. Therefore, $h(y - y') = hf(z) = 0$, so $h(y) = h(y')$ if $g(y) = g(y') = x$. Consequently, we can define a map $h': M_3 \to X$ by setting $h'(x) = h(y)$ where y is any element in $g^{-1}(x)$. The reader can show that the map $h': M_3 \to X$ is a morphism of R-modules and that $h'g = h$. That the morphism h' is the only one having the property that $h'g = h$ follows from the fact that g is an epimorphism. Hence, (a) implies (b).

(b) implies (c) is obvious.

(c) implies (a). It has already been seen that the exactness of $0 \longrightarrow \mathrm{Hom}_R(M_3, X) \xrightarrow{\mathrm{Hom}_R(g, X)} \mathrm{Hom}_R(M_2, X)$ for all R-modules implies that $M_2 \xrightarrow{g} M_3 \longrightarrow 0$ is exact (see Basic Properties 3.3). The fact that $gf = 0$ follows from Basic Properties 3.3 because the composition $\mathrm{Hom}_R(M_3, X) \to \mathrm{Hom}_R(M_2, X) \xrightarrow{\mathrm{Hom}_R(f, X)} \mathrm{Hom}_R(M_1, X)$ is zero for all R-modules X. Hence, to show that $M_3 \xrightarrow{f} M_2 \xrightarrow{g} M_1 \longrightarrow 0$ is exact, we need only show that $\mathrm{Ker}\, g$ is contained in $\mathrm{Im}\, f$.

To this end, consider the R-module $X = M_2/\mathrm{Im}\, f$ and the canonical epimorphism $k: M_2 \to M_2/\mathrm{Im}\, f$. We claim that $\mathrm{Hom}_R(f, X)(k) = 0$, for $\mathrm{Hom}_R(f, X)(k) = kf$ and $kf = 0$ because $\mathrm{Im}\, f = \mathrm{Ker}\, k$. Thus, there is a unique morphism $k': M_3 \to X$ such that $\mathrm{Hom}_R(g, X)(k') = k$ since we are assuming that $0 \to \mathrm{Hom}_R(M_3, X) \xrightarrow{\mathrm{Hom}_R(g, X)} \mathrm{Hom}_R(M_2, X) \xrightarrow{\mathrm{Hom}_R(f, X)} \mathrm{Hom}_R(M_1, X)$ is exact for every R-module X. This means that $k = k'g$.

Now suppose that x is in $\mathrm{Ker}\, g$. Then $k(x) = k'(g(x)) = k'(0) = 0$, so x is in $\mathrm{Ker}\, k$. Because $\mathrm{Ker}\, k = \mathrm{Im}\, f$, we see that x is in $\mathrm{Im}\, f$ and we have therefore shown that $\mathrm{Ker}\, g$ is contained in $\mathrm{Im}\, f$. This completes the proof that $M \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$ is exact.

5. ISOMORPHISM THEOREMS

We have already developed for groups and rings a body of theorems known as isomorphism theorems. In this section we point out their analogs for modules over a ring. Some of these results are then used to study the relationships given by a morphism $f: M \to N$ of R-modules between various submodules of M and N. Since another convenient notion to have is that of a set of generators for a module, we begin with the following easily verified results on which the definition of a set of generators for a module is based.

Basic Properties 5.1
Let M be an R-module, let S be a subset of M, and let $\{M_i\}_{i\in I}$ be the family of submodules of M consisting of all submodules M_i of M which contain S. Then the submodule $\bigcap_{i\in I} M_i$ of M consists precisely of all elements of M which can be written as a finite sum $\Sigma r_j s_j$ with r_j in R and s_j in S.

Definition
Let S be a subset of an R-module M.

(a) The submodule M' of M consisting of finite sums $\Sigma r_j s_j$ with r_j in R and s_j in S is called the **submodule of M generated by the set S.**
(b) The subset S of M is said to **generate** the module M if the submodule of M generated by S is all of M.
(c) A module M is said to be **finitely generated** if there is a finite subset of M which generates M.

Before developing the basic properties related to these definitions, we give examples to illustrate some of these notions.

Example 5.2 Let V be a vector space over a field K. Then any basis for V is also a set of generators for V. Hence, V is a finite-dimensional vector space over K if and only if it is a finitely generated module over K.

Example 5.3 Suppose A is a finite abelian group. Then A is a finitely generated module over the ring of integers $\mathbf{Z}$ because the whole set A is clearly a finite set which generates the $\mathbf{Z}$-module A.

Example 5.4 Let R be a ring. Earlier in this chapter we described how we view R as an R-module. We recall that the additive group of this R-module is the additive group of the ring and that the R-module structure on this module is the map $R\times R\to R$ given by $(r_1, r_2)\to r_1 r_2$ for all r_1 and r_2 in R. Then any unit in the ring R is a generator for this R-module R. In particular, 1 is a generator for R as an R-module. Letting R be $\mathbf{Z}$, the ring of integers, we see that $\mathbf{Z}$ is a finitely generated $\mathbf{Z}$-module which is not a finite abelian group.

Example 5.5 We have already observed that if R is a commutative ring, then a subset I of R is an ideal of the ring R if and only if it is a submodule of R when R is viewed as an R-module in the usual way. The reader should have no difficulty verifying that the ideal generated by a subset S of R is the same thing as the submodule of R generated by S. In particular, an ideal in R is finitely generated if and only if it is a finitely generated R-module.

Example 5.6 Let $\mathbf{Z}$ be the ring of integers and let $\mathbf{Q}$ be the field of rational numbers. Since under addition $\mathbf{Q}$ is an abelian group, $\mathbf{Q}$ is also a $\mathbf{Z}$-module. The $\mathbf{Z}$-module $\mathbf{Q}$ is not a finitely generated $\mathbf{Z}$-module.

PROOF: We prove that $\mathbf{Q}$ is not finitely generated by showing that the submodule generated by any finite number of elements in $\mathbf{Q}$ is not all of $\mathbf{Q}$.

Suppose $z_1/z_1', \ldots, z_n/z_n'$ is a finite set of elements in $\mathbf{Q}$. Let $s = \prod_{i=1}^{n} z_i'$. Because there are an infinite number of prime elements in $\mathbf{Z}$ and only a finite number of

them divide s, we know there is a prime number p which does not divide s. We claim that the element $1/p$ in $\mathbf{Q}$ is not in the submodule generated by $z_1/z_1', \ldots, z_n/z_n'$. For if it were, we would have $x_1, \ldots, x_n$ in $\mathbf{Z}$ such that $1/p = \Sigma_{i=1}^n x_i(z_i/z_i')$. Then $1/p = \Sigma_{i=1}^n x_i z_i/z_i' = z/\prod_{i=1}^n z_i' = z/s$ for some z in $\mathbf{Z}$. Hence, $zp = s$, which means that p divides s. But p was chosen to be a prime not dividing s. This contradiction shows that the element $1/p$ of $\mathbf{Q}$ is not in the submodule of $\mathbf{Q}$ generated by $z_1/z_1', \ldots, z_n/z_n'$. Hence, no finitely generated submodule of $\mathbf{Q}$ is all of $\mathbf{Q}$ which shows that $\mathbf{Q}$ is not a finitely generated $\mathbf{Z}$-module.

Returning to our general discussion of generators for modules, we have the following.

Basic Properties 5.7

Let S be a subset of an R-module M.

(a) The following statements are equivalent:
  (i) M is generated by S.
  (ii) M is the only submodule of M containing S.
  (iii) If $f: X \to M$ is a morphism of R-modules and $S \subset \operatorname{Im} f$, then f is surjective.
  (iv) If $f: M \to X$ is a morphism of R-modules and $f(s) = 0$ for all s in S, then $f = 0$.
  (v) If $f_1, f_2: M \to X$ are morphisms of R-modules such that $f_1(s) = f_2(s)$ for all s in S, then $f_1 = f_2$.

(b) If $f: M \to X$ is a surjective morphism of R-modules and S generates M, then $f(S)$ generates X. In particular, if M is finitely generated, then so is X.

(c) If $f: M \to X$ is a surjective morphism, then S generates M if and only if $f(S)$ generates X and $\operatorname{Ker} f$ is contained in the submodule of M generated by S.

(d) Let $0 \longrightarrow M' \xrightarrow{i} M \xrightarrow{f} M'' \longrightarrow 0$ be an exact sequence of R-modules. Suppose the module M is generated by the subset S' and the module M'' is generated by the subset S''. For each s in S'', choose one element m_s in $f^{-1}(s)$ and let T be the subset of M consisting of all elements m_s.
  Then:
  (i) $\operatorname{card}(S') = \operatorname{card}(i(S'))$.
  (ii) $\operatorname{card}(T) = \operatorname{card}(S'')$.
  (iii) $i(S') \cup T$ generates M.

(e) If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules and M' and M'' are finitely generated R-modules, then M is also a finitely generated R-module.

PROOF: (a) (i) implies (ii) is obvious.

(ii) implies (iii). Suppose $f: X \to M$ is a morphism of R-modules such that $S \subset \operatorname{Im} f$. Because $\operatorname{Im} f$ is a submodule of M containing S and, by hypothesis, M is the only submodule of M containing S, it follows that $\operatorname{Im} f = M$ which means that $f: X \to M$ is a surjective morphism of R-modules.

(iii) implies (i). Let M' be the submodule of M generated by S. The inclusion morphism $\operatorname{inc}: M' \to M$ obviously has the property that $\operatorname{Im}(\operatorname{inc})$ contains S. Hence, by hypothesis the morphism $\operatorname{inc}: M' \to M$ is surjective which means that $M' = M$ or, what is the same thing, M is generated by S.

Thus, we have shown that (i), (ii), and (iii) are equivalent. We finish the proof of (a) by showing that (ii), (iv), and (v) are equivalent.

(ii) implies (iv). Suppose $f: M \to X$ is a morphism of R-modules such that $f(s) = 0$ for all s in S. Then S is contained in Ker f which must be all of M since M is the only submodule of M containing S. Therefore, f is the zero morphism.

(iv) implies (ii). Suppose M' is the submodule of M generated by S. Then the canonical surjective morphism $k_{M/M'}: M \to M/M'$ has the property that Ker $k_{M/M'} = M'$. So $k_{M/M'}(s) = 0$ for all s in S. Hence, by hypothesis, $k_{M/M'}: M \to M/M'$ is the zero map. But this implies $M/M' = 0$ since the morphism $k_{M/M'}$ is surjective. Therefore, $M' = M$, which means that S generates M. We have already seen that this implies that M is the only submodule of M containing S. So the proof that (ii) implies (iv) is complete, which establishes the equivalence of (ii) and (iv). We finish the proof of (a) by establishing the equivalence of (iv) and (v).

(iv) implies (v). Suppose $f_1, f_2: M \to X$ are two morphisms of R-modules such that $f_1(s) = f_2(s)$ for all s in S. Then the morphism $f_1 - f_2: M \to X$ has the property $(f_1 - f_2)(s) = 0$ for all s in S. Hence, by hypothesis, $f_1 - f_2 = 0$ or equivalently $f_1 = f_2$.

(v) implies (iv). Suppose $f: M \to X$ is a morphism of R-modules such that $f(s) = 0$ for all s in S. The zero morphism $0: M \to X$ has the property that $0(s) = 0 = f(s)$ for all s in S. Therefore, by hypothesis, $f = 0$ because they agree on S. This shows that (v) implies (iv) which completes the proof of the equivalence of (iv) and (v) as well as the proof of (a).

(b) Suppose $f: M \to X$ is a surjective morphism of R-modules and S generates M. We want to show that $f(S)$ generates X. Let X' be the submodule of X generated by $f(S)$. Then $f^{-1}(X')$ is a submodule of M containing S. Hence, $f^{-1}(X') = M$ because M is the only submodule of M containing S. Because $f(f^{-1}(X')) \subset X'$, it follows that $f(M) \subset X'$. Combining this with the fact that $f: M \to X$ is surjective, we have that $X \subset X'$. Hence, $X = X'$, which means that $f(S)$ generates X. The rest of (b) is obvious.

(c) Suppose $f: M \to X$ is a surjective morphism and S is a subset of M with the property that $f(S)$ generates X and the submodule M' generated by S contains Ker f. We want to show that $M' = M$. To do this it suffices to show that if m is in M, there are finite sets $s_1, \ldots, s_n$ of elements in S and $r_1, \ldots, r_n$ in R such that $m = \sum_{i=1}^{n} r_i s_i$.

Suppose m is in M. Because $f(S)$ generates X we know there are a finite number of elements $s_1, \ldots, s_t$ in S and $r_1, \ldots, r_t$ in R such that $f(m) = \sum_{i=1}^{t} r_i f(s_i)$. From this it follows that the element $m - \sum_{i=1}^{t} r_i s_i$ is in Ker f. Because Ker f is contained in the submodule generated by S, we know that there are finite sets of elements $s_{t+1}, \ldots, s_n$ in S and $r_{t+1}, \ldots, r_n$ in R such that $m - \sum_{i=1}^{t} r_i s_i = \sum_{j=t+1}^{n} r_j s_j$. Thus, $m = \sum_{k=1}^{n} r_k s_k$, which is our desired result.

The rest of (c) follows from previously established results and is left as an exercise.

(d) also follows from previously established results and is likewise left as an exercise.

We end this discussion of generators for modules with the following.

Definition
Let $\{M_i\}_{i\in I}$ be a family of submodules of an R-module M. The submodule of M generated by the set $\bigcup_{i\in I} M_i$ is called the **submodule of M generated by the family** $\{M_i\}_{i\in I}$ of submodules of M.

We leave it to the reader to verify the following.

Basic Property 5.8
Let $\{M_i\}_{i\in I}$ be a family of submodules of an R-module M. The submodule generated by the family $\{M_i\}_{i\in I}$ of submodules of M consists of all possible finite sums $m_1+\cdots+m_r$ where each m_i is in M_j for some j in I.

Having established the basic facts concerning generators for modules, we now study the relationships given by a morphism $f: M\to N$ of R-modules between the submodules of M and those of N. We begin with the following.

Proposition 5.9
Let $f: M\to N$ be a morphism of R-modules with $\operatorname{Ker} f = K$. Suppose M' is a submodule of M. Then:

(a) $f(M')$ is a submodule of N.
(b) $f^{-1}(f(M'))$ is the submodule of M generated by M' and K. Consequently:
(c) $f^{-1}(f(M')) = M'$ if and only if $M'\supset K$.
(d) If we denote by $M'+K$, the submodule of M generated by M' and K, then:
  (i) The morphism $g: M'+K\to f(M')$ given by $g(x)=f(x)$ for all x in $M'+K$ is a surjective morphism whose kernel is K. Hence:
  (ii) The morphism $j_g: (M'+K)/K\to f(M')$ given by the coimage analysis of g is an isomorphism.
(e) The morphism $h: M'\to f(M')$ given by $h(m)=f(m)$ for all m in M' is a surjective morphism with kernel $K\cap M'$. Hence, the morphism $j_h: M'/M'\cap K\to f(M')$ given by the coimage analysis of h is an isomorphism.
(f) The isomorphism $t: M'/M'\cap K\to (M'+K)/K$ which is the composition of isomorphisms $M'/M'\cap K \xrightarrow{j_h} f(M') \xrightarrow{j_g} (M'+K)/K$ can be described by $t(m+M'\cap K)=m+K$ for all m in M.

PROOF: Because the proofs of these results are essentially the same as their obvious analogs for groups, we leave the verification of these facts to the reader to carry out.

Specializing the above results to the special case when the R-module morphism $f: M\to N$ is surjective, we obtain the following stronger conclusions.

Proposition 5.10
Let $f: M\to N$ be a surjective morphism of R-modules with kernel K.

(a) For each submodule N' of N we have $f(f^{-1}(N'))=N'$.
(b) For each submodule M' of M containing K, we have that $f^{-1}(f(M'))=M'$. Hence:
(c) If $\mathcal{T}$ is the set of submodules of N and $\mathcal{S}$ is the set of submodules of M

containing K, the maps of sets $\mathcal{S} \to \mathcal{T}$ given by $M' \to f(M')$ for all M' in $\mathcal{S}$ and $\mathcal{T} \to \mathcal{S}$ given by $N' \to f^{-1}(N')$ for all N' in $\mathcal{T}$, are isomorphisms of sets which are inverse of each other. Finally:

(d) For each submodule M' of M containing K, we have that the coimage analysis of the surjective morphism $M' \to f(M')$ given by $m \mapsto f(m)$ for all m in M' yields the canonical isomorphism $M'/K \to f(M')$ given by $m + K \mapsto f(m)$ for all m in M'.

6. NOETHERIAN AND ARTINIAN MODULES

We now apply these results to obtain some preliminary information concerning noetherian and artinian modules, notions which, for modules, are analogous to those we have already discussed for ideals in rings. (See Chapter 5, Section 5.)

Definitions

Let M be an R-module.

(a) M is said to be a **noetherian module**, or to **satisfy the ascending chain condition**, if the set of all submodules of M is noetherian. That is, M is noetherian if given any ascending chain of submodules $M_0 \subset M_1 \subset \cdots \subset M_i \subset M_{i+1} \subset \cdots$ there is an integer n such that $M_i = M_n$ for all $i \geq n$.

(b) M is said to be **artinian**, or to **satisfy the descending chain condition**, if the set of all submodules of M is artinian. That is, M is artinian if given any descending chain $M_0 \supset M_1 \supset \cdots \supset M_i \supset M_{i+1} \supset \cdots$ of submodules, there is an integer n such that $M_i = M_n$ for all $i \geq n$.

Before giving examples of noetherian and artinian modules, we develop some of their basic properties.

Basic Properties 6.1

Let M be an R-module.

(a) The following statements are equivalent:
  (i) M is noetherian.
  (ii) Every submodule of M is finitely generated.
  (iii) Every nonempty subset of submodules of M has a maximal element.

(b) If $0 \to M' \xrightarrow{i} M \xrightarrow{f} M'' \to 0$ is an exact sequence of R-modules, then M is noetherian if and only if M' and M'' are noetherian.

PROOF: (a) We have already proven that if R is a commutative ring, then R is noetherian if and only if all ideals of R are finitely generated. We know that the ideals of R are precisely the submodules of R. The reader should therefore have no difficulty in translating the proof of Chapter 5, Proposition 5.1 for the special module R to an arbitrary module M over an arbitrary ring R.

(b) Suppose M is noetherian and $0 \to M' \xrightarrow{i} M \xrightarrow{f} M'' \to 0$ is an exact sequence of R-modules. We show that M'' and M' are noetherian by showing that

every submodule of M'' and M' is finitely generated. Suppose X is a submodule of M''. Because the morphism $f: M \to M''$ is surjective, we know that $f(f^{-1}(X)) = X$ by Proposition 5.10. Since M is noetherian, the submodule $f^{-1}(X)$ of M is finitely generated. From this it follows that $X = f(f^{-1}(X))$ is also finitely generated (see Basic Properties 5.7). Hence, if M is noetherian, so is M''.

We now show that every submodule of M' is also finitely generated. From the fact that $i: M \to M$ is a monomorphism, it follows that each submodule of M' is isomorphic to a submodule of M. Hence, each submodule of M' is finitely generated because it is isomorphic to a submodule of M which we know is finitely generated because M is noetherian. Therefore, if M is a noetherian module and $0 \to M' \to M \to M'' \to 0$ is exact, then M and M'' are noetherian.

We now complete the proof of (b) by showing that if $0 \to M' \xrightarrow{i} M \xrightarrow{f} M'' \to 0$ is an exact sequence of R-modules with M' and M'' noetherian modules, then M is noetherian. Suppose X is a submodule of M. Then the morphism $g: X \to f(X)$ given by $g(x) = f(x)$ for all x in X is a surjective morphism with kernel $K \cap X$ where $K = \ker f = \operatorname{Im} i$. Because $0 \to M' \xrightarrow{i} M \xrightarrow{f} M'' \to 0$ is exact, it follows that $i(i^{-1}(K \cap X)) = K \cap X$ and so the sequence $0 \to i^{-1}(K \cap X) \to X \xrightarrow{g} f(X) \to 0$ is exact where the morphism $i^{-1}(K \cap X) \to X$ is given by $y \mapsto i(y)$ for all y in $i^{-1}(K \cap X)$. But the fact that $i^{-1}(K \cap X)$ and $f(X)$ are submodules of the noetherian modules M' and M'', respectively, implies that both $i^{-1}(K \cap X)$ and $f(X)$ are finitely generated. This, combined with the fact that $0 \to i^{-1}(K \cap X) \to X \to f(X) \to 0$ is exact, implies that X is finitely generated (see Proposition 5.7). Because this is true for each submodule X of M, we have that M is noetherian, which completes the proof of (b).

We now use these various descriptions of noetherian modules to give some examples of noetherian modules.

Example 6.2 Suppose R is a principal ideal domain. Then R, viewed as an R-module in the usual way, is a noetherian R-module. We have already seen that the submodules of R are nothing more than the ideals of R and that an ideal is generated by a set S if and only if as a module it is generated by S. Because each ideal in R is generated by one element, each submodule of R is also generated by one element. Hence, every submodule of R is finitely generated, which means that R is a noetherian module.

Example 6.3 Suppose A is a finite abelian group. Then viewing A as a module over the integers $\mathbf{Z}$ in the usual way, we have that A is a noetherian $\mathbf{Z}$-module. For the submodules of A are nothing more than the subgroups of A and because each submodule has only a finite number of elements, it is finitely generated.

Example 6.4 Let V be a finite-dimensional vector space over a field K. Now a submodule V' of V is nothing more than a subvector space of V. Because subvector space V' of V is finitely generated, each finite-dimensional vector space over a field is a noetherian module.

Having developed some of the basic properties of noetherian modules, we turn our attention to artinian modules.

Basic Properties 6.5
Let M be an R-module.

(a) The following statements are equivalent:
  (i) M is an artinian module.
  (ii) Every nonempty set of submodules of M has a minimal element.
(b) If $0 \to M' \xrightarrow{i} M \xrightarrow{f} M'' \to 0$ is exact, then M is an artinian R-module if and only if M' and M'' are artinian R-modules.

PROOF: (a) This was already established when we showed in the last chapter that a set $\mathcal{T}$ of subsets of a set X is artinian if and only if every nonempty subset of $\mathcal{T}$ has a minimal element (see Chapter 5, Basic Properties 6.2).

(b) Suppose M is an artinian module and $0 \to M' \xrightarrow{i} M \xrightarrow{f} M'' \to 0$ is an exact sequence of modules. We want to show that this implies that M'' and M' are artinian. We begin by showing that M'' is artinian.

Suppose

$$X_0 \supset X_1 \supset \cdots \supset X_i \supset X_{i+1} \supset \cdots$$

is a descending chain of submodules of M''. Then

$$f^{-1}(X_0) \supset f^{-1}(X_1) \supset \cdots \supset f^{-1}(X_i) \supset f^{-1}(X_{i+1}) \supset \cdots$$

is a descending chain of submodules of M. Because M is artinian, there is an integer n such that $f^{-1}(X_i) = f^{-1}(X_n)$ for all $i \geq n$. The fact that $f: M \to M''$ is surjective implies that $f(f^{-1}(X_i)) = X_i$ for all i. In particular, $X_i = ff^{-1}(X_i)) = f(f^{-1}(X_n)) = X_n$ for all $i \geq n$. Hence, the fact that M is artinian and $f: M \to M'' \to 0$ is exact implies that M'' is artinian.

We now show that M' is artinian. Suppose

$$Y_0 \supset Y_1 \supset \cdots \supset Y_j \supset Y_{j+1} \supset \cdots$$

is a descending chain of submodules of M'. Then

$$i(Y_0) \supset i(Y_1) \supset \cdots \supset i(Y_j) \supset i(Y_{j+1}) \supset \cdots$$

is a descending chain of submodules of M. Because M is artinian, there is an integer n such that $i(Y_j) = i(Y_n)$ for all $j \geq n$. Because $i: M' \to M$ is a monomorphism, we know that $i^{-1}(i(Y_j)) = Y_j$ for all j. In particular, $Y_j = i^{-1}(i(Y_j)) = i^{-1}(i(Y_n)) = Y_n$ for all $j \geq n$. Hence, the fact that $0 \to M' \xrightarrow{i} M$ is exact and M is artinian implies that M' is also artinian.

To finish the proof of (b) we have to show that if $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules with the property that M' and M'' are artinian modules, then M is also artinian. Suppose

$$X_0 \supset X_1 \supset \cdots \supset X_j \supset X_{j+1} \supset \cdots$$

is a descending chain of submodules of M. Then

$$f(X_0) \supset f(X_1) \supset \cdots \supset f(X_j) \supset f(X_{j+1}) \supset \cdots$$

is a descending chain of submodules of M'' whereas

$$i^{-1}(X_0) \supset i^{-1}(X_i) \supset \cdots \supset i^{-1}(X_j) \supset i^{-1}(X_{j+1}) \supset \cdots$$

is a descending chain of submodules of M'. Because M'' and M' are artinian, we know there is an integer n'' such that $f(X_j) = f(X_n'')$ for all $j \geq n''$. Similarly, there is an integer n' such that $i^{-1}(X_j) = i^{-1}(X_{n'})$ for all $j \geq n'$. Therefore, if we let n = maximum of n' and n'', we have that $f(X_j) = f(X_n)$ and $i^{-1}(X_j) = i^{-1}(X_n)$ for all $j \geq n$.

Now for each j in $\mathbf{N}$ define $g_j: X_j \to f(X_j)$ by $g_j:(x) = f(x)$ for all x in X_j. Clearly, g_j is a surjective morphism with kernel $K \cap X_j$ where $K = \operatorname{Ker} f$. Because $0 \to M' \xrightarrow{i} M \xrightarrow{f} M'' \to 0$ is exact, $i: M' \to M$ is a monomorphism with $i(M') = K$. Therefore, $i^{-1}(X_j) = i^{-1}(K \cap X_j)$ and so $i(i^{-1}(X_j)) = K \cap X_j$. Hence, if we define $h_j: i^{-1}(X_j) \to X_j$ by $h_j(x) = i(x)$ for all x in $i^{-1}(X_j)$, then the sequence $0 \to i^{-1}(X_j) \xrightarrow{h_j} X_j \xrightarrow{g_j} f(X_j) \to 0$ is exact for all j in $\mathbf{N}$.

Suppose $j \geq n$. Then we have the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & i^{-1}(X_j) & \xrightarrow{h_j} & X_j & \xrightarrow{g_j} & f(X_j) & \longrightarrow & 0 \\ & & \| & & \downarrow \text{inc} & & \| & & \\ 0 & \longrightarrow & i^{-1}(X_n) & \xrightarrow{h_n} & X_n & \xrightarrow{g_n} & f(X_n) & \longrightarrow & 0 \end{array}$$

with exact rows. From this it follows that $\operatorname{Im} h_j = \operatorname{Im} h_n = \operatorname{Ker} g_n$, so that X_j is a submodule of X_n containing $\operatorname{Ker} g_n$. Therefore, we know that $g_n^{-1}(g_n(X_j)) = X_j$. But $g_n(X_j) = g_j(X_j) = f(X_j) = f(X_n)$, so that we also have $g_n^{-1}(g_n(X_j)) = g_n^{-1}(f(X_n)) = X_n$. Therefore, it follows that $X_j = X_n$ for each $j \geq n$. Thus, the descending chain of submodules of M

$$X_0 \supset X_i \supset \cdots \supset X_i \supset X_{i+1} \supset \cdots$$

has the property that there is an integer n in $\mathbf{N}$ such that $X_i = X_n$ for all $j \geq n$. Because this is true for any descending chain of submodules of M, we have shown that M is an artinian module if there is an exact sequence $0 \to M' \xrightarrow{i} M \xrightarrow{f} M'' \to 0$ of R-modules with M' and M'' artinian modules.

We now give some examples of artinian modules.

Example 6.6 Every finite abelian group A is an artinian $\mathbf{Z}$-module.

Example 6.7 Every finite-dimensional vector space V over a field K is an artinian K-module.

PROOF: Let $\mathcal{T}$ be a nonempty set of submodules of V. Because the submodules of V are the same thing as the subvector spaces of V, we know that $\dim_K(V') \leq \dim_K(V)$ for all submodules V' of V in $\mathcal{T}$, where $\dim_K(V)$ stands for the dimension of the vector space V over K. Because $\mathbf{N}$ is well ordered, we know there is a submodule V_0 of V in $\mathcal{T}$ such that $\dim_K(V_0) \leq \dim_K(V')$ for all V' in $\mathcal{T}$. Then V_0 is a minimal element of $\mathcal{T}$. For if V' in $\mathcal{T}$ is contained in V_0, then $\dim_K(V') \leq \dim_K(V_0) \leq \dim(V')$. Hence, $\dim_K(V') = \dim_K(V_0)$. Because V' is a subspace of the finite-dimensional vector space V_0 of the same dimension as V_0, it follows that $V' = V_0$. This shows that V_0 is indeed a minimal element of $\mathcal{T}$. Hence, V is an artinian K-module because every nonempty subset of submodules of V has a minimal element.

Example 6.8 Let R be a principal ideal domain and I a nonzero ideal in R. Then R/I is an artinian R-module as well as an artinian R/I-module.

PROOF: This follows easily from the result established in the previous chapter (see Chapter 5, Proposition 6.3) that R/I is an artinian ring.

7. FREE R-MODULES

Undoubtedly, the modules that are most familiar to the reader are vector spaces over a field. Probably the most distinctive feature of the theory of vector spaces is that every vector space has a basis. After generalizing the notion of a basis from vector spaces over fields to modules over arbitrary rings, we introduce the notion of a free module over an arbitrary ring. Namely, an R-module is a free R-module if and only if it has a basis.

We recall that a subset S of a vector space V over a field K is said to be linearly independent if each finite subset of distinct elements $s_1, \ldots, s_n$ in S has the property that given $k_1, \ldots, k_n$ in K such that $\Sigma_{i=1}^{n} k_i s_i = 0$, then each $k_i = 0$ for $i = 1, \ldots, n$. This suggests the following general definition.

Definition

Let M be an R-module. A subset S of M is said to be a **linearly independent subset** of M if each finite subset of distinct elements $s_1, \ldots, s_n$ in S has the property that given $r_1, \ldots, r_n$ in R such that $\Sigma_{i=1}^{n} r_i s_i = 0$, then each $r_i = 0$ for $i = 1, \ldots, n$.

Before giving examples of linearly independent subsets of modules, it is convenient to have the following easily verified properties.

Basic Properties 7.1

Let M be an R-module.

(a) The empty set is a linearly independent subset of M.

(b) A subset S of M consisting of a single element m is a linearly independent subset of M if and only if for any r in R we have $rm = 0$ implies $r = 0$. Hence, the subset $\{m\}$ is linearly independent if and only if the morphism of R-modules $R \to M$ given by $r \to rm$ is a monomorphism.

(c) If S is a linearly independent subset of M, then every subset of S is also a linearly independent subset of M.

(d) For a subset S of M, the following statements are equivalent:
  (i) S is a linearly independent subset of M.
  (ii) Each finite subset of S is a linearly independent subset of M.
  (iii) If $\{r_s\}_{s \in S}$ is an almost zero family of elements of R such that $\Sigma_{s \in S} r_s s = 0$, then $r_s = 0$ for each s in S.
  (iv) If $\{r_s\}_{s \in S}$ and $\{r'_s\}_{s \in S}$ are two almost zero families of elements of R such that $\Sigma_{s \in S} r_s s = \Sigma_{s \in S} r'_s s$, then $r_s = r'_s$ for all s in S.

We now give some examples to illustrate various types of linearly independent subsets of modules.

Example 7.2 Suppose the $\mathbf{Z}$-module A is a finite abelian group. Then the empty set is the only linearly independent subset of A.

PROOF: By Basic Properties 7.1, we know that every subset of a linearly independent subset S of A must be linearly independent. Hence, if there are any nonempty linearly independent subsets of A, there must be one consisting of a single element. But again by Basic Properties 7.1, we know that an element a in A is linearly independent if and only if the morphism of $\mathbf{Z}$-modules $\mathbf{Z} \to A$ given by $z \to za$ is a monomorphism. But this cannot be an injective morphism for any a in A because $\mathbf{Z}$ is an infinite set and A is a finite set. Hence, the empty set is the only linearly independent subset of A if A is a finite group.

Lest the reader be misguided into thinking by this example that the finiteness of an abelian group has too much to do with the fact that the empty set is the only linearly independent subset of the abelian group, we now give an example of an infinite abelian group with the property that the empty set is its only linearly independent subset.

Example 7.3 Because the field $\mathbf{Q}$ of rational numbers contains the ring $\mathbf{Z}$ of integers, the additive group of $\mathbf{Q}$, which we also denote by $\mathbf{Q}$, contains the additive group of $\mathbf{Z}$, which we also denote by $\mathbf{Z}$. Then the abelian group $\mathbf{Q}/\mathbf{Z}$ is an infinite group with the property that the empty set is its only linearly independent subset.

PROOF: We first show that $\mathbf{Q}/\mathbf{Z}$ is an infinite group. To see this, let $p_1, p_2, \ldots, p_n, \ldots$ be the distinct positive prime numbers. We have already seen that there are an infinite number of primes. It is easily checked that if p_i and p_j are distinct positive primes, then the cosets $1/p_i+\mathbf{Z}$ and $1/p_j+\mathbf{Z}$ are distinct elements of $\mathbf{Q}/\mathbf{Z}$. Hence, the cosets $1/p_i+\mathbf{Z}$, $1/p_2+\mathbf{Z}$, $\ldots$, $1/p_n+\mathbf{Z}$, $\ldots$ are all distinct elements of $\mathbf{Q}/\mathbf{Z}$ and so $\mathbf{Q}/\mathbf{Z}$ is an infinite group.

Suppose $z_1/z_2+\mathbf{Z}$ is an element of $\mathbf{Q}/\mathbf{Z}$. Then $z_2 \neq 0$ and $z_2(z_1/z_2+\mathbf{Z}) = z_1z_2/z_2+\mathbf{Z} = z_1+\mathbf{Z} = \mathbf{Z}$ which is the zero element of $\mathbf{Q}/\mathbf{Z}$. Hence, given any element x in $\mathbf{Q}/\mathbf{Z}$, there is a nonzero element z in $\mathbf{Z}$ such that $zx = 0$. Therefore, no single element in $\mathbf{Q}/\mathbf{Z}$ is a linearly independent subset of $\mathbf{Q}/\mathbf{Z}$ which means that the empty set is the only subset of $\mathbf{Q}/\mathbf{Z}$ which is linearly independent.

Example 7.4 Let R be a nonzero ring. Then an element x in R is linearly independent if and only if $rx = 0$ implies $r = 0$. Hence, R always has at least one linearly independent element, for example, 1 or any other unit in R.

Example 7.5 Let R be a commutative nonzero ring. Then an element x in R is linearly independent if and only if x is a regular element in R. Moreover, any subset S of R which contains two distinct nonzero elements in R is not linearly independent.

PROOF: We only prove the last assertion of the example. Suppose x and y are distinct nonzero elements in a subset S of R. Then $xy+(-y)x = 0$. Because x and $-y$ are both not zero, it follows that the subset $\{x, y\}$ and hence the set S is not a linearly independent subset of R.

Having generalized the notion of a linearly independent subset of a vector space over a field to arbitrary modules, we can now generalize the notion of a basis for vector spaces over a field to arbitrary R-modules.

Definition
A subset B of an R-module M is said to be a **basis** for M if B is a linearly independent subset of M which also generates M.

An R-module M is said to be a **free R-module** if M has a basis.

As a first step in studying free R-modules, we point out the following easily verified properties.

Basic Properties 7.6
Let R be a ring.

(a) A subset B of an R-module M is a basis for M if and only if given an element m in M there is a unique almost zero family $\{r_b\}_{b\in B}$ of elements in R such that $\Sigma_{b\in B} r_b b = m$.
(b) A subset B' of a basis B of an R-module M is all of B if and only if B' generates M.
(c) Suppose M is a free R-module with basis B. If $f: M \to N$ is an isomorphism of R-modules, then N is a free R-module with basis $f(B)$.
(d) The zero module is a free module with the empty set as basis.
(e) An R-module M is isomorphic to the R-module R if and only if M is a free R-module which has a basis consisting of a single element.

PROOF: (a), (b), (c), and (d) are left as exercises because they follow immediately from the definitions involved.

(e) We have already seen that 1 in R is a linearly independent subset of R. Because it also generates R, we know that R is a free R-module with a basis consisting of a single element, namely, 1. Hence, if $f: R \to N$ is an isomorphism of R-modules, then by (c) N is a free R-module with basis consisting of the single element $f(1)$.

On the other hand, suppose N is a free R-module with basis consisting of the single element n. Then it is easily seen that the map $f: R \to N$ given by $f(r) = rn$ is an isomorphism of R-modules.

Recalling that the ideals of a commutative ring R are the same thing as the submodules of the R-module R, we leave it to the reader to establish as an application of our discussion of free modules the following characterization of when a commutative ring is a PID.

Proposition 7.7
A commutative ring R is a PID if and only if every submodule of R is a free R-module. Further, if R is a PID, then every basis of a nonzero submodule of R has precisely one element.

Again generalizing from vector spaces, we make the following definition.

Definition
Let B be a basis for the free R-module M. For each element m in M, the unique almost zero family $\{r_b\}_{b\in B}$ of elements in R with the property that $m=\Sigma\, r_b b$ is called the set of **coordinates with respect to the basis B** of the element m. If $\{r_b\}_{b\in B}$ is the set of coordinates with respect to B of the element m in M, then for each b in B the element r_b is called the **bth coordinate** of m. For each m in M the bth coordinate of m will usually be denoted by m_b. In this notation the set of coordinates of an element m in M with respect to B is the almost zero family $\{m_b\}_{b\in B}$ of elements of R. Clearly, $m=\Sigma_{b\in B}\, m_b b$ for each m in M.

The reader should have no difficulty verifying that the coordinates of elements in a free R-module have the following.

Basic Properties 7.8
Let B be a basis for the free R-module M.

(a) If m is an element of M, then $m=0$ if and only if $m_b=0$ for all b in B.
(b) Two elements m and m' in M are the same if and only if $m_b=m'_b$ for all b in B.
(c) If $\{r_b\}_{b\in B}$ is an almost zero family of elements of R, then there is a unique element m in M such that $m_b=r_b$ for all b in B, namely, $m=\Sigma_{b\in B}\, r_b b$.
(d) An element m in M is the element x in B if and only if $m_b=0$ for $b\neq x$ and $m_x=1$.
(e) For each element m in M and r in R we have $(rm)_b=r(m_b)$ for each b in B. In particular, for each m in M we have $(-m)_b=-(m_b)$ for all b in B.
(f) For each pair of elements m and m' in M we have $(m+m')_b=m_b+m'_b$ for all b in B.

These rules for calculating with coordinates of elements in a free R-module suggest the following way of constructing a free R-module $F(B)$ starting from a set B. As a set, $F(B)$ is the set of all almost zero families $\{r_b\}_{b\in B}$ of elements in R. We define the addition in $F(B)$ by $\{r_b\}_{b\in B}+\{r'_b\}_{b\in B}=\{r_b+r'_b\}_{b\in B}$. Simple calculations show that $F(B)$ with this addition is an abelian group where the element $\{r_b\}_{b\in B}$ satisfying $r_b=0$ for all b in B is the zero element and $-\{r_b\}_{b\in B}=\{-r_b\}_{b\in B}$. It is also easily checked that the map $R\times F(B)\to F(B)$ given by $(r,\{r_b\}_{b\in B})\mapsto r\{r_b\}_{b\in B}=\{rr_b\}_{b\in B}$ for all r in R and $\{r_b\}_{b\in B}$ in $F(B)$ is an R-module structure on $F(B)$. We now show that the R-module $F(B)$ we just constructed is a free R-module.

For each x in B, let us denote by δ_x the almost zero family $\{r_b\}_{b\in B}$ of elements of R given by $r_b=0$ for $b\neq x$ and $r_x=1$. Let B' be the set of all δ_x in B. It is then not difficult to see that B' is a basis for the R-module $F(B)$. Hence, $F(B)$ is a free R-module with basis B'. Finally, we observe that the map $B\to B'$ given by $x\mapsto\delta_x$ for all x in B is an isomorphism of sets which we usually consider an identification. Obviously, this identification simply consists of writing x for δ_x for every element x in B.

We now summarize this discussion in the following.

Definition
Let R be a ring and B an arbitrary set. By the **free R-module generated by B** we mean the free R-module $F(B)$ whose elements are the almost zero families $\{r_b\}_{b\in B}$

of elements in R, whose addition is given by $\{r_b\}_{b\in B}+\{r'_b\}_{b\in B}=\{r_b+r'_b\}_{b\in B}$ and whose R-module structure is given by $r\{r_b\}_{b\in B}=\{rr_b\}_{b\in B}$.

Further, the map $B\mapsto F(B)$ given by $x\mapsto\delta_x$ for all x in B, where δ_x stands for the almost zero family $\{r_b\}_{b\in B}$ satisfying $r_b=0$ if $b\neq x$ and $r_x=1$, is an injective map which we will usually consider an inclusion map simply by identifying the element δ_x of $F(B)$ with the element x in B. In this way, B becomes a basis for the free module $F(B)$.

We now point out some important features of free modules and bases of free modules.

Proposition 7.9
Let R be a ring.

(a) If B is a basis for the free R-module M and X is an arbitrary R-module, then given any map g from B to the underlying set of X there is one and only one morphism $f:M\to X$ of R-modules such that $f|B=g$. This uniquely determined morphism of R-modules $f:M\to X$ is given by $f(m)=\Sigma_{b\in B}m_bg(b)$ for all m in M where $\{m_b\}_{b\in B}$ is the set of coordinates of m relative to the basis B of M.

(b) Suppose S is a subset of the R-module M and suppose $f:F(S)\to M$ is the unique morphism of R-modules with the property that $f|S:S\to M$ is the inclusion map. Then:
  (i) The submodule of M generated by S is $\operatorname{Im}f$; hence:
  (ii) S generates M if and only if $f:F(S)\to M$ is surjective. Thus, every module is a factor module of a free module.
  (iii) The set S is linearly independent if and only if the map $f:F(S)\to M$ is a monomorphism.
  (iv) S is a basis for M if and only if $f:F(S)\to M$ is an isomorphism of R-modules.

(c) A subset S of an R-module M is a basis for M if and only if given any R-module X and any map of sets $g:S\to X$, there is one and only one morphism $f:M\to X$ of R-modules such that $f|S=g$.

PROOF: The proofs of (a) and (b) are left as exercises for the reader.

(c) Suppose S is a subset of the R-module M with the property that given any R-module X and any map $g:S\to X$, there is one and only one morphism of R-modules $f:M\to X$ such that $f|S=g$. Stated more symbolically, S has the property that if we denote by (S, X) the set of all maps from the set S to the underlying set of the R-module X, then the map $\alpha_X:\operatorname{Hom}_R(M, X)\to(S, X)$ given by $\alpha(f)=f|S$ for all f in $\operatorname{Hom}_R(M, X)$ is an isomorphism of sets.

Now let $f:F(S)\to M$ be the unique morphism of R-modules such that $f|S:S\to M$ is the inclusion map. If we show that $f:F(S)\to M$ is an isomorphism of R-modules, then it will follow from (b) that S is a basis for M.

Suppose X is an R-module. Because $S\subset F(S)$ is a basis for the free R-module $F(S)$, we know by (a) that given any R-module X the set (S, X) of all maps from S to the underlying set of X is isomorphic to the set $\operatorname{Hom}_R(F(S), X)$ by means of the map $\beta_X:\operatorname{Hom}_R(F(S), X)\to(S, X)$ given by $\beta(f)=f|S$ for all f in $\operatorname{Hom}_R(F(S), X)$. Now it is easy to check that for each R-module X, the diagram

$$\begin{array}{ccc} \mathrm{Hom}_R(M, X) & \xrightarrow{\mathrm{Hom}_R(f, X)} & \mathrm{Hom}_R(F(S), X) \\ & {\scriptstyle \alpha_X} \searrow \quad \swarrow {\scriptstyle \beta_X} & \\ & (S, X) & \end{array}$$

of maps commutes. For every R-module X, because both the maps α_X and β_X are isomorphisms of sets, it follows that $\mathrm{Hom}_R(f, X): \mathrm{Hom}_R(M, X) \to \mathrm{Hom}_R(F(S), X)$ is also an isomorphism of sets for every R-module X. Hence, $\mathrm{Hom}_R(f, X): \mathrm{Hom}_R(M, X) \to \mathrm{Hom}_R(F(S), X)$ is an isomorphism of $C(R)$-modules for every R-module X. But we have already seen that a morphism of R-modules $f: F(S) \to M$ is an isomorphism of R-modules if and only if $\mathrm{Hom}_R(f, X): \mathrm{Hom}_R(M, X) \to \mathrm{Hom}_R(F(S), X)$ is an isomorphism of $C(R)$-modules for all R-modules X (see Basic Properties 3.3). Therefore, the morphism $f: F(S) \to M$ is an isomorphism of R-modules which shows that S is a basis for M.

As an application of this last result we prove:

Proposition 7.10

Suppose M is a free R-module and $0 \to X' \xrightarrow{f} X \xrightarrow{g} X'' \to 0$ is an exact sequence of R-modules. Then the sequence of $C(R)$-modules

$$0 \longrightarrow \mathrm{Hom}_R(M, X') \xrightarrow{\mathrm{Hom}_R(M, f)} \mathrm{Hom}_R(M, X) \xrightarrow{\mathrm{Hom}_R(M, g)} \mathrm{Hom}_R(M, X'') \longrightarrow 0$$

is exact.

PROOF: We have already shown (see Theorem 4.7) that if $0 \to X' \xrightarrow{f} X \xrightarrow{g} X'' \to 0$ is an exact sequence of R-modules, then for any R-module M, the sequence of $C(R)$-modules

$$0 \longrightarrow \mathrm{Hom}_R(M, X') \xrightarrow{\mathrm{Hom}_R(M, f)} \mathrm{Hom}_R(M, X) \xrightarrow{\mathrm{Hom}_R(M, g)} \mathrm{Hom}_R(M, X'')$$

is exact. Therefore, to finish the proof of the proposition it suffices to show that if M is a free R-module and $X \xrightarrow{g} X'' \to 0$ is an exact sequence of R-modules, then the morphism $\mathrm{Hom}_R(M, g): \mathrm{Hom}_R(M, X) \to \mathrm{Hom}_R(M, X'')$ of $C(R)$-modules is surjective.

Let B be a basis for the free R-module M. Suppose $h: M \to X''$ is a morphism of R-modules. Let $t = h|B$. Because $X \xrightarrow{g} X''$ is a surjective morphism of R-modules, we know that there is a map $u: B \to X$ such that the composition $B \xrightarrow{u} X \xrightarrow{g} X''$ is the map $t: B \to X''$. Simply define $u(b)$ to be an element of the nonempty set $g^{-1}(t(b))$ for each b in B. Because B is a basis for the free R-module M, we know there is a morphism $d: M \to X$ such that $d|B = u$. Consider the R-module morphism $M \to X''$ which is the composition $M \xrightarrow{d} X \xrightarrow{g} X''$. For each b in B we have that $gd(b) = gu(b) = t(b)$. Hence, the composition $gd: M \to X''$ has the property that $(gd)|B = t = h|B$. Because the morphisms gd and h from M to X'' agree on the basis B of M, they must be the same. But $gd = \mathrm{Hom}_R(M, g)(d)$. Hence, $\mathrm{Hom}_R(M, g)(d) = h$, which implies that if M is free and $X \xrightarrow{g} X' \to 0$ is exact, then $\mathrm{Hom}_R(M, g): \mathrm{Hom}_R(M, X) \to \mathrm{Hom}_R(M, X'') \to 0$ is exact.

8. CHARACTERIZATION OF DIVISION RINGS

We now turn our attention to describing those rings R with the property that every R-module is a free R-module. The reader is already familiar with the fact that fields have this property. We now show that division rings, which are the natural generalization of the notation of a field to arbitrary, not necessarily commutative, rings, also have the same property.

Definition
A ring R is called a **division ring** if it is not the zero ring and every nonzero element in R is a unit in R.

Obviously, a commutative ring is a division ring if and only if it is a field. So fields are special cases of division rings.

In order to show that every module over a division ring has a basis, it is convenient to have the notion of a maximal linearly independent subset of a module over an arbitrary ring R.

Definition
A subset S of an R-module M is said to be a **maximal linearly independent subset of M** if S is linearly independent and S is not contained in any larger linearly independent subset of M.

The main result about maximal linearly independent subsets of a module M is that every linearly independent subset of M is contained in a maximal such subset of M. The proof of this fact is the burden of the following.

Basic Properties 8.1
Let M be an R-module.

(a) If $\{S_i\}_{i\in I}$ is a totally ordered family of linearly independent subsets of M, then $S = \bigcup_{i\in I} S_i$ is a linearly independent subset of M.
(b) Every linearly independent subset S of M is contained in a maximal linearly independent subset of M.
(c) M has a maximal linearly independent subset.
(d) If M is a free R-module, then every basis for M is a maximal linearly independent subset of M.

PROOF: (a) Suppose $\{S_i\}_{i\in I}$ is a totally ordered family of linearly independent subsets of M. We now show that $S = \bigcup_{i\in I} S_i$ is a linearly independent subset of M by showing that each finite subset of S is linearly independent. Suppose $s_1, \ldots, s_n$ is a finite set of distinct elements in S. Then there is a finite subset $\{i_1, \ldots, i_n\}$ of I such that s_j is in S_{i_j} for all $j = 1, \ldots, n$. Because the family $\{S_i\}_{i\in I}$ is totally ordered by inclusion, this implies that there is a maximal element, say S_{i_k}, which contains all the other S_{i_j}. Hence, the set $\{s_1, \ldots, s_n\}$ is contained in S_{i_k} which means that the set $\{s_1, \ldots, s_n\}$ is linearly independent because it is a subset of the linearly independent subset S_{i_k} of M. Because each finite subset of S is linearly independent, it follows that S is linearly independent.

(b) Suppose S is a linearly independent subset of M. Let $\mathcal{T}$ be the set consisting of all linearly independent subsets of M containing S. We consider $\mathcal{T}$ an ordered set by inclusion. Then, by our previous result, $\mathcal{T}$ is an inductive set. For if $\{S_i\}_{i\in I}$ is a nonempty totally ordered subset of $\mathcal{T}$, then $S = \bigcup_{i\in I} S_i$ is in $\mathcal{T}$ and is an upper bound for $\{S_i\}_{i\in I}$. Therefore, by Zorn's lemma, we know that $\mathcal{T}$ has a maximal element, say T. Then T is a linearly independent subset of M containing S which is a maximal linearly independent subset of M. For if $T' \supset T$ and T is a linearly independent subset of M, then T' is clearly in $\mathcal{T}$. Because T is a maximal element of $\mathcal{T}$, it follows that $T = T'$. This shows that T is a maximal linearly independent subset of M containing S.

(c) Because every R-module has at least one linearly independent subset, namely, the empty set, it follows from (b) that every R-module has a maximal linearly independent subset.

(d) Left as an exercise for the reader.

Although the reader has just seen that a basis for a free R-module is a maximal linearly independent set, it is not generally true that every maximal linearly independent set in an R-module, even a free R-module, need be a basis for the R-module. For example, let $R = \mathbf{Z}$ and consider $\mathbf{Z}$ as a module over $\mathbf{Z}$. Then $\mathbf{Z}$ is a free Z-module with basis $\{1\}$. Because $\mathbf{Z}$ is a domain, we know that any nonzero element, say 2, is a linearly independent subset of $\mathbf{Z}$. But $\{2\}$ is a maximal linearly independent subset of $\mathbf{Z}$ because we have already seen that no two distinct nonzero elements in a commutative ring are ever linearly independent (see Example 7.5). Nonetheless, 2 is clearly not a basis for $\mathbf{Z}$ even though it is a maximal linearly independent subset of $\mathbf{Z}$. However, for the special case of division rings we do have the following.

Proposition 8.2
Let D be a division ring. Then the following statements are equivalent for a subset B of a D-module M:

(a) B is a basis for M.
(b) B is a maximal linearly independent subset of M. Since every module has a maximal linearly independent subset, every module over a division ring D has a basis and is therefore a free D-module.

PROOF: Because the reader has already shown that every basis of a module is a maximal linearly independent subset of the module, we only have to show that (b) implies (a). Suppose B is a maximal linearly independent subset of the D-module M. We want to show that B is a basis for M, or what is the same thing, B generates M. Thus, we want to show that if m is in M, then there is a finite set of elements $b_1, \ldots, b_t$ in B and a finite set of elements $r_1, \ldots, r_t$ in D such that $m = \Sigma_{i=1}^{t} r_i b_i$.

If $m = 0$ or is in B, there is obviously nothing to prove. So suppose $m \neq 0$ and m is not in B. Then the set $B \cup \{m\}$ is not linearly independent because it contains B properly and B is a maximal linearly independent subset of M. Hence, there must be some finite subset B' of $B \cup \{m\}$ which is not linearly independent, for we have already seen that a set is linearly independent if and only if every finite

subset is linearly independent. For the same reason we know that the subset B' is not contained in B. Hence, the subset $B' = \{m, b_1, \ldots, b_t\}$ with the $b_1, \ldots, b_t$ in B. Because this set is not linearly independent, we know that there is a sum $rm + \Sigma_{i=1}^{t} r_i b_i = 0$ with the elements $r, r_1, \ldots, r_t$ in D and not all are zero. If $r = 0$, then we would have $\Sigma_{i=1}^{t} r_i b_i = 0$ with some $r_i \neq 0$. But this cannot happen because the fact that $\{b_1, \ldots, b_t\} \subset B$ and B is linearly independent shows that the set $\{b_1, \ldots, b_t\}$ is also linearly independent. Hence, $r \neq 0$. Because D is a division ring, r is a unit and thus has an inverse $1/r$. Multiplying the equation $rm = \Sigma_{i=1}^{t} (-r_i) b_i$ by $1/r$ on the left, we see that $m = \sum_{i=1}^{t} (1/r(-r_i)) b_i$. Therefore, m is in the submodule generated by B. This finishes the proof that a maximal linearly independent subset B of a D-module M is a basis for M because it generates M.

The rest of the proposition is left as an exercise for the reader.

Having established that all modules over division rings are free, we will have a complete description of all nonzero rings R with the property that all R-modules are free if we show that any nonzero ring with this property must be a division ring. Because we are trying to describe when a ring is a division ring in terms of its module theory, it is reasonable to expect that a module-theoretic description of when a ring is a division ring would be helpful. We do this now in terms of the properties of the R-module R.

Suppose R is a division ring. We claim that the R-module R has the following properties: (a) $R \neq (0)$ and (b) (0) and R are the only submodules of R. By definition, a division ring R is not zero so (a) is trivially satisfied. Suppose now that M is a nonzero submodule of a division ring R. Then there is a nonzero x in M. Because R is a division ring there is a y in R such that $yx = 1$. Because yx is in M, it follows that 1 is in M and so $r = r1$ is in M for all r in R, which means that $M = R$. So we see that a division ring R also satisfies (b), that is, it has the property that (0) and R are the only submodules of R.

On the other hand, it is not difficult to see that a nonzero ring R which has the property that (0) and R are its only submodules, is a division ring. To show this we first show that if x is a nonzero element of R and $yx = 0$, then $y = 0$. The set M of all y in R such that $yx = 0$ is a submodule of R, because it is the kernel of the morphism of R-module $R \to R$ given by $r \mapsto rx$ for all r in R. Now $M \neq R$ because 1 is not in R (remember R is not the zero ring). Therefore, $M = (0)$ because (0) and R are the only submodules of R. Hence, if $yx = 0$, then $y = 0$ because it is in M and $M = (0)$.

Next we observe that if x is a nonzero element of R, then there is a y in R such that $yx = 1$. For the subset Rx is a submodule of R which is not the zero submodule of R because it contains the nonzero element x. Hence, $Rx = R$ because (0) and R are the only submodules of R and $Rx \neq 0$. This means that there is a y in R such that $yx = 1$.

We now show that these two observations imply that a nonzero ring R is a division ring if (0) and R are the only submodules of R. To do this we must show that if x is a nonzero element of R, then there is a y in R such that $yx = 1 = xy$. By what we have just shown we know that if x is a nonzero element of R, then there is a y in R such that $yx = 1$. Multiplying both sides of this equation by y on the right, we obtain $yxy = y$ or, equivalently, $y(xy - 1) = 0$. The fact that R is not the

zero ring means that $1 \neq 0$. Because $yx = 1$, it follows that $y \neq 0$. But this, combined with the fact that $y(xy-1)=0$, implies that $xy-1=0$. For if $xy-1 \neq 0$, then by previous observation we would have $y(xy-1) \neq 0$ because both y and $xy-1$ are different from zero. Hence, $xy=1$ which gives our desired result that $yx = 1 = xy$. Thus, we have shown that a nonzero ring R is a division ring if (0) and R are the only submodules of R.

We summarize our discussion up to this point in the following.

Proposition 8.3
A ring R is a division ring if and only if the R-module R is a nonzero module satisfying the condition that (0) and R are the only submodules of R.

This result suggests that for an arbitrary ring R the nonzero R-modules M with the property that (0) and M are the only submodules of M, might be worth considering. In fact they play an important role in all of ring theory and for this reason are given a special name.

Definition
Let R be an arbitrary ring. An R-module M is called a **simple** R-module if $M \neq (0)$ and (0) and M are the only submodules of M.

In this terminology our previous result becomes: A ring R is a division ring if and only if the R-module R is a simple R-module. We leave it to the reader to verify the following characterization of simple R-modules.

Basic Properties 8.4
Let R be an arbitrary ring and M a nonzero R-module. The following conditions are equivalent:

(a) M is generated by each nonzero element in M.
(b) For every R-module X, every morphism $f: X \to M$ is either zero or an epimorphism.
(c) For every R-module X, every morphism $f: M \to X$ is either zero or a monomorphism.

As an immediate consequence of these basic properties, we have the following.

Corollary 8.5
Let M be a simple R-module. Then every endomorphism of M is either zero or an automorphism. Hence, $\mathrm{End}_R(M)$, the ring of endomorphisms of M, is a division ring.

The main point to establish about simple modules in connection with our problem of showing that a nonzero ring R is a division ring if every R-module is free is that every nonzero ring R has at least one simple R-module.

Suppose we know that our nonzero ring R, which has the property that every R-module is free, also has a simple R-module M. Then the simple R-module M must have a basis B since M is a free R-module. Because $M \neq (0)$, we know that B is not empty. We now show that B consists of exactly one element. Let b be an

element of B. Then by one of our characterizations of simple modules (Basic Property 8.4), we know that the element b generates M since $b \neq 0$. By Basic Property 7.6, it follows that $\{b\} = B$. Hence, B consists of a single element.

But we have already shown that a free module over a ring R has a basis consisting of one element if and only if it is isomorphic to R. Hence, the simple R-module M is isomorphic to R which means that R is a simple R-module. Hence, R is a division ring because we have already seen that a ring R is a division ring if and only if the R-module R is a simple R-module. Thus, our problem of showing that a nonzero ring R is a division ring if every R-module is free is solved once we establish that every nonzero ring has a simple module. To this end it is convenient to have the following.

Definition
Let M be a nonzero R-module. A submodule M' of M is said to be a **maximal submodule** of M if and only if $M' \neq M$ and M' and M are the only submodules of M containing M'.

The following characterization of maximal submodules of a module is an almost immediate consequence of the definition.

Basic Property 8.6
A submodule M' of the R-module M is a maximal submodule of M if and only if M/M' is a simple R-module.

PROOF: This is a direct consequence of the isomorphism established by the canonical surjective morphism $k_{M/M'}: M \to M/M'$ between the set of submodules of M containing M' and the set of submodules of M/M'.

Hence, in order to show that a nonzero ring R has simple modules, it suffices to show that the R-module R has a maximal submodule M because in that case R/M is a simple R-module.

Proposition 8.7
Let R be a nonzero ring. Then every submodule M' of R, different from R, is contained in a maximal submodule of R. Consequently, the ring R has at least one maximal submodule M which means that R also has the simple R-module R/M.

PROOF: Let M' be a submodule of R different from R. Let $\mathcal{T}$ be the set of all submodules of R different from R and containing M'. Then $\mathcal{T}$ is not empty because M' is in $\mathcal{T}$. We now show that viewing $\mathcal{T}$ as an ordered set under inclusion, $\mathcal{T}$ is an inductive set.

To do this we must show that a nonempty totally ordered subset $\mathcal{T}'$ of $\mathcal{T}$ has an upper bound in $\mathcal{T}$. Because $\mathcal{T}'$ is a totally ordered set of submodules of M we know that $N = \bigcup_{X \in \mathcal{T}'} X$ is a submodule of M. Because each X in T' contains M', we know that N contains M'. We also claim that $N \neq R$. If $N = R$, then 1 is in N which means that 1 is in X for some X in $\mathcal{T}'$ because $N = \bigcup_{X \in \mathcal{T}'} X$. But then that X would be all of R, which is a contradiction. Hence, $N \neq R$ which means that N is in $\mathcal{T}$ and is an upper bound for $\mathcal{T}'$.

Because $\mathcal{T}$ is an inductive set, it must have a maximal element M by Zorn's lemma. We leave it to the reader to verify that M is a maximal submodule of R. Because M obviously contains M', the first part of the proposition is proven.

In the light of this result, to see that R contains at least one maximal submodule, all we have to do is find some submodule M' of R different from R. Because R is not the zero ring, the zero submodule of R will do.

The rest of the proposition now follows trivially from our previous characterization of maximal submodules.

In the light of this discussion, we have also established the following.

Theorem 8.8
For a nonzero ring R, the following statements are equivalent:

(a) R is a division ring.
(b) Every R-module is a free R-module.
(c) Every nonzero R-module generated by a single element is a free R-module.

As our final remark in this preliminary discussion of simple modules we point out the following.

Proposition 8.9
Let R be a nonzero ring. An R-module M is simple if and only if M is isomorphic to R/N where N is a maximal submodule of R.

PROOF: Suppose M is a simple R-module. Then $M \neq (0)$ and M is generated by any nonzero element m in M. Let m be a nonzero element in M. Then the morphism of R-modules $f: R \to M$ given by $f(r) = rm$ for all r in R is surjective since m generates M. Hence, M is isomorphic to $R/\text{Ker } f$. Because $R/\text{Ker } f$ is a simple R-module, it follows that $\text{Ker } f$ is a maximal submodule of R.

The rest of the proposition follows from previous results and is therefore left as an exercise.

9. RANK OF FREE MODULES

The topic that we consider in this section is the analog of dimension for vector spaces. It is well known that if a vector space has a basis with n elements (n a nonnegative integer), then any other basis also has n elements. Although this is not the case for free modules over arbitrary rings, it is true for a large class of rings known as left noetherian rings. This section is devoted to establishing this fact.

We begin by introducing some convenient terminology.

Definitions
Let R be a ring.

(a) A submodule of R is called a **left ideal of R**.
(b) R is said to be a **left noetherian ring** if the set of left ideals of R is noetherian.
(c) R is said to be **left artinian** if the set of left ideals is artinian.

The terminology left ideal for a ring R comes from the fact that a left ideal is a subgroup I of the additive group of R closed under multiplication on the left by elements of R, that is, $rI \subset I$ for all r in R. Subgroups I of a ring R closed under multiplication on the ring, that is, $Ir \subset I$ for all r in R, are called **right ideals** of R, a notion we shall return to later on. In this terminology an ideal of R is a subgroup of R which is both a left and a right ideal. Finally, in the case of commutative rings, left ideals, right ideals, and ideals are all the same. Hence, a commutative ring is left noetherian (left artinian) if and only if it is noetherian (artinian) in the sense discussed in Chapter 5. This observation gives a ready supply of examples of left noetherian (left artinian) rings, namely, the commutative noetherian and artinian rings we discussed in Chapter 5. In addition, we point out that division rings R are both left noetherian and left artinian rings because the only left ideals of R are (0) and R.

The principal property of left noetherian rings that interests us at the moment is the following.

Basic Property 9.1
If R is a left noetherian (left artinian) ring, then every finitely generated R-module is noetherian (artinian).

PROOF: We will show that all finitely generated modules over a left noetherian ring are noetherian. The proof that all finitely generated modules over a left artinian ring are artinian proceeds in an analogous way and is left as an exercise.

Saying that a ring R is left noetherian is the same thing as saying that R is a noetherian R-module. Suppose now that M is a finitely generated R-module. We show by induction on the number of generators of M that M is a noetherian R-module. The proof depends heavily on the fact established in Section 6 that if $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules, then M is noetherian if M' and M'' are noetherian (see Basic Properties 6.1).

Suppose M can be generated by one element M. Then the map $R \to M$ given by $r \to rm$ is a surjective morphism of R-modules. Because the sequence $R \to M \to 0$ of R-modules is exact and R is noetherian, it follows that M is noetherian. Hence, any R-module that can be generated by one element is noetherian.

Suppose now that we know that all R-modules that can be generated by n elements ($n \geq 1$) are noetherian, and suppose M can be generated by the $n+1$ elements $m_1, \ldots, m_n, m_{n+1}$. Let M' be the submodule of M generated by $m_1, \ldots, m_n$. Then M' is noetherian because M' can be generated by n elements. Let $k_{M/M'}: M \to M/M'$ be the canonical surjective morphism. Because $m_1, \ldots, m_{n+1}$ generate M and $k_{M/M'}$ is a surjective morphism, it follows that $k_{M/M'}(m_1), \ldots, k_{M/M'}(m_{n+1})$ generate M/M'. But the fact that $m_1, \ldots, m_n$ are in M' means that $k_{M/M'}(m_1), \ldots, k_{M/M'}(m_n)$ are all zero. Hence, M/M' is generated by the one element $k_{M/M'}(m_{n+1})$, that is, M'' is generated by one element and is therefore noetherian by the first step in our inductive proof. Because the sequence $0 \to M' \xrightarrow{\text{inc}_{M'}} M \xrightarrow{k_{M/M'}} M/M' \to 0$ is exact and both M' and M/M' are noetherian, it follows that M is noetherian. This completes the proof.

By combining this result with the following general fact concerning noetherian (artinian) modules, we will obtain our desired result concerning the number of elements in a basis for a free module over a left noetherian ring.

Proposition 9.2
Let M be a noetherian (artinian) module over an arbitrary ring R. If $f: M \to M$ is a surjective (injective) morphism of R-modules, then f is an isomorphism.

PROOF: Assume that M is a noetherian module and $f: M \to M$ is a surjective morphism. For each integer $i \geq 1$, let f^i be the composition $ff, \ldots, f$ of i-times. Then each f^i is a surjective morphism because it is the composition of a finite number of surjective morphisms. Also, for each i we have $\operatorname{Ker} f^i \subset \operatorname{Ker} f^{i+1}$. Therefore, we have the ascending chain of submodules

$$\operatorname{Ker} f \subset \operatorname{Ker} f^2 \subset \cdots \subset \operatorname{Ker} f^i \subset \operatorname{Ker} f^{i+1} \subset \cdots$$

Because M is noetherian we know that there is some n such that $\operatorname{Ker} f^n = \operatorname{Ker} f^{n+1}$. We now show that this implies that $\operatorname{Im} f^n \cap \operatorname{Ker} f = 0$.

For suppose x is in $\operatorname{Im} f^n \cap \operatorname{Ker} f$. Then $x = f^n(y)$ for some y in M because x is in $\operatorname{Im} f^n$. Because x is also in $\operatorname{Ker} f$, we have $f(x) = f^{n+1}(y) = 0$ and hence y is in $\operatorname{Ker} f^{n+1}$. But $\operatorname{Ker} f^n = \operatorname{Ker} f^{n+1}$ which means that $f^n(y) = 0$ or, equivalently, $x = 0$. Therefore, we have $\operatorname{Im} f^n \cap \operatorname{Ker} f = 0$. However, $\operatorname{Im} f^n = M$ because $f^n: M \to M$ is surjective. Therefore, $0 = M \cap \operatorname{Ker} f = \operatorname{Ker} f$. Hence, f is an isomorphism because it is injective as well as surjective.

The proof that if M is an artinian module and $f: M \to M$ is an injective morphism, then f is an isomorphism proceeds in a similar fashion. The hypothesis that $f: M \to M$ is an injective morphism implies that $f^i: M \to M$ is also injective for all integers $i \geq 1$ since they are compositions of injective morphisms. Obviously, $\operatorname{Im} f^i \supset \operatorname{Im} f^{i+1}$ for all $i \geq 1$ and so we have the descending chain of submodules

$$\operatorname{Im} f \supset \operatorname{Im} f^2 \supset \cdots \supset \operatorname{Im} f^i \supset \operatorname{Im} f^{i+1} \supset \cdots$$

Because M is an artinian module, we know that there is an integer n such that $\operatorname{Im} f^n = \operatorname{Im} f^{n+1}$. We now show that this fact combined with the hypothesis that f is injective implies that f is surjective and hence an isomorphism.

Let x be an element of M. Then $f^n(x)$ is in $\operatorname{Im} f^n$. Because $\operatorname{Im} f^n = \operatorname{Im} f^{n+1}$, there is a y in M such that $f^n(x) = f^{n+1}(y) = f^n(f(y))$. The fact that f^n is injective shows that $f(y) = x$ because $f^n(x - f(y)) = 0$. Hence, the injective morphism $f: M \to M$ is also surjective and therefore an isomorphism.

We now apply these results to prove the following.

Proposition 9.3
If a free module M over a left noetherian ring R has a finite basis with n elements, then all bases of M are finite with n elements.

PROOF: We first show that if R is a left noetherian ring and B is a finite basis for a free module M, then any other basis B' of M has $\operatorname{card}(B') \leq \operatorname{card}(B)$.

Because the free R-module M has the finite basis B, M is a finitely generated R-module and is therefore a noetherian R-module. Suppose B' is another basis and that $\operatorname{card}(B') > \operatorname{card}(B)$. Then there is a surjective map $g: B' \to B$ which is not injective. Let $f: M \to M$ be the unique morphism of R-modules such that $f(b') = g(b')$ for all b' in B. Then $\operatorname{Im} f = M$ because $B \subset \operatorname{Im} f$ and B generates M. Because M is a noetherian R-module, it follows from Proposition 9.2 that the surjective morphism $f: M \to M$ is an isomorphism. Because $f(b') = g(b')$ for all b' in B', the

fact that f is injective implies that $g: B' \to B$ is injective, which contradicts the fact that $g: B' \to B$ is a surjective map which is not injective. Consequently, our assumption that $\text{card}(B') > \text{card}(B)$ is false or, equivalently, $\text{card}(B') \leq \text{card}(B)$.

From this it follows that if a free module M over a left noetherian ring R has a finite basis, then all bases are finite. Further, if B and B' are two bases for M, we have $\text{card}(B') \leq \text{card}(B)$ and $\text{card}(B) \leq \text{card}(B')$ because both B and B' are finite. Therefore, all bases of M are finite with the same number of elements.

In connection with this result, we make the following definition.

Definition
Let R be an arbitrary ring and M a free R-module. We define the **rank of M** to be the nonnegative integer n, if every basis of M has cardinality n. If M has a well-defined rank, we denote it by $\text{rank}_R(M)$.

Proposition 9.3 shows that if R is a left noetherian ring, then every finitely generated free R-module has a well-defined rank.

As noted earlier, although this result is not true for arbitrary rings, the following considerably weaker result does hold for arbitrary rings.

Proposition 9.4
Let M be a free module over an arbitrary ring R. Then the following statements are equivalent:

(a) There is a basis B for M with a finite number of elements.
(b) M is a finitely generated R-module.
(c) Every basis of M has a finite number of elements.

PROOF: (a) implies (b) is obvious.

(b) implies (c). Suppose B is a basis for M and suppose $m_1, \ldots, m_n$ is a finite set of elements of M which generates M. Then for each $i = 1, \ldots, n$, there is a finite subset B_i of B such that $m_i = \sum_{b_j \in B_i} r_{ij} b_j$ with the r_{ij} in R. Then the set $B' = \bigcup_{i=1}^{n} B_i$ is a finite subset of B which generates M, because the submodule generated by B' contains all the elements $m_1, \ldots, m_n$. But we have already seen (Basic Properties 7.6) that a subset of a basis of an R-module M which generates M is the whole basis. Hence, $B' = B$ which means that B is a finite set.

(c) implies (a) is trivial.

As a final comment about cardinality of bases of free modules, we have the following.

Proposition 9.5
Let R be an arbitrary ring and let M and M' be free R-modules with bases B and B', respectively. If $\text{card}(B) = \text{card}(B')$, then M and M' are isomorphic.

10. COMPLEMENTARY SUBMODULES OF A MODULE

This section is devoted to generalizing the familiar notion of complementary subspaces of a vector space to modules over arbitrary rings. We recall that if V' is a

subspace of the vector space V over a field K, then a complement of V' is a subspace V'' of V such that $V' \cap V'' = (0)$ and the whole space V is generated by V' and V''. Generalizing this notion to modules over arbitrary rings we obtain the following.

Definition
Let M' be a submodule of the R-module M. A **complement of M'** in M is a submodule M'' of M such that (a) $M' \cap M'' = 0$ and (b) M is generated by M' and M''. A submodule M' of M is said to be a **summand of M** if it has a complement in M.

Before giving some examples of summands of modules we establish the following criteria for when a submodule M'' of an R-module M is the complement of a submodule M' of M.

Basic Properties 10.1
Let M' be a submodule of the R-module M.

(a) A submodule M'' of M is a complement of M' if and only if M' is a complement of M'' in M.
(b) A submodule M'' is a complement of M' in M if and only if the canonical surjective morphism $k_{M/M'}: M \to M/M'$ has the property that $k_{M/M'}|M'': M'' \to M/M'$ is an isomorphism of R-modules. Hence:
(c) If M' is a summand of M, then all complements of M' in M are isomorphic R-modules because they are all isomorphic to the R-module M/M'.

PROOF: **(a)** Trivial.

(b) Let M'' be a submodule of M. Then the morphism $f: M'' \to M/M'$ given by $f(x) = x + M'$ for all x in M'' is the same as the morphism $k_{M/M'}|M''$. Because $f(x) = x + M'$ is the zero element of M/M' if and only if x is in M', we have $\operatorname{Ker} f = M'' \cap M'$. Therefore, $f: M'' \to M/M'$ is a monomorphism if and only if $M'' \cap M' = 0$.

Also, because $k_{M/M'}: M \to M/M'$ is surjective with $\operatorname{Ker} k_{M/M'} = M'$, we know by the basic properties of surjective morphisms that $k_{M/M'}^{-1}(k_{M/M'}(M''))$ is the submodule of M generated by M' and M''. Combining this fact with the fact that $f(M'') = k_{M/M'}(M'')$, we see that $f(M'') = M/M'$ if and only if M' and M'' generate M. Therefore, the morphism $f: M'' \to M/M'$ is injective and surjective if and only if $M' \cap M'' = (0)$ and M is generated by M' and M''. Thus, we have our desired result that the morphism $f: M'' \to M/M'$ is an isomorphism of R-modules if and only if M'' is a complement of the submodule M' of M.

(c) Follows trivially from **(b)**.

For two submodules M' and M'' of a module M, because the relationships M' is a complement of M'' in M and M'' is a complement for M' in M are equivalent, we can simplify our terminology as follows.

Definition
We shall say that two submodules M' and M'' of M are **complementary** in M if M' is a complement of M'' in M or, equivalently, M'' is a complement for M' in M.

We now give some examples of summands of modules.

Example 10.2 If M is an R-module, then (0) and M are summands of M whose complements are M and (0), respectively.

A more interesting situation is illustrated in the following.

Example 10.3 Suppose B is a basis for a free R-module M. Let B' be a subset of B and M' the submodule of M generated by B'. Then M'', the submodule of M generated by $B'' = B - B'$, is a complement of M' in M. Hence, M' is a summand of M. Further, each of the R-modules M' and M'' is a free R-module.

PROOF: Follows readily from definitions and is therefore left as an exercise.

As a consequence of this example, we obtain the following generalization to modules over division rings of the fact that subspaces of vector spaces over fields have complements.

Example 10.4 Let D be a division ring. Then every submodule of a D-module M is a summand of M.

PROOF: Let M' be a submodule of the D-module M. Because D is a division ring, M' is a free D-module and so has a basis B'. Hence, B' is a linearly independent subset, not only of M', but of M also. Therefore, there exists a maximal linearly independent subset B of M containing B'. But this maximal linearly independent subset B of M is a basis of M because D is a division ring. Hence, the submodule M' of M is generated by the subset B' of the basis B of the free D-module M. This implies that M' is a summand of M as the reader has just seen in our previous example.

In a later chapter, we shall obtain a description of all rings R having the property that every submodule of an R-module M is a summand of M. Such rings are called semisimple rings.

We now return to our general discussion by relating the notion of complementary submodules of a module to properties of exact sequences.

Proposition 10.5
Suppose $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is an exact sequence of R-modules. Then the following statements are equivalent:

(a) $\operatorname{Im} f$ is a summand of M.
(b) There is a morphism $h : M'' \to M$ such that $gh : M'' \to M''$ is $\mathrm{id}_{M''}$.
(c) There is a morphism $t : M \to M'$ such that $tg : M' \to M'$ is $\mathrm{id}_{M'}$.

PROOF: (a) implies (b). Because the sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is exact we know that g is a surjective morphism with $\operatorname{Ker} g = \operatorname{Im} f$. Hence, the morphism $j_g : M/\operatorname{Ker} g \to M''$ given by the cokernel analysis of g is an isomorphism. The fact that $\operatorname{Im} f = \operatorname{Ker} g$ together with the fact that $\operatorname{Im} f$ is a summand of M means that there is a complementary submodule N of $\operatorname{Ker} g$ in M. By Basic Property 10.1, we know that the restriction of $k_{M/\operatorname{Ker} g} : M \to M/\operatorname{Ker} g$ to N is an isomorphism. Denoting this isomorphism by $u : N \to M/\operatorname{Ker} g$, it is not difficult to check that the morphism $h : M'' \to M$ which is the composition

$M'' \xrightarrow{f_0^{-1}} M/\text{Ker } g \xrightarrow{w^{-1}}$ has the property that composition $M'' \xrightarrow{h} M \xrightarrow{g} M''$ is $\text{id}_{M''}$.

(b) implies **(c)**. Suppose $h: M'' \to M$ is a morphism with the property that the composition $M'' \xrightarrow{h} M \xrightarrow{g} M''$ is $\text{id}_{M''}$. Let $s: M \to M$ be the composition $M \xrightarrow{g} M'' \xrightarrow{h} M$. Consider the morphism $\text{id}_M - s: M \to M$ given by $(\text{id}_M - s) \times (m) = m - s(m)$ for all m in M.

First of all, it is easily seen that $(\text{id}_m - s)(m') = m'$ for all m' in $\text{Im } f$. For if m' is in $\text{Im } f = \text{Ker } g$, then $(\text{id}_m - s)(m') = m' - hg(m') = m'$.

Secondly, we claim that $\text{Im}(\text{id}_m - s) \subset \text{Im } f$. Because $\text{Im } f = \text{Ker } g$, it suffices to show that $g(\text{id}_M - s) = 0$. But for each m in M we have $g(\text{id}_M - s)(m) = g(m - hg(m)) = g(m) - g(hg(m)) = g(m) - (gh)(gm) = g(m) - g(m) = 0$ since $gh = \text{id}_{M''}$.

Combining these two facts, we see that we obtain a morphism $w: M \to \text{Im } f$ by defining $w(m) = (\text{id}_M - s)(m)$ for all m in M which has the property that the composition $\text{Im } f \xrightarrow{\text{inc}} M \xrightarrow{w} \text{Im } f$ is $\text{id}_{\text{Im } f}$. Because $f: M' \to M$ is a monomorphism, we know that $f_0: M' \to \text{Im } f$ is an isomorphism. It now follows that the morphism $t: M \to M'$ which is the composition $M \xrightarrow{w} \text{Im } f \xrightarrow{f_0^{-1}} M'$ has the property that the composition $M' \xrightarrow{f} M \xrightarrow{t} M'$ is $\text{id}_{M'}$.

(c) implies **(a)**. Suppose that $t: M \to M'$ is a morphism such that the composition $M' \xrightarrow{f} M \xrightarrow{t} M'$ is id_M. We show that $\text{Im } f$ is a summand of M by showing that $\text{Ker } t$ is a complement for $\text{Im } f$ in M. Because $M' \xrightarrow{f} M \xrightarrow{t} M'$ is id_M, it follows that if x is in $\text{Im } f \cap \text{Ker } t$, then $x = f(m')$ for some m' in M', and $t(x) = tf(m') = 0$. But $tf(m') = m'$ because $tf = \text{id}_{M'}$ which implies that $m' = 0$ and hence $x = f(m') = 0$. Therefore, we have that $\text{Im } f \cap \text{Ker } t = 0$. Hence, we are done if we show that M is generated by $\text{Im } f$ and $\text{Ker } t$. But the submodule generated by $\text{Im } f$ and $\text{Ker } t$ is $t^{-1}(t(\text{Im } f))$. The fact that $tf = \text{id}_{M'}$ implies that $t(\text{Im } f) = M'$. Hence, $t^{-1}(t(\text{Im } f)) = t^{-1}(M') = M$, which shows that M is generated by $\text{Im } f$ and $\text{Ker } t$.

In connection with this result we point out the following often used terminology.

Definitions

Let $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ be an exact sequence of R-modules.

(a) The monomorphism f is said to be a **splittable monomorphism** if there is a morphism $t: M \to M'$ such that the composition $M' \xrightarrow{f} M \xrightarrow{t} M'$ is id_M. Any morphism $t: M \to M'$ such that $tf = \text{id}_{M'}$ is called a **splitting for** f.

(b) The epimorphism $g: M \to M''$ is said to be a **splittable epimorphism** if there is a morphism $s: M'' \to M$ such that the composition $M'' \xrightarrow{s} M \xrightarrow{g} M''$ is $\text{id}_{M''}$. Any morphism $s: M'' \to M$ with the property $gs = \text{id}_{M''}$ is called a **splitting for the epimorphism** g.

(c) The exact sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is said to be **splittable** if either f is a splittable monomorphism or, equivalently, g is a splittable epimorphism.

On the basis of what has already been established, the reader should not have difficulty in proving the following.

Basic Properties 10.6

(a) Let $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ be an exact sequence of R-modules.
  (i) A morphism $t : M \to M'$ is a splitting for the monomorphism f if and only if $\operatorname{Ker} t$ and $\operatorname{Im} f$ are complementary submodules of M.
  (ii) A morphism $s : M'' \to M$ is a splitting for the epimorphism $g : M \to M''$ if and only if $\operatorname{Im} s$ and $\operatorname{Ker} g$ are complementary submodules of M.

(b) $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is a splittable exact sequence of R-modules if and only if there are morphisms $t : M \to M'$ and $s : M'' \to M$ satisfying:
  (i) $gf = 0$.
  (ii) $tf = \mathrm{id}_M$, and $gs = \mathrm{id}_{M''}$.
  (iii) $ts = 0$.
  (iv) $ft + sg = \mathrm{id}_M$.

PROOF: (a) Left as an exercise.

(b) Suppose $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is an exact splittable sequence of R-modules. Then let $s : M'' \to M$ be a splitting for g. For every m in M the element $m - sg(m)$ is in $\operatorname{Ker} g = \operatorname{Im} f$. Then the map $t : M \to M'$ defined by $t(m) = f_0^{-1}(m - sg(m))$ is our desired morphism, which we leave to the reader to verify as well as the rest of (b).

These criteria for when an exact sequence is splittable yield the following useful proposition.

Proposition 10.7

Let $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ be an exact sequence of R-modules. Then the following statements are equivalent:

(a) The exact sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is splittable.

(b) For each R-module X the sequence of $C(R)$-modules

$$0 \longrightarrow \operatorname{Hom}_R(X, M') \xrightarrow{\operatorname{Hom}_R(X, f)} \operatorname{Hom}_R(X, M) \xrightarrow{\operatorname{Hom}_R(X, g)} \operatorname{Hom}_R(X, M'') \longrightarrow 0$$

is exact.

(c) For each R-module X, the sequence of $C(R)$-modules

$$0 \longrightarrow \operatorname{Hom}_R(M'', X) \xrightarrow{\operatorname{Hom}_R(g, X)} \operatorname{Hom}_R(M, X) \xrightarrow{\operatorname{Hom}_R(f, X)} \operatorname{Hom}_R(M', X) \longrightarrow 0$$

is exact.

PROOF: We only prove the equivalence of (a) and (b). The equivalence of (a) and (c) proceeds in an analogous fashion and is left as an exercise.

(a) implies (b). Suppose the exact sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is splittable. Then there is a splitting for the epimorphism g, that is, a morphism $s : M'' \to M$ such that the composition $M'' \xrightarrow{s} M \xrightarrow{g} M''$ is $\mathrm{id}_{M''}$. Hence, for each R-module X, the composition of $C(R)$-modules

$$\operatorname{Hom}_R(X, M'') \xrightarrow{\operatorname{Hom}_R(X, s)} \operatorname{Hom}_R(X, M) \xrightarrow{\operatorname{Hom}_R(X, g)} \operatorname{Hom}_R(X, M'')$$

is the identity because $\operatorname{Hom}_R(X, g)\operatorname{Hom}_R(X, s) = \operatorname{Hom}_R(X, gs) = \operatorname{Hom}_R(X, \mathrm{id}_{M''}) = \mathrm{id}_{\operatorname{Hom}_R(X, M'')}$. Therefore, for each R-module X, the $C(R)$-morphism

$\mathrm{Hom}_R(X, g): \mathrm{Hom}_R(X, M) \to \mathrm{Hom}_R(X, M'')$ is an epimorphism. Combining this with the fact that the sequence

$$0 \longrightarrow \mathrm{Hom}_R(X, M') \xrightarrow{\mathrm{Hom}_R(X, f)} \mathrm{Hom}_R(X, M) \xrightarrow{\mathrm{Hom}_R(X, g)} \mathrm{Hom}_R(X, M'')$$

is always exact, we obtain that the sequence of $C(R)$-modules

$$0 \longrightarrow \mathrm{Hom}_R(X, M') \xrightarrow{\mathrm{Hom}_R(X, f)} \mathrm{Hom}_R(X, M) \xrightarrow{\mathrm{Hom}_R(X, g)} \mathrm{Hom}_R(X, M'') \longrightarrow 0$$

is exact for all R-modules X.

(b) implies **(a)**. If, for all X, the sequence of $C(R)$-modules

$$0 \longrightarrow \mathrm{Hom}_R(X, M') \xrightarrow{\mathrm{Hom}_R(X, f)} \mathrm{Hom}_R(X, M) \xrightarrow{\mathrm{Hom}_R(X, g)} \mathrm{Hom}_R(X, M'') \longrightarrow 0$$

is exact, then in particular, the sequence

$$\mathrm{Hom}_R(M'', M) \xrightarrow{\mathrm{Hom}_R(M'', g)} \mathrm{Hom}_R(M'', M'') \longrightarrow 0$$

is exact. Therefore, there is a morphism $t: M'' \to M$ such that $\mathrm{Hom}_R(M'', g)(t) = \mathrm{id}_{M''}$. But $\mathrm{Hom}_R(M'', g)(t) = gt$. So we see that $gt = \mathrm{id}_{M''}$, which means that $t: M'' \to M$ is a splitting for the epimorphism $g: M \to M''$. Hence, the sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is splittable. The reader should notice that condition (b) in Proposition 10.7 simply states that the sequence is exact. However, the exactness for all X immediately implies that the sequence splits as an exact sequence of $C(R)$-modules.

As an interesting example of how all these notions can be used, we will prove that if M' is a submodule of a free module M of rank n over a PID, then M' is a free module of rank at most n. We begin with the following preliminary result which we leave to the reader to prove.

Proposition 10.8
Let M be an R-module and let R be an arbitrary ring. Suppose M' and M'' are submodules of M which are free R-modules with bases B' and B'', respectively. Then M' is a complement of M'' in M if and only if:

(a) $B' \cap B'' = \emptyset$.
(b) M is a free R-module with basis $B' \cup B''$.

We now turn our attention to the main point of this digression.

Theorem 10.9
Let R be a PID. Suppose N is a submodule of the finitely generated free R-module M of rank n. Then N is a finitely generated free R-module of rank at most n.

PROOF: The proof is based on the fact established earlier that since R is a PID, every submodule of R is a free R-module of rank at most one. Keeping this fact in mind, we prove the theorem by induction on n, the rank of M.

If $n = 0$, there is nothing to prove because $M = (0)$. Suppose $n = 1$. If b is the unique element of the basis of M, then the morphism of R-modules $f: R \to M$

given by $f(r) = rb$ for all r in R is an isomorphism of R-modules. Hence, the submodule M' of M is isomorphic to the submodule $f^{-1}(M')$ of R. By our initial observation, $f^{-1}(M')$ is a free R-module of rank at most one which means that M' is a free submodule of M of rank at most one. Thus, the theorem has been established for $n = 1$.

Suppose now that the theorem is true for all free R-modules of rank at most n ($n \geq 1$) and suppose M' is a submodule of the free R-module M with rank $M = n+1$. Let $\{b_1, \ldots, b_{n+1}\}$ be a basis for M and let $f: M \to R$ be the unique morphism with the property $f(b_1) = 1$ and $f(b_i) = 0$ for all $i \geq 2$. Then an element $m = \sum_{i=1}^{n+1} r_i b_i$ in M is in Ker f if and only if $r_1 = 0$. This implies that Ker f is the submodule of M generated by the linearly independent set $\{b_2, \ldots, b_{n+1}\}$. Therefore, Ker f is a free module of rank n.

Now letting $g: N \to R$ be the morphism $f|N$, we know that Ker $g = N \cap$ Ker f. Hence, Ker g is contained in the free R-module Ker f of rank n. Therefore, by our inductive hypothesis we know that Ker g is a free R-module of rank $m \leq n$. Because Im g is a submodule R, it is a free R-module of rank 0 or 1. Hence, we have the exact sequence $0 \to \text{Ker } g \xrightarrow{\text{inc}} N \xrightarrow{g_0} \text{Im } g \to 0$ where Im g is a free R-module of rank 0 or 1.

If rank Im $g = 0$, then Im $g = (0)$ which means that $N = \text{Ker } g$ and so N is a free R-module of rank $m \leq n$ and we are done.

Suppose rank Im $g = 1$. Then Im g has a basis consisting of one element b. Because $g_0: N \to \text{Im } g$ is surjective, there is an element x in N such that $g_0(x) = b$. Let $h: \text{Im } g \to N$ be the morphism of R-modules with the property that $h(b) = x$. Then the composition $\text{Im } g \xrightarrow{h} N \xrightarrow{g_0} \text{Im } g$ has the property $g_0 h(b) = b$. Because $\{b\}$ is a basis for Im g and the two morphisms $g_0 h$ and $\text{id}_{\text{Im } g}$ agree on that basis, they are the same, that is, $g_0 h = \text{id}_{\text{Im } g}$.

Therefore, it follows from Proposition 10.5 that Im h is a complement for Ker g in N. Further, $h_0: \text{Im } g \to \text{Im } h$ is an isomorphism and $x = h_0(b)$. Hence, $\{x\}$ is a basis for Im h since $\{b\}$ is a basis for Im g. Therefore, if we let B' be a basis for Ker g, we have by Proposition 10.8 that N is a free R-module with basis $B' \cup \{x\}$ where x is not in B'. Because B' has m elements and $m \leq n$, it follows that N is a free R-module with rank $N = m + 1 \leq n + 1 = \text{rank } M$, which is our desired result.

Corollary 10.10
Let R be a PID. If M is an R-module which can be generated by n elements, then every submodule of M can also be generated by m elements where $m \leq n$.

PROOF: Suppose S is a set of generators for M with n elements. Then the free R-module $F(S)$ generated by S has rank n and the unique morphism $f: F(S) \to M$ such that $f|S = \text{inc}$ is an epimorphism because S generates M (see Proposition 7.9). Let N be a submodule of M. Then by Theorem 10.9, we know that $f^{-1}(N)$ is a free submodule of M of rank $m \leq n$. Hence, $f^{-1}(N)$ can be generated by m elements, which implies that $f(f^{-1}(N)) = N$ can also be generated by m elements, where $m \leq n$.

11. SUMS OF MODULES

In an earlier chapter the notion of a sum of an indexed family of objects in an arbitrary category was introduced. This section is devoted to discussing the notion of sums in the category of R-modules for an arbitrary ring R. We start by looking at complements of submodules from the point of view of sums of modules. To this end, it is useful to have yet another easily verified description of complementary submodules of a module.

Proposition 11.1
Let M' and M'' be submodules of an R-module M.

(a) $M' \cap M'' = (0)$ if and only if whenever $m_1' + m_1'' = m_2' + m_2''$ with m_1' and m_2' in M' and m_1'' and m_2'' in M'', we have $m_1' = m_2'$ and $m_1'' = m_2''$.
(b) M' and M'' are complementary submodules in M if and only if every element m in M can be written in one and only one way as a sum $m' + m''$ with m' in M' and m'' in M''.

We now use this result to describe when a submodule M' of a module M is a complement for a submodule M'' of M in terms of the morphisms from M', M'', and M to various R-modules.

Proposition 11.2
Let M' and M'' be submodules of the R-module M. Then the following statements are equivalent:

(a) M' and M'' are complementary submodules of M.
(b) For each R-module X and pair of morphisms $f' : M' \to X$ and $f'' : M'' \to X$, there is one and only one morphism $f : M \to X$ such that $f|M' = f'$ and $f|M'' = f''$.
(c) There are morphisms $p_{M'} : M \to M'$ and $p_{M''} : M \to M''$ satisfying:
  (i) The composition $M' \xrightarrow{\text{inc}} M \xrightarrow{p_{M'}} M'$ is $\mathrm{id}_{M'}$ and $M'' \xrightarrow{\text{inc}} M \xrightarrow{p_{M''}} M'$ is the zero morphism.
  (ii) The composition $M'' \xrightarrow{\text{inc}} M \xrightarrow{p_{M''}} M''$ is $\mathrm{id}_{M''}$ and $M' \xrightarrow{\text{inc}} M \xrightarrow{p_{M'}} M''$ is the zero morphism.
  (iii) For each m in M we have $m = p_{M'}(m) + p_{M''}(m)$.

PROOF: (a) implies (b). Suppose M'' is a complement of M' in M. Then, by our previous result, we know that each element m in M can be written in one and only one way as a sum $m' + m''$ with m' in M' and m'' in M''. Hence, given any R-module X and morphisms $f' : M' \to X$ and $f'' : M'' \to X$, we can define a map $f : M \to X$ by $f(m) = f'(m') + f''(m'')$ where m' and m'' are the unique elements in M' and M'', respectively, such that $m = m' + m''$. It is easy to check that f is actually a morphism of R-modules with the properties $f|M' = f'$ and $f|M'' = f''$. That this is the only morphism with these properties follows from the fact that $M' \cup M''$ generates M.

(b) implies (c). Suppose M' and M'' are submodules of the R-module M with the property that given any R-module X and morphisms $f' : M' \to X$ and $f'' : M'' \to X$, there is a unique $f : M \to X$ such that $f|M' = f'$ and $f|M'' = f''$. Then define

$p_{M'}: M \to M'$ to be the unique morphism with the property that $p_{M'}|M': M' \to M'$ is the identity and $p_{M'}|M'': M'' \to M'$ is the zero morphism. Then $p_{M'}: M \to M'$ clearly satisfies part (i). Similarly, define $p_{M''}: M \to M''$ to be the unique morphism with the property that $p_{M''}|M'': M'' \to M''$ is the identity and $p_{M''}|M': M' \to M''$ is the zero morphism. Then $p_{M''}: M \to M''$ clearly satisfies part (ii). Therefore, in order to finish the proof, we must show that $p_{M'}$ and $p_{M''}$ satisfy condition (iii).

To this end, consider the morphism $h: M \to M$ given by $h(m) = p_{M'}(m) + p_{M''}(m)$ for all m in M. Now for each m' in M', we have $h(m') = p_{M'}(m') + p_{M''}(m') = p_{M'}(m') = m'$ and similarly, $h(m'') = m''$ for each m'' in M''. In other words, the morphism $h: M \to M$ has the property that $h|M': M' \to M$ is the inclusion morphism and $h|M'': M'' \to M$ is the inclusion morphism. But $\mathrm{id}_M: M \to M$ also has these properties. Therefore, $h = \mathrm{id}_M$ because the submodules M' and M'' of M by hypothesis have the property that two morphisms with domain M are the same if their restrictions to M' and M'' are the same. Thus, we have $m = h(m) = p_{M'}(m) + p_{M''}(m)$ for each m in M, which finishes the proof that (b) implies (c).

(c) implies (a). Left as an exercise.

Restating part (b) of this proposition in slightly different terms, we see that M' and M'' are complementary submodules of M if and only if for each R-module X, the map

$$\varphi_X: \mathrm{Hom}_R(M, X) \to \mathrm{Hom}_R(M', X) \times \mathrm{Hom}_R(M'', X)$$

given by $\varphi_X(f) = (f|M', f|M'')$ is an isomorphism of sets. What this in essence says is that M together with the morphisms $M' \xrightarrow{\text{inc}} M$ and $M'' \xrightarrow{\text{inc}} M$ is the sum of the R-modules M' and M'' in the category of R-modules if and only if M' and M'' are complementary submodules of M.

Specializing the definition of a sum of an arbitrary family of objects in an arbitrary category to the category of R-modules, we obtain the following.

Definition

Let $\{M_i\}_{i \in I}$ be a family of R-modules. A module M together with a family of morphisms $\{g_i: M_i \to M\}_{i \in I}$ is said to be a **sum of the family** of R-modules if and only if for each R-module X, the map of sets $\varphi_X: \mathrm{Hom}_R(M, X) \to \prod_{i \in I} \mathrm{Hom}_R(M_i, X)$ given by $\varphi_X(f) = \{fg_i\}_{i \in I}$ for all f in $\mathrm{Hom}_R(M, X)$ is an isomorphism of sets.

We recall that when discussing sums in arbitrary categories, we explained in what sense a sum of a family of objects in a category is unique. Specializing that discussion to the category R-modules, we have the following.

Proposition 11.3

Let $\{M_i\}_{i \in I}$ be a family of R-modules. Suppose that R-modules M and M' together with the family of morphisms $\{g_i: M_i \to M\}_{i \in I}$ and $\{g'_i: M_i \to M'\}_{i \in I}$ are sums for $\{M_i\}_{i \in I}$. Then there is a unique morphism $h: M \to M'$ such that $hg_i = g'_i$ for all i in I and this uniquely determined morphism is an isomorphism.

Having defined, as well as having established, the uniqueness of a sum of a family of R-modules $\{M_i\}_{i \in I}$, it is natural to ask if every family of R-modules has a sum.

We recall that this question was answered in the affirmative in the special case when R is the ring of intergers $\mathbf{Z}$. Given a family $\{A_i\}_{i\in I}$ of $\mathbf{Z}$-modules, or, what is the same thing, abelian groups, we constructed the abelian group $\coprod_{i\in I} A_i$ as follows. As a set $\coprod_{i\in I} A_i$ is the subset of $\prod_{i\in I} A_i$ consisting of all elements $(a_i)_{i\in I}$ satisfying $a_i = 0$ for all but a finite number of i in I. The addition in $\coprod_{i\in I} A_i$ is given by $(a_i)_{i\in I} + (a'_i)_{i\in I} = (a_i + a'_i)_{i\in I}$ for all $(a_i)_{i\in I}$ and $(a'_i)_{i\in I}$ in $\coprod_{i\in I} A_i$, which makes $\coprod_{i\in I} A_i$ an abelian group. Also, for each k in I we defined the morphism $f_k : A_k \to \coprod_{i\in I} A_i$ by $f_k(a)$ is the element $(a_i)_{i\in I}$ with the property $a_i = 0$ for $i \neq k$ and $a_i = a$ for $i = k$ for each element a in A_k. Finally, we showed that the family of morphisms $\{f_k : A_k \to \coprod_{i\in I} A_i\}_{k\in I}$ is a sum for the given family of abelian groups $\{A_i\}_{i\in I}$. We recall that the morphisms of abelian groups $f_k : A_k \to \coprod_{i\in I} A_i$ are monomorphisms which we called the canonical injections of A_k into $\coprod_{i\in I} A_i$. This suggests the following construction of the sum $\coprod_{i\in I} A_i$ of a family of R-modules $\{A_i\}_{i\in I}$ over an arbitrary ring R. Because each R-module A_i is an abelian group, we can form their sum $\coprod_{i\in I} A_i$ as abelian groups in the manner just described. We leave it to the reader to check that the map $R \times \coprod_{i\in I} A_i \to \coprod_{i\in I} A_i$ given by $(r, (a_i)_{i\in I}) \mapsto (ra_i)_{i\in I}$ for all r in R and $(a_i)_{i\in I}$ in $\coprod_{i\in I} A_i$ is an R-module structure. The abelian group $\coprod_{i\in I} A_i$ together with this R-module structure will also be denoted by $\coprod_{i\in I} A_i$.

Now it is easily checked that for each k in I, the canonical injection $f_k : A_k \to \coprod_{i\in I} A_i$ is not only a morphism of abelian groups, but is also an R-module morphism. Finally, it is not difficult to check that the family $\{f_k : A_k \to \coprod_{i\in I} A_i\}_{k\in I}$ of morphisms is a sum of the family of R-modules $\{A_i\}_{i\in I}$. For the sake of completeness we outline a proof of this fact.

Suppose X is an arbitrary R-module and $\{g_k : A_k \to X\}_{i\in I}$ is an arbitrary family of R-module morphisms. We have to show that there is a unique morphism $f : \coprod_{i\in I} A_i \to X$ with the property that for each k in I, the composition $A_k \xrightarrow{f_k} \coprod_{i\in I} A_i \xrightarrow{f} X$ is the same as the given morphism $g_k : A_k \to X$. To this end we first observe that if $(a_i)_{i\in I}$ is an element of $\coprod_{i\in I} A_i$, then the family $(g_i(a_i))_{i\in I}$ of elements in X is an almost zero family of elements in X because all but a finite number of the a_i are zero. Hence, we can define a map $f : \coprod_{i\in I} A_i \to X$ by $f((a_i)_{i\in I}) = \Sigma_{i\in I} g_i(a_i)$ for all (a_i) in $\coprod_{i\in I} A_i$. We leave it to the reader to show that this map f is an R-module morphism having our desired property that for each k in I, the composition $A_k \xrightarrow{f_k} \coprod_{i\in I} A_i \xrightarrow{f} X$ is the given morphism $g_k : A_k \to X$. The fact that this is the only morphism from $\coprod_{i\in I} A_i$ to X satisfying this condition follows from the observation that $\coprod_{i\in I} A_i$ is generated by the family $\{\operatorname{Im} f_k\}_{k\in I}$ of submodules of $\coprod_{i\in I} A_i$.

We summarize part of this discussion in the following.

Definitions
Let $\{A_i\}_{i\in I}$ be a family of R-modules.

(a) We denote by $\coprod_{i\in I} A_i$ the R-module consisting of the abelian group $\coprod_{i\in I} A_i$ together with the R-module structure $R \times \coprod_{i\in I} A_i \to \coprod_{i\in I} A_i$ given by $(r, (a_i)_{i\in I}) \to (ra_i)_{i\in I}$ for all r in R and $(a_i)_{i\in I}$ in $\coprod_{i\in I} A_i$.

(b) The R-module morphisms $f_k : A_k \to \coprod_{i\in I} A_i$ defined by $f_k(a) = (a_i)_{i\in I}$ where $a_i = 0$ for $i \neq k$ and $a_i = a$ for $i = k$ for all a in A_k, is called the **kth injection morphism** from A_k to $\coprod_{i\in I} A_i$.

(c) The sum of the family $\{A_i\}_{i\in I}$ of R-modules consisting of the family of morphisms $\{f_k : A_k \to \coprod_{i\in I} A_i\}_{k\in I}$ is called the **standard sum of the family** $\{A_i\}_{i\in I}$ of R-modules.

(d) Given any R-module X and any family $\{g_k : A_k \to X\}_{k\in I}$ of R-module morphisms we denote by $\coprod_{i\in I} g_i : \coprod_{i\in I} A_i \to X$ the unique morphism from $\coprod_{i\in I} A_i$ to X with the property that $g_k = (\coprod_{i\in I} g_i) f_k$ for all k in I. The morphism $\coprod_{i\in I} g_i$ is called the **sum of the family of morphisms** $\{g_k : A_k \to X\}$ and can be described by $\coprod_{i\in I} g_i((a_i)_{i\in I})$ is the sum $\Sigma_{i\in I} g_i(a_i)$ of the almost zero family $(g_i(a_i))_{i\in I}$ of elements in X for each element $(a_i)_{i\in I}$ of $\coprod_{i\in I} A_i$.

Having shown that every family of R-modules has a sum in the category of R-modules, we now turn our attention to some basic properties of sums of families of R-modules. Although these will be stated in terms of arbitrary sums, the reader may find it helpful to see the form these results take in the specific case of standard sums.

Basic Properties 11.4
Suppose $\{f_i : A_i \to A\}_{i\in I}$ is a sum for the family $\{A_i\}_{i\in I}$ of R-modules. Then:

(a) If $g, h : A \to X$ are two morphisms of R-modules, then $g = h$ if and only if $gf_i = hf_i$ for all i in I.
(b) The R-module A is generated by the family of submodules $\{\text{Im } f_i\}_{i\in I}$.
(c) For each k in I, there is one and only one morphism $p_k : A \to A_k$ such that $p_k f_i = 0$ if $k \neq i$ and $p_i f_i = \text{id}_{A_i}$.
(d) For each k in I, Ker p_k is generated by the set of all submodules Im f_i of A with $i \neq k$.
(e) For each k in I, the morphism $f_k : A_k \to A$ is a splitting monomorphism having the property that Ker p_k is a complement for Im f_k.

PROOF: (a) Follows from the definition of a sum of a family of modules.
(b) It follows from (a) that the subset $S = \bigcup_{i\in I} \text{Im } f_i$ of A has the property that two morphisms $g, h : A \to X$ are the same if $g(x) = h(x)$ for all x in S. We have already seen that this property of the set S assures that S generates A.

(c) Because $\{f_i : A_i \to A\}_{i\in I}$ is a sum of the family $\{A_i\}_{i\in I}$ of R-modules, we know that given any R-module X and any family $\{g_i : A_i \to X\}_{i\in I}$ of R-module

morphisms there is a unique morphism $g: A \to X$ such that $gf_i = g_i$ for all i in I. For each k in I define the family of morphisms $\{g_{ik}: A_i \to A_k\}_{i \in I}$ by $g_{ik} = 0$ if $k \neq i$ and $g_{ii} = \mathrm{id}_{A_i}$. Then by our initial observation, there is a unique R-module morphism $p_k: A \to A_k$ satisfying $p_k f_i = g_{ik}$ for all i in I. Hence, $p_k f_k = \mathrm{id}_{A_k}$ while $p_k f_i = 0$ for $i \neq k$, which establishes (c).

(d) Because $p_k f_i = 0$ for $i \neq k$, it follows that $\mathrm{Im}\, f_i \subset \mathrm{Ker}\, p_k$ for each $i \neq k$. Hence, to show that $\mathrm{Ker}\, p_k$ is generated by the set of all submodules $\mathrm{Im}\, f_i$ with $i \neq k$, it suffices to show that for each element x in $\mathrm{Ker}\, p_k$ there is a finite set of distinct elements $i_1, \ldots, i_n$ in I, all of which are different from k, such that $x = \sum_{j=1}^{n} f_{i_j}(a_{i_j})$ where each a_{i_j} is in A_{i_j}. Because A is generated by the family of submodules $\{\mathrm{Im}\, f_i\}_{i \in I}$, we know that given an element x in $\mathrm{Ker}\, p_k$, there is a finite set $i_1, \ldots, i_n$ of distinct elements in I such that $x = \sum_{j=1}^{n} f_{i_j}(a_{i_j})$ where each a_{i_j} is in A_{i_j}. If none of the $i_j = k$, then we are done. Suppose one of the $i_j = k$, say $i_1 = k$. Then $p_k(x) = p_k f_{i_1}(a_{i_1}) + \sum_{j>1} p_k f_{i_j}(a_{i_j}) = a_{i_1}$, because $p_k f_{i_1} = p_k f_k = \mathrm{id}_{A_k}$ and $p_k f_{i_j} = 0$ for all $j > 1$. Hence, $a_{i_1} = p_k(x) = 0$ because x is in $\mathrm{Ker}\, p_k$. Therefore, $x = \sum_{j>1} f_{i_j}(a_{i_j})$ where no $i_j = k$. This shows that each element in $\mathrm{Ker}\, p_k$ can be written as a finite sum of elements in $\mathrm{Im}\, f_i$ with $i \neq k$ or, what is the same thing, $\mathrm{Ker}\, p_k$ is generated by the set of all submodules $\mathrm{Im}\, f_i$ with $i \neq k$.

(e) Because the morphism $p_k: A \to A_k$ has the property $p_k f_k = \mathrm{id}_{A_k}$, it follows that $f_k: A_k \to A$ is a splitting monomorphism and $\mathrm{Ker}\, p_k$ is a complement for $\mathrm{Im}\, f_k$.

Because the morphisms $p_i: A \to A_i$ described in the above basic properties play an important role in studying sums of R-modules, we make the following definition.

Definition

Let $\{A_i\}_{i \in I}$ be a family of R-modules and $\{f_i: A_i \to A\}_{i \in I}$ a sum for this family. For each k in I, the **kth projection morphism is the** unique morphism $p_k: A \to A_k$ having the properties $p_k f_k = \mathrm{id}_{A_k}$ and $p_k f_i = 0$ for $i \neq k$.

We now point out the following basic properties for projection morphisms.

Basic Properties 11.5

Suppose the family $\{f_i: A_i \to A\}_{i \in I}$ of R-module morphisms is a sum for the family $\{A_i\}_{i \in I}$ of R-modules. Suppose $\{p_i: A \to A_i\}_{i \in I}$ are the projection morphisms for this sum. Then:

(a) For each x in A, there are only a finite number of i in I such that $p_i(x) \neq 0$.

(b) For each x in A, we have $x = \sum_{i \in I} f_i p_i(x)$. Hence, two elements x and y in A are equal if and only if $p_i(x) = p_i(y)$ for all i in I.

(c) The map $t: A \to \coprod_{i \in I} A_i$ given by $t(x) = (p_i(x))_{i \in I}$ for each x in A is an isomorphism of R-modules.

PROOF: **(a)** Let x be in A. Because A is generated by the family of submodules $\{\mathrm{Im}\, f_i\}_{i \in I}$, it follows that there is a finite set $i_1, \ldots, i_n$ of elements in I such that $x = \sum_{j=1}^{n} f_{i_j}(a_{i_j})$ with the a_{i_j} in A_{i_j}. Hence, if k is not one of the $i_1, \ldots, i_n$, then $p_k(x) = \sum_{j=1}^{n} p_k f_{i_j}(a_{i_j}) = 0$ since $p_k f_{i_j} = 0$ because $k \neq i_j$. Therefore, $p_k(x) = 0$ for all but a finite number of k in I.

(b) For each x in A, because the elements $p_k(x)$ in A_k are zero for all but a

finite number of k, it follows that for each a in A we know that $(f_i p_i(a))_{i\in I}$ is an almost zero family of elements in A. Hence, we can form the sum $\Sigma_{i\in I} f_i p_i(a)$ for each a in A. Thus, we can define the map $f: A \to A$ by $f(a) = \Sigma_{i\in I} f_i p_i(a)$ for each i in I which is easily seen to be a morphism of R-modules. Our aim is to show that $f = \mathrm{id}_A$, which will give our desired result that $a = \Sigma_{i\in I} f_i p_i(a)$ for each a in A.

Now id_A obviously has the property $\mathrm{id}_A f_i = f_i$ for all i in I. Therefore, if we show that f also has the property $ff_i = f_i$ for all i in I, then we will have that $f = \mathrm{id}_A$ since we have already seen that two morphisms g, $h: A \to X$ are the same if $gf_i = hf_i$ for all i in I. Let a_k in A_k. Then $f(f_k(a_k)) = \Sigma_{i\in I} f_i p_i(f_k(a_k)) = f_k p_k(f_k(a_k))$ because $p_i f_k = 0$ if $i \neq k$. Because $p_k f_k = \mathrm{id}_{A_k}$, it then follows that $f(f_k(a_k)) = f_k(a_k)$ for all a_k in A_k or, equivalently, $ff_k = f_k$. Because this is true for each k in I, we have the result that $ff_i = f_i$ for all i in I and hence $f = \mathrm{id}_A$. The rest of (b) is obvious.

(c) For each x in A, because we have that $p_i(x) = 0$ for all but a finite number of i in I, it follows that $(p_i(x))_{i\in I}$ is an element of $\coprod_{i\in I} A_i$ for each x in A. Hence, we can define a map $t: A \to \coprod_{i\in I} A_i$ by $t(x) = (p_i(x))_{i\in I}$ for each x in A. We leave it to the reader to check that this map is a morphism of R-modules.

We also can define the map $s: \coprod_{i\in I} A_i \to A$ by $s((a_i)_{i\in I}) = \Sigma_{i\in I} f_i(a_i)$, since $(f_i(a_i))_{i\in I}$ is an almost zero family of elements in A for each element $(a_i)_{i\in I}$ in $\coprod_{i\in I} A_i$. We leave it to the reader to check that this map is a morphism of R-modules. Now for each x in A, we have $st(x) = s(p_i(x))_{i\in I} = \Sigma_{i\in I} f_i p_i(x) = x$. Hence, $st = \mathrm{id}_A$. Hence, we will have shown that t is an isomorphism with inverse s, if we show that $ts = \mathrm{id}_{\coprod_{i\in I} A_i}$.

But for each $(a_i)_{i\in I}$ in $\coprod_{i\in I} A_i$ we have $ts((a_i)_{i\in I}) = t(\Sigma_{i\in I} f_i(a_i)) = (p_k(\Sigma_{i\in I} f_i(a_i)))_{k\in I}$. Because $p_k \Sigma_{i\in I} f_i(a_i) = \Sigma_{i\in I} p_k f_i(a_i)$ and $p_k f_i = 0$ for $i \neq k$ and $p_k f_k = \mathrm{id}_{A_k}$, it follows that $p_k(\Sigma_{i\in I} f_i(a_i)) = a_k$, for each k in I. Hence, $ts = \mathrm{id}_{\coprod_{i\in I} A_i}$, which finishes the proof of (c).

In connection with part (c) of these basic properties, we outline the following alternate proof of the fact that the morphism $t: A \to \coprod_{i\in I} A_i$ given by $t(a) = (p_i(a))_{i\in I}$ for each a in A is an isomorphism. This alternate proof is based on the fact that for each k in I the composition $tf_k: A_k \to \coprod_{i\in I} A_i$ is the kth inclusion morphism $j_k: A_k \to \coprod_{i\in I} A_i$ since $tf_k(a_k) = (p_i f_k(a_k))_{i\in I}$ while $p_i f_k(a_k) = 0$ for $k \neq i$ and $p_k f_k(a_k) = a_k$. Because $\{f_i: A_i \to A\}_{i\in I}$ and $\{j_i: A_i \to \coprod_{i\in I} A_i\}_{i\in I}$ are both sums for the family $\{A_i\}_{i\in I}$ of R-modules, we know by the uniqueness theorem for sums of R-modules that there is only one morphism $s: A \to \coprod_{i\in I} A_i$ such that $sf_i = j_i$ for all i in I and this uniquely determined morphism is an isomorphism. Because we have shown that $tf_i = j_i$ for all i in I, we know that $t = s$ and is therefore an isomorphism.

We now point out an important special type of sum of a family of R-modules.

Proposition 11.6

Let M be an R-module and $\{M_i\}_{i\in I}$ a family of submodules of M. Then the following statements are equivalent:

(a) The family $\{\text{inc}: M_i \to M\}_{i \in I}$ of morphisms is a sum for the family $\{M_i\}_{i \in I}$ of R-modules.

(b) The family $\{M_i\}_{i \in I}$ of submodules of M satisfies:
  (i) M is generated by $\{M_i\}_{i \in I}$.
  (ii) For each j in I, let N_j be the submodule of M generated by $\{M_i\}_{i \in I-\{j\}}$. Then $N_j \cap M_j = 0$ for each j in I.

(c) For each m in M, there is a unique almost zero family of elements $\{m_i\}_{i \in I}$ in M satisfying:
  (i) m_i is in M_i for each i in I.
  (ii) $m = \Sigma_{i \in I}\, m_i$.

PROOF: (a) implies (b). The fact that M is generated by the family $\{M_i\}_{i \in I}$ of submodules of M follows from Basic Properties 11.4. Again by 11.4 we know that the projection morphisms $p_k : M \to M_k$ have the property that $p_k | M_i = 0$ for $i \neq j$ and $p | M_k = \text{id}_{M_k}$ for each k in I. Suppose now that m is in $N_j \cap M_j$ for some j in I. Then $p_j(m) = m$ because m is in M_j. On the other hand, $m = m_1 + \cdots + m_n$ where each m_k is in some M_i with $i \neq j$ because N_j is generated by $\{M_i\}_{i \in I-\{j\}}$. Consequently, $p_j(m) = p_j(m_1) + \cdots + p_j(m_n) = 0$ because $p_j | M_i = 0$ for $i \neq j$. Therefore, we have $m = p_j(m) = 0$ which shows that $N_j \cap M_j = 0$ for all j in I.

(b) implies (c). The fact that M is generated by the family $\{M_i\}_{i \in I}$ of submodules of M implies that given any m in M there is an almost zero family $\{m_i\}_{i \in I}$ of elements in M satisfying $m = \Sigma_{i \in I}\, m_i$ and m_i is in M_i for each i in I. Suppose now that $\{m'_i\}_{i \in I}$ is another almost zero family of elements of M satisfying $m = \Sigma_{i \in I}\, m'_i$ and m'_i is in M_i for each i in I. Then $0 = m - m = \Sigma_{i \in I}\, m_i - \Sigma_{i \in I}\, m'_i = \Sigma_{i \in I}\, m_i - m'_i$. From this equation it follows that $m_j - m'_j$ is in $N_j \cap M_j = 0$ for each j in I. Hence, we have $m_j = m'_j$ for all j in I. Therefore, given any m in M, there is one and only one almost zero family $\{m_i\}_{i \in I}$ of elements in M satisfying $m = \Sigma_{i \in I}\, m_i$ where each m_i is in M_i.

(c) implies (a). For each m in M, let $\{m_i\}_{i \in I}$ denote the unique almost zero family of elements in M satisfying $m = \Sigma_{i \in I}\, m_i$ where each m_i is in M_i. Now suppose we are given a family $\{f_i : M_i \to X\}_{i \in I}$ of morphisms of R-modules. Then define the map $f : M \to X$ by $f(m) = \Sigma_{i \in I}\, f(m_i)$ for each m in M. It is easily checked that f is a morphism of R-modules with the property $f | M_i = f_i$ for each i in I. Moreover, because M is generated by the family $\{M_i\}_{i \in I}$ of submodules of M (why?), $f : M \to X$ is the only morphism of R-modules such that $f | M_i = f_i$. This implies that the family of morphisms $\{\text{inc}: M_i \to M\}_{i \in I}$ is a sum for the family $\{M_i\}_{i \in I}$ of R-modules, as the reader can readily verify.

In connection with this result we make the following definition.

Definition

Suppose $\{M_i\}_{i \in I}$ is a family of submodules of an R-module M. We say that M is the sum of the family $\{M_i\}_{i \in I}$ if the family of morphisms $\{\text{inc}: M_i \to M\}_{i \in I}$ is a sum. We denote the fact that M is the sum of the family $\{M_i\}_{i \in I}$ by writing $M = \coprod_{i \in I} M_i$. If I is a finite set, say $[1, \ldots, n]$, then we will often write $M_1 \amalg \cdots \amalg M_n$ for $\coprod_{i \in I} M_i$.

We end this section with some examples of sums of R-modules.

Example 11.7 Let $\{R_i\}_{i \in I}$ be an indexed family of R-modules with each

$R_i = R$ for all i in I. For each j in I, let $\{\delta_{ij}\}_{i\in I}$ be the element of $\coprod_{i\in I} R_i$ satisfying $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{jj} = 1$. Then $\coprod_{i\in I} R_i$ is a free R-module with basis B consisting of the elements $\{\delta_{ij}\}_{i\in I}$ for all j in I.

PROOF: Follows immediately from the fact that $\coprod_{i\in I} R_i = F(I)$, the free R-module generated by I.

Because the type of bases, as well as the free R-modules, described in this example occur frequently, we make the following definition.

Definition
Let $\coprod_{i\in I} R_i$ be the standard sum of the family $\{R_i\}_{i\in I}$ of R-modules with $R_i = R$ for all i in I. Then the **standard basis for** $\coprod_{i\in I} R_i$ is the basis B of $\coprod_{i\in I} R_i$ consisting of all elements $\{r_i\}_{i\in I}$ with the property that there is a j in I such that $r_i = 0$ for $i \neq j$ and $r_j = 1$.

Example 11.8. Let M be a free R-module with basis B. Then letting $\{R_b\}_{b\in B}$ be the family of R-modules with $R_b = R$ for all b in B, the map $f\colon \coprod_{b\in B} R_b \to M$ given by $f(\{r_b\}_{b\in B}) = \Sigma_{b\in B} r_b b$ is an isomorphism of R-modules.

PROOF: See Proposition 7.9.

Combining these two examples we have the following.

Example 11.9 An R-module M is a free R-module if and only if there is an indexed family $\{R_i\}_{i\in I}$ of R-modules with $R_i = R$ for all i in I such that $M \approx \coprod_{i\in I} R_i$.

The reader is probably familiar with the fact that if V is an n-dimensional vector space over a field K, then the ring of endomorphisms of V can be represented as the ring of $n \times n$ matrices over K. We generalize this representation here in order to explain our next example.

Let R be an arbitrary ring, and M a free R-module with basis $B = \{b_1, \ldots, b_n\}$. If $f\colon M \to M$ is an R-endomorphism, then $f(b_i) = \Sigma_{j=1}^{n} r_{ij} b_j$ for all $i = 1, \ldots, n$. The elements r_{ij} are uniquely determined elements of R. Hence, associated with each f in $\mathrm{End}_R(M)$ is an $n \times n$ square array (r_{ij}) of elements of R, called a square matrix of order n over R. We shall denote by $M_n(R)$ the set of all square matrices over R of order n. We make $M_n(R)$ a ring by defining the following laws of composition:

$$(r_{ij}) + (r'_{ij}) = (r_{ij} + r'_{ij})$$
$$(r_{ij}) \cdot (r'_{ij}) = (s_{ij})$$

where

$$s_{ij} = \sum_{k=1}^{n} r_{ik} r'_{kj}$$

The reader should verify that $M_n(R)$, together with these laws of addition and multiplication, is a ring. The zero element is the matrix (r_{ij}) where $r_{ij} = 0$ for all i

and j. The one of $M_n(R)$ is the matrix (r_{ij}) where $r_{ij}=0$ if $i\neq j$, and $r_{ij}=1$ if $i=j$. Finally, the reader should verify that the map $\rho:\mathrm{End}_R(M)\to M_n(R)$ given by $\rho(f)=(r_{ij})$ for all f in $\mathrm{End}_R(M)$, where $f(b_i)=\sum_{j=1}^n r_{ij}b_j$, is an isomorphism of rings.

For each $k=1,\ldots,n$, let C_k be the subset of $M_n(R)$ consisting of all matrices (r_{ij}) such that $r_{ij}=0$ if $i\neq k$. The reader should verify that:

(a) Each C_k is a left ideal of $M_n(R)$ but not a right ideal of $M_n(R)$.
(b) Let e_{kl} be the matrix (s_{ij}) where $s_{ij}=0$ if $k\neq i$ or $l\neq j$ and $s_{kl}=1$. Show that each C_k is generated, as a left ideal, by the element e_{kl}.
(c) Show that $C_k e_{lk}$ is contained in C_l and that the map $g:C_k\to C_l$ defined by $g(x)=xe_{lk}$ is an isomorphism of $M_n(R)$-modules.

With these preliminary remarks out of the way, we can state the following.

Example 11.10 The ring $M_n(R)$ is the sum of the family of left ideals $\{C_k\}_{k=1,\ldots,n}$. That is, $M_n(R)=\coprod C_k$.

PROOF: Show that the family $\{C_k\}$ satisfies condition (c) of Proposition 11.6.

12. CHANGE OF RINGS

Until now we have considered modules over a fixed ring R. In this section we examine the connections that a ring morphism $f:R\to S$ gives between the modules over S and the modules over R.

Suppose $f:R\to S$ is a morphism of rings. If M is an S-module, then it is easily seen that the map $R\times M\to M$ given by $(r, m)\to f(r)m$ is an R-module structure on the underlying abelian group of M.

Definition

Let $f:R\to S$ be a morphism of rings. Suppose M is an S-module. The R-module consisting of the underlying abelian group of M together with the R-module structure given by $(r, m)\mapsto f(r)m$ is called the **R-module induced by f**. The R-module induced by f is usually denoted by the same symbol M and the R-module structure $(r, m)\mapsto f(r)m$ is written more simply as $(r, m)\mapsto rm$.

We now point out some obvious properties.

Basic Properties 12.1

Let $f:R\to S$ be a morphism of rings.

(a) Suppose M is an S-module. Then each S-submodule of M is also an R-submodule of the induced R-module M.
(b) Suppose $g:M\to M'$ is a morphism of S-modules. Then g is also a morphism of the induced R-modules M and M'. Further, Ker g, Im g, Coim g, Coker g are the same whether g is regarded as a morphism of S-modules or a morphism of R-modules. Hence:
(c) If $M'\to M\to M''$ is an exact sequence of S-modules, then it is also an exact sequence of R-modules.
(d) Suppose M and M' are S-modules. Because each S-morphism $g:M\to M'$ is

also an R-morphism we have that $\mathrm{Hom}_S(M, M') \subset \mathrm{Hom}_R(M, M')$. In fact, $\mathrm{Hom}_S(M, M')$ is a subgroup of $\mathrm{Hom}_R(M, M')$.

(e) For each S-module M, we have that $\mathrm{End}_S(M)$ is a subring of $\mathrm{End}_R(M)$.

(f) Suppose $\{g_i : M_i \to M\}_{i \in I}$ is a sum of the family of S-modules $\{M_i\}_{i \in I}$. Then $\{g_i : M_i \to M\}$ is also a sum of the family of R-modules $\{M_i\}_{i \in I}$.

PROOF: (a) through (e) are obvious.

(f) Clearly, it suffices to prove (f) in the case that $\{g_i : M_i \to M\}$ is the standard sum of the family $\{M_i\}_{i \in I}$ of S-modules. But in this special case the result is obvious.

In connection with this list of basic properties it is worthwhile considering the following.

Example 12.2 Let $\mathbf{Z} \to \mathbf{Z}[X]$ be the inclusion morphism of rings.

(a) The $\mathbf{Z}$-submodule $\mathbf{Z}$ of $\mathbf{Z}[X]$ is not a $\mathbf{Z}[X]$-submodule of $\mathbf{Z}[X]$.

(b) It is easily seen that the set $\{X^n\}_{n \in \mathbf{N}}$ is a basis for $\mathbf{Z}[X]$ viewed as a $\mathbf{Z}$-module. Hence, there is a unique $\mathbf{Z}$-module morphism $f : \mathbf{Z}[X] \to \mathbf{Z}[X]$ such that $f(X^i) = i$ for all i in N. It is easily seen that this is not a morphism of $\mathbf{Z}[X]$-modules. Hence:

(c) The subring $\mathrm{End}_{\mathbf{Z}[X]}(\mathbf{Z}[X])$ of $\mathrm{End}_{\mathbf{Z}}(\mathbf{Z}[X])$ is not all of $\mathrm{End}_{\mathbf{Z}}(\mathbf{Z}[X])$.

Thus, we see that although a ring morphism $f : R \to S$ gives some connections between the category of S-modules and the category of R-modules, these connections are not terribly strong in general. However, these connections are much stronger in the case $f : R \to S$ is a surjective morphism of rings as we now see in the following.

Basic Properties 12.3

Suppose $f : R \to S$ is a surjective morphism of rings.

(a) Suppose X is a subset of the S-module M. Then X is an S-submodule of M if and only if X is an R-submodule of the induced R-module M. In particular:

(b) An S-module M is a simple S-module if and only if the induced R-module M is a simple R-module.

(c) A map $f : M \to M'$ of S-modules is a morphism of S-modules if and only if, viewed as a map of the induced R-modules M and M', f is a morphism of R-modules. Hence:

(d) For each pair of S-modules M_1 and M_2, we have $\mathrm{Hom}_S(M_1, M_2) = \mathrm{Hom}_R(M_1, M_2)$. In particular:

(e) For each S-module M we have $\mathrm{End}_S(M) = \mathrm{End}_R(M)$.

PROOFS: Left as exercises.

It is not true, in general, that a ring morphism $f : R \to S$ carries the center of R into the center of S. However, if $f(C(R)) \subset C(S)$, then associated with the ring morphism $f : R \to S$ is the ring morphism $f' : C(R) \to C(S)$ given by $f'(x) = f(x)$ for all x in $C(R)$. Hence, for each pair of S-modules M_1, M_2 we have that $\mathrm{Hom}_S(M_1, M_2)$ is not only a $C(S)$-module, but also a $C(R)$-module; that is, $\mathrm{Hom}_S(M_1, M_2)$ is the $C(R)$-module induced by the ring morphism $f' : C(R) \to C(S)$.

In this case, it is not difficult to check that $\mathrm{Hom}_S(M_1, M_2)$ is a $C(R)$-submodule of $\mathrm{Hom}_R(M_1, M_2)$.

It is worth noting, however, that if $f: R \to S$ is a surjective morphism of rings, then it is automatically true that $f(C(R)) \subset C(S)$, although, even in this case, $f(C(R))$ need not be all of $C(S)$. Therefore, if $f: R \to S$ is surjective, we have that $\mathrm{Hom}_S(M_1, M_2) = \mathrm{Hom}_R(M_1, M_2)$ has two $C(R)$-module structures. One is given by the usual operation of $C(R)$ on $\mathrm{Hom}_R(M_1, M_2)$ and the other is given by the operation of $C(R)$ on $\mathrm{Hom}_S(M_1, M_2)$ induced by the ring morphism $f': C(R) \to C(S)$. The reader should check that these two operations are really the same.

In order to look more deeply into the connections between the S-modules and R-modules given by a surjective ring morphism $f: R \to S$, we need to introduce a few new concepts.

Suppose M is an R-module and I is a left ideal of R. The subset IM, consisting of all finite sums $\Sigma\, r_i m_i$ with r_i in I and m_i in M, is easily seen to be a submodule of M. Hence, in particular, if I_1 and I_2 are left ideals of R, then $I_1 I_2$ is a left ideal of R contained in I_2. Moreover, $I_1(I_2 M) = (I_1 I_2)M$ for all left ideals I_1 and I_2 in R. Finally, $RM = M$ for all R-modules M.

Next, suppose M is an R-module. Then it is easily seen that the subset $\mathrm{ann}(M)$ of R, consisting of all r in R such that $rm = 0$ for all m in M, is an ideal of R.

Definition
For each R-module M, the ideal $\mathrm{ann}_R(M)$ of R is called the **annihilator** of M. The R-module M is called a **faithful R-module** if $\mathrm{ann}_R(M) = 0$.

Basic Properties 12.4
Let $f: R \to S$ be a ring morphism.

(a) If M is an S-module, then $f^{-1}(\mathrm{ann}_S(M))$ is the annihilator of the R-module M induced by f. In particular:
(b) If M is an S-module, then $\mathrm{Ker}\, f$ is contained in the annihilator of the R-module M.

Suppose M is an R-module and I is an ideal contained in $\mathrm{ann}(M)$. Then M is an R/I-module where the R/I-module structure $R/I \times M \to M$ is given by $(r + I, m) \to rm$ for all r in R. This R/I-module has the property that the R-module structure induced by the canonical ring morphism $k: R \to R/I$ is the R-module structure we started with on M. The R/I-module structure we just defined in M is the only way we consider M an R/I-module. Further, the reader should check that M is a faithful R/I-module if and only if $I = \mathrm{ann}(M)$.

Using these observations, it is not difficult to see that the category of R/I-modules is the full subcategory of R-modules whose annihilators contain I.

Suppose I is an ideal in R. If M is an R-module, then $\mathrm{ann}(M/IM) \supset I$. Hence, M/IM is an R/I-module. It is not hard to show that the canonical R-morphism $k: M \to M/IM$ has the following property. For each R/I-module N, the morphism $\mathrm{Hom}_R(k, N): \mathrm{Hom}_R(M/IM, N) \to \mathrm{Hom}_R(M, N)$ is an isomorphism. Recalling that $\mathrm{Hom}_R(M/IM, N) = \mathrm{Hom}_{R/I}(M/IM, N)$, we have the isomorphism $\mathrm{Hom}_{R/I}(M/IM, N) \to \mathrm{Hom}_R(M, N)$ which we consider an identification.

The following is another important property of the R-morphism $k: M \to M/IM$. Suppose $f: M_1 \to M_2$ is a morphism of R-modules. Because $f(IM_1) \subset IM_2$, there is a unique morphism $\bar{f}: M_1/IM_1 \to M_2/IM_2$ of R-modules, and hence of R/I-modules, such that the diagram

$$\begin{array}{ccc} M_1 & \xrightarrow{f} & M_2 \\ \downarrow{\scriptstyle k_1} & & \downarrow{\scriptstyle k_2} \\ M_1/IM_1 & \xrightarrow{\bar{f}} & M_2/IM_2 \end{array}$$

commutes.

Basic Properties 12.5
Let M_1, M_2, and M_3 be R-modules and let I be an ideal of R.

(a) The map $\mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_{R/I}(M_1/IM_1, M_2/IM_2)$ given by $f \to \bar{f}$ is a group morphism.
(b) If $f: M_1 \to M_2$ and $g: M_2 \to M_3$ are R-morphisms, then $\overline{gf} = \bar{g}\bar{f}$. Hence, if $gf = 0$, then $\bar{g}\bar{f} = 0$.
(c) If $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \to 0$ is an exact sequence of R-modules, then $M_1/IM_1 \xrightarrow{\bar{f}} M_2/IM_2 \xrightarrow{\bar{g}} M_3/IM_3 \to 0$ is an exact sequence of R/I-modules.
(d) If $\{f_\alpha: M_\alpha \to M\}_{\alpha \in A}$ is a sum for the family $\{M_\alpha\}_{\alpha \in A}$ of R-modules, then $\{\bar{f}_\alpha: M_\alpha/IM_\alpha \to M/IM\}_{\alpha \in A}$ is a sum of the family of R/I-modules $\{M_\alpha/IM_\alpha\}_{\alpha \in A}$.

PROOF: (a) through (c) are straightforward.
(d) Use the standard sum.

13. TORSION MODULES OVER PID'S

As an illustration of the ideas introduced in Sections 11 and 12, we develop a useful structure theorem for torsion modules over PID's. Throughout this section R is a PID.

Let x be an element of an R-module M. The **annihilator of x** is the ideal $\mathrm{ann}(x)$ of R consisting of all r in R such that $rx = 0$. It is obvious that $\mathrm{ann}(x)$ is the kernel of the R-morphism $f: R \to M$ given by $f(r) = rx$. Hence, $R/\mathrm{ann}(x) \approx (x)$ where (x) is the submodule of M generated by x. The element x is said to be a **torsion element** if $\mathrm{ann}(x) \neq 0$. Using the fact that R is an integral domain, it is not difficult to show that the subset $t(M)$ of M consisting of all the torsion elements of M is a submodule of M called the **torsion submodule** of M. Finally, M is said to be a **torsion module** if $t(M) = M$.

Suppose M is a torsion R-module. For each prime ideal (p) of R we denote by $M_{(p)}$ the subset of M consisting of all m in M such that $p^n m = 0$ for some n in $\mathbf{N}$. The reader can easily see that $M_{(p)}$ is a submodule of M for each prime ideal (p) of R. We now show that $M = \coprod M_{(p)}$ where (p) ranges over $PPD(R)$.

We do this by showing: (a) For each prime ideal (p) we have $M_{(p)} \cap N_{(p)} = 0$ where $N_{(p)}$ is the submodule generated by all $M_{(q)}$ where (q) ranges over $PPD(R) - \{(p)\}$ and (b) M is generated by the family of submodules $\{M_{(p)}\}_{(p) \in PPD(R)}$ (see Proposition 11.6).

(a) Suppose $x \in M_{(p)} \cap N_{(p)}$. Then $p^n x = 0$ for some n. Because x is in $N_{(p)}$, $x = x_1 + \cdots + x_n$ where each x_i has the property that $q_i^{n_i} x_i = 0$ for some prime element q_i such that $(q_i) \neq (p)$. Hence, $rx = 0$ where $r = \prod_{i=1}^{n} q_i^{n_i}$. Because r and p^n are relatively prime, we have $sr + tp^n = 1$ for some s and t in R. Thus, $x = srx + tp^n x = 0$. Therefore, $M_{(p)} \cap N_{(p)} = 0$ for all (p) in $PPD(R)$.

(b) Let x be in M and $(a) = \operatorname{ann}(x)$. We have already seen that there is an isomorphism $f: R/(a) \to (x)$. Now $R/(a) = \prod_{i=1}^{t} R/(p_i^{n_i})$ where $a = u \prod_{i=1}^{t} p_i^{n_i}$ for some unit u in R (see Chapter 5). Thus, every element of $R/(a)$ may be written as a t-tuple $(\bar{r}_1, \ldots, \bar{r}_t)$ where $\bar{r}_i$ denotes the coset of the element r_i in $R/(p_i^{n_i})$. If we set $y_i = (0, \ldots, \bar{1}, 0, \ldots, 0)$ where $\bar{1}$ is in the ith coordinate, then $p_i^{n_i} y_i = 0 = p_i^{n_i} f(y_i)$. Hence, $f(y_i)$ is in $M_{(p_i)}$. Moreover, it is easy to see that $x = \Sigma f(y_i)$. This proves (b).

Hence, we have almost proven the following.

Theorem 13.1
Let M be a torsion module over a PID R. Then $M = \amalg M_{(p)}$ as (p) ranges over $PPD(R)$. Further, M is a finitely generated R-module if and only if each $M_{(p)}$ is finitely generated and $M_{(p)} = 0$ for all but a finite number of (p) in $PPD(R)$.

PROOF: The first part of the theorem has already been established. The second part is a consequence of the following general observation. Suppose a module $X = \amalg_{\alpha \in A} Y_\alpha$. Then X is finitely generated if and only if $Y_\alpha = 0$ for all but a finite number of α in A and each Y_α is a finitely generated module. The proof of this is left to the reader.

As an application of this theorem we develop the basic facts concerning partial fractions, a subject the reader was undoubtedly introduced to in connection with techniques of integration.

Proposition 13.2
Let K be the field of quotients of the PID R and let $\{p_\alpha\}_{\alpha \in A}$ be a representative family of prime elements of R. If x is an element of K, there exists a finite subset A' of A such that $x = r_0 + \Sigma_{\alpha \in A'} r_\alpha p_\alpha^{-n(\alpha)}$ where:

(a) r_0 and r_α are in R.
(b) r_α is not divisible by p_α for all $\alpha \in A'$.
(c) $n(\alpha) > 0$ for each $\alpha \in A'$.

Moreover, the subset A' and the integers $n(p_\alpha)$ are uniquely determined by these conditions.

PROOF: It is easily seen that $M = K/R$ is a torsion R-module. Hence, by our previous theorem we have $M = \amalg_{\alpha \in A} M_{(p_\alpha)}$. Let $k: K \to K/R$ be the canonical epimorphism. Then $k(x) = \Sigma_{\alpha \in A} m_\alpha$ with each $m_\alpha \in M_{(p_\alpha)}$ where only a finite number of $m_\alpha \neq 0$. Let A' be the subset of A consisting of all $\alpha \in A$ such that $m_\alpha \neq 0$. Then $k(x) = \Sigma_{\alpha \in A'} m_\alpha$ and $m_\alpha \neq 0$ for all α in A'.

If $k(x_\alpha) = m_\alpha$, then $p_\alpha^{m_\alpha} x_\alpha \in R$ for some $m_\alpha > 0$. Let n_α be the smallest integer such that $p_\alpha^{n_\alpha} x_\alpha \in R$. We know that $n_\alpha > 0$ because $m_\alpha \neq 0$. Let $r_\alpha = p_\alpha^{n_\alpha} x_\alpha$. Then $k(x - \Sigma_{\alpha \in A'} r_\alpha p_\alpha^{-n_\alpha}) = 0$ so that $x - \Sigma_{\alpha \in A'} r_\alpha p_\alpha^{-n_\alpha} = r_0$ in R. Thus, we have $x = r_0 + \Sigma_{\alpha \in A'} r_\alpha p_\alpha^{-n_\alpha}$ and conditions (a), (b), and (c) are easily verified.

The proof of the rest of the proposition is left to the reader.

The above expression for x is not the one that is usually used. For example, it is different from the expression used in integrating rational functions. We now describe how to modify it.

First, let p be a prime element of R and let $s: R/(p) \to R$ be a map of sets such that $ks = \mathrm{id}_{R/(p)}$ where $k: R \to R/(p)$ is the canonical ring morphism. If r is an element of R, then for each $n > 0$ there exist unique elements v_i in $R/(p)$ such that $r - \Sigma_{i=0}^{n-1} s(v_i) p^i$ is in (p^n). When $n = 1$, this is clear. Assuming by induction that the statement is true for $n-1$, we prove it for n.

Let $v_0, \ldots, v_{n-2}$ be such that $x - \Sigma_{i=0}^{n-2} s(v_i) p^i = b p^{n-1}$. Let v_n in $R/(p)$ be such that $b - s(v_n) = tp$ for some t in R. Then $x = \Sigma_{i=0}^{n-1} s(v_i) p^i + t p^n$. Hence, $x - \Sigma_{i=0}^{n-1} s(v_i) p^i$ is in (p^n). The uniqueness of the v_i is easy to verify.

We can restate this result as follows. For each element x in R and integer $n > 1$, we have that x can be written uniquely in the form

$$x = \sum_{i=0}^{n-1} r_{i,n} p^i + t_n p^n$$

where $r_{i,n} \in \mathrm{Im}\, s$ and t_n is in R. Multiplying by p^{-n}, we have that $x p^{-n}$ can be written

$$x p^{-n} = t_n + \sum_{k=1}^{n-1} r_{k,n} p^{-k}$$

with t_n in R and $r_{k,n}$ in Im s in one and only one way. Hence, if we define $r_k = r_{k,n}$ for $k \leq n-1$ and $r_k = 0$ for $k \geq n$, we obtain the unique expression

$$x p^{-n} = t_n + \sum_{k=1}^{\infty} r_k p^{-k} \qquad (*)$$

where t_n is in R and where the finite number of r_k which are not zero are in Im s.

Returning to Proposition 13.2 we know that each x in K, where K is the field of quotients of the PID R, can be written uniquely as $x = r_0 + \Sigma_{\alpha \in A'} r_\alpha p_\alpha^{-n_\alpha}$ where A' is a finite subset of A and $n_\alpha > 0$ for all $\alpha \in A'$. Suppose for each α in A we choose a map of sets $s_\alpha: R/(p_\alpha) \to R$ such that $k_\alpha s_\alpha = \mathrm{id}_{R/(p_\alpha)}$. Then applying formula (*) to $r_\alpha p_\alpha^{-n_\alpha}$ we have

$$r_\alpha p_\alpha^{-n_\alpha} = t_{n_\alpha} + \sum_{k=1}^{\infty} r_{k,\alpha} p_\alpha^{-k}$$

where t_{n_α} is in R, all but a finite number of the elements $r_{k,\alpha}$ in R are zero, and $r_{k,\alpha} \in \mathrm{Im}\, s_\alpha$ if $r_{k,\alpha} \neq 0$. Further, this expression is unique. Hence,

$$x = r_0 + \sum_{\alpha \in A'} t_{n_\alpha} + \sum_{\alpha \in A'} \left(\sum_{k=1}^{\infty} r_{k,\alpha} p_\alpha^{-k} \right)$$

where the t_{n_α} are in R, all but a finite number of the $r_{k,\alpha}$ are zero, and $r_{k,\alpha}$ is in Im s_α if $r_{k,\alpha} \neq 0$. Hence, we have our desired version of Proposition 13.2.

Proposition 13.3
Let R be a PID, K its field of quotients, and $\{p_\alpha\}_{\alpha\in A}$ a representative family of prime elements of R. Suppose, further, that for each α in A we have a map $s_\alpha: R/(p_\alpha)\to R$ such that $k_\alpha s_\alpha = \mathrm{id}_{R/(p_\alpha)}$ where $k_\alpha: R\to R/(p_\alpha)$ is the canonical epimorphism.

Then each element x in K can be written uniquely in the form

$$x = r_0 + \sum_{\alpha\in A}\left(\sum_{k=1}^{\infty} r_{k,\alpha} p_\alpha^{-k}\right)$$

where

(a) r_0, $r_{k,\alpha}$ are elements of R.
(b) All but a finite number of $r_{k,\alpha}=0$.
(c) If $r_{k,\alpha}\neq 0$, then $r_{k,\alpha}\in \mathrm{Im}\, s_\alpha$.

We now give two classical applications of this proposition.

Let $R=\mathbf{Z}$, the ring of integers and $K=\mathbf{Q}$, the field of rational numbers. For each positive prime number p define $s_p: \mathbf{Z}/(p)\to\mathbf{Z}$ by $s_p(x)$ is the smallest positive integer in the coset x. Clearly, $k_p s_p = \mathrm{id}_{\mathbf{Z}/(p)}$ where $k_p: \mathbf{Z}\to Z/(p)$ is the canonical epimorphism. Equally obviously, $\mathrm{Im}\, s_p = [0, p-1]$ for each positive prime number p. Proposition 3.3 yields the following in this context.

Proposition 13.4
Each rational number x can be written uniquely in the form

$$x = r_0 + \sum_{p}\left(\sum_{k=1}^{\infty} r_{p,k} p^{-k}\right)$$

where p ranges over all positive prime integers and where:

(a) r_0 and $r_{p,k}$ are in $\mathbf{Z}$.
(b) All but a finite number of $r_{p,k}\neq 0$.
(c) If $r_{p,k}\neq 0$, then $r_{p,k}\in[0, p-1]$.

Another classical application of Proposition 13.3 is the one we alluded to in the beginning of this section concerning rational functions.

Let L be a field. Let R be the PID $L[X]$ and $K=L(X)$, the field of quotients of $L[X]$, which is also called the field of **rational functions** over L. The set A of all monic irreducible elements of $K[X]$ is a representative family of prime elements of $K[X]$. Suppose $p(X)$ is in A. Then it is not very difficult to show, using the Euclidean algorithm, that if degree $p(X)=n$, then each coset in $K[X]/(p(X))$ contains either the element 0 or precisely one monic polynomial of degree less than n. Thus, we obtain a map $s_{p(X)}: K[X]/(p(X))\to K[X]$ where $s_{p(X)}(Y)=0$ or is the unique monic polynomial of degree $<n$ in Y for each coset Y in $K[X]/(p(X))$. Obviously, $k_{p(X)}s_{p(X)} = \mathrm{id}_{k[X]/(p(X))}$. Equally obvious is the fact that Im $s_{p(X)}$ is the set of all monic polynomials of degree $<$ degree $p(X)$ union (0). Thus, we have in this context the following.

Proposition 13.5
Let L be a field, $L[X]$ the polynomial ring over L, and $L(X)$ the field of rational functions over L. Then each rational function $r(X)$ in $L(X)$ can be written uni-

quely in the form

$$r(X) = a(X) + \sum_{p(X)} \left(\sum_{k=1}^{\infty} l_{p(X),k}(X) p(X)^{-k} \right)$$

where $p(X)$ ranges over all monic irreducible polynomials in $L[X]$ and:

(a) $a(X)$ and $l_{p(X),k}(X)$ are in $L[X]$.
(b) All but a finite number of $l_{p(X),k}(X)$ are zero.
(c) If $l_{p(X),k}(X)$ is not zero, then deg $(l_{p(X),k}(X)) <$ deg $p(X)$.

In the particular case $L = \mathbf{R}$, the field of real numbers, it is well known that an irreducible element $p(X)$ of $\mathbf{R}[X]$ has degree at most 2. Hence, in this case, the polynomials $l_{p(X),k}(X)$ entering into the formula are either constants, that is, elements of $\mathbf{R}$, or linear, that is, of the form $X - a$ for some a in $\mathbf{R}$. With this remark in mind, the reader should have no difficulty seeing that when specialized to the case $L = \mathbf{R}$, Proposition 13.5 gives the usual theorem about partial fractions used in calculus to integrate rational functions.

14. PRODUCTS OF MODULES

This section is devoted to studying the elementary properties of products of indexed families of modules over an arbitrary ring R.

Because we have already defined and proved the uniqueness of products in arbitrary categories, provided they exist, there is no need to go into these matters again. However, for ease of reference we recall the following.

Definition
Let $\{M_i\}_{i \in I}$ be an indexed family of R-modules. A family $\{f_i : M \to M_i\}_{i \in I}$ is called **a product of the family** $\{M_i\}_{i \in I}$ if given any family of morphisms $\{g_i : X \to M_i\}$ there is a unique morphism $g : X \to M$ such that $f_i g = g_i$ for each i in I.

Basic Properties 14.1
Let $\{f_i : M \to M_i\}_{i \in I}$ be a family of morphisms.

(a) $\{f_i : M \to M_i\}_{i \in I}$ is a product for the indexed family $\{M_i\}_{i \in I}$ of R-modules if and only if for each R-module X, the map

$$\varphi : \mathrm{Hom}_R(X, M) \to \prod_{i \in I} \mathrm{Hom}_R(X, M_i)$$

given by $\varphi(g) = (f_i g)_{i \in I}$ for all R-morphisms $g : X \to M$ is an isomorphism of abelian groups.

(b) If $\{f_i : M \to M_i\}_{i \in I}$ and $\{f_i' : M' \to M_i\}_{i \in I}$ are both products for the indexed set $\{M_i\}_{i \in I}$, then there is a unique R-morphism $h : M \to M'$ such that $f_i' h = f_i$ for all i in I. This unique R-morphism h is an isomorphism.

It now remains for us to show that each indexed family of R-modules has a product.

Suppose $\{M_i\}_{i \in I}$ is an indexed family of R-modules. Viewing the M_i solely as abelian groups, we saw in Chapter 3 that the abelian group consisting of the set

$\prod_{i\in I} M_i$ together with the addition defined by $\{m_i\}_{i\in I} + \{m'_i\}_{i\in I} = \{m_i + m'_i\}_{i\in I}$ is a product in the category of abelian groups of the family $\{M_i\}_{i\in I}$. It is easily checked that the map $R \times \prod_{i\in I} M_i \to \prod_{i\in I} M_i$ given by $(r, \{m_i\}) \to \{rm_i\}_{i\in I}$ for all r in R and $\{m_i\}_{i\in I}$ in $\prod_{i\in I} M_i$ is an R-module structure in the abelian group $\prod_{i\in I} M_i$. We denote this R-module also by $\prod_{i\in I} M_i$.

It is easy to check that for each j in I, the map $\text{proj}_j : \prod_{i\in I} M_i \to M_j$ given by $\text{proj}_j(\{m_i\}_{i\in I}) = m_j$ is a surjective morphism of R-modules. Finally, it is also easy to show, just as we did for sets and groups, that $\{\text{proj}_j : \prod_{i\in I} M_i \to M_j\}_{j\in I}$ is a product for the indexed family $\{M_i\}_{i\in I}$ of R-modules. We only point out here that if we are given a family $\{g_i : X \to M_i\}_{i\in I}$ of R-morphisms, then the unique morphism $\prod_{i\in I} g_i : X \to \prod_{i\in I} M_i$ such that $f(\prod_{i\in I} g_i) = g_i$ for all i in I is given by $\Pi g_i(x) = \{g_i(x)\}_{i\in I}$. This shows that every indexed family of R-modules has a product in the category of R-modules.

This discussion suggests the following.

Definitions

Let $\{M_i\}_{i\in I}$ be an indexed family of R-modules. The R-module $\prod_{i\in I} M_i$ described above is called the **product** of the indexed family $\{M_i\}_{i\in I}$. For each j in I, the R-morphisms $\text{proj}_j : \prod_{i\in I} M_i \to M_j$ given by $\text{proj}_j(\{m_i\}) = m_j$ is called the jth **projection morphism**. The product $\{\text{proj}_j : \Pi M_i \to M_j\}$ is called the **standard product** of the family $\{M_i\}_{i\in I}$ of R-modules.

Basic Properties 14.2

Let $\{f_i : M \to M_i\}$ be a product for the indexed family $\{M_i\}_{i\in I}$ of R-modules.

(a) The map $h : M \to \prod_{i\in I} M_i$ given by $h(m) = \{\text{proj}_j(m)\}_{i\in I}$ is an R-module isomorphism.

(b) If $t : R' \to R$ is a ring morphism, then $\{f_i : M \to M_i\}_{i\in I}$ is also an R'-product for the family $\{M_i\}_{i\in I}$ of R'-modules.

(c) For each R-module X, we know that $\text{Hom}_R(X, M)$ and $\text{Hom}_R(M, X_i)$ are $C(R)$-modules for each i in I. Hence, $\prod_{i\in I} \text{Hom}_R(C, X_i)$ is also a $C(R)$-module and the map $\varphi : \text{Hom}_R(X, M) \to \prod_{i\in I} \text{Hom}_R(X, M_i)$ given by $\varphi(g) = \{f_i g\}_{i\in I}$ is an isomorphism of $C(R)$-modules.

(d) Suppose K is an ideal in the ring R. Then $\{M_i/KM_i\}_{i\in I}$ is a family of R/K-modules as well as R-modules. Then $\{\bar{f}_i : M/KM \to M_i/KM_i\}_{i\in I}$ is a product of $\{M_i/KM_i\}_{i\in I}$ in the category of R-modules as well as in the category of R/K-modules.

We now show how the sums and products of an indexed family of R-modules are related. We deal mainly with standard sums and products. We leave it to the reader to generalize these results to arbitrary sums and products.

Suppose $\{M_i\}_{i\in I}$ is a family of R-modules. Then by definition, the sum $\coprod_{i\in I} M_i$

consists of all the elements $\{m_i\}_{i\in I}$ in $\prod_{i\in I} M_i$ with the property that $m_i = 0$ for all but a finite number of i in I. Thus, as a set, we have $\coprod_{i\in I} M_i \subset \prod_{i\in I} M_i$. It is easy to see that this subset $\coprod_{i\in I} M_i$ of $\prod_{i\in I} M_i$ is actually a submodule of $\prod_{i\in I} M_i$. Hence, we have that $\coprod_{i\in I} M_i$ is the submodule of $\prod_{i\in I} M_i$ consisting of all elements $\{m_i\}_{i\in I}$ satisfying $m_i = 0$ for all but a finite number of i in I. From this observation it follows that $\coprod_{i\in I} M_i = \prod_{i\in I} M_i$ if I is a finite set. Thus, we have:

Proposition 14.3
Let $\{M_i\}_{i\in I}$ be a family of R-modules. Then:

(a) $\coprod_{i\in I} M_i$ is a submodule of $\prod_{i\in I} M_i$.
(b) If I is a finite set, then $\coprod_{i\in I} M_i = \prod_{i\in I} M_i$.

More generally, suppose $\{f_i : M \to M_i\}_{i\in I}$ and $\{g_i : M_i \to N\}_{i\in I}$ are a product and a sum for the family $\{M_i\}_{i\in I}$ of R-modules. Then in Basic Properties 11.4 we saw that for each k in I there are unique morphisms $p_k : N \to M_k$ such that $p_k g_k = \mathrm{id}_{M_k}$ and $p_k g_i = 0$ for $i \neq k$. Thus, there is a unique morphism $h : N \to M$ such that $f_i h = p_i$ for each i in I. We leave it to the reader to verify that this uniquely determined morphism $h : N \to M$ is always a monomorphism and is an isomorphism if I is a finite set. Also, the reader should show that if $\{f_i : M \to M_i\}_{i\in I}$ and $\{g_i : M_i \to N\}_{i\in I}$ are the standard product and sum of $\{M_i\}_{i\in I}$, then $h : N \to M$ is the inclusion morphism described in Proposition 14.3.

We now conclude this discussion of products by pointing out the following useful criterion for when a module M is isomorphic to a sum and, hence a product, of a finite family $\{M_i\}_{i\in I}$ of R-modules.

Proposition 14.4
Let $\{M_i\}_{i\in I}$ be a finite family of R-modules. A module M is isomorphic to a sum, or equivalently, a product, of the finite family $\{M_i\}_{i\in I}$ if and only if there exist morphisms $p_k : M \to M_k$ and $i_k : M_k \to M$ for each k in I satisfying:

(a) $p_k i_j = 0$ if $j \neq k$ and $p_k i_k = \mathrm{id}_{M_k}$.
(b) $\mathrm{id}_M = \Sigma_{k\in I} i_k p_k$; that is, $m = \Sigma_{k\in I} i_k p_k(m)$ for all m in M.

PROOF: Left as an exercise for the reader.

EXERCISES

(1) Let R be a ring. Let $\{F_i\}_{i\in I}$ be a family of free R-modules and suppose $B_i \subset F_i$ is a basis for F_i. Show that if $\{f_i : F_i \to M\}$ is a sum of the family $\{F_i\}_{i\in I}$, then M is a free R-module with basis $\bigcup_{i\in I} f_i(B_i)$.

(2) Let R be a ring. Suppose $\{M_i\}_{i\in I}$ is a family of R-modules and that for each i in I we are given a set J_i and a family $\{f_{ij} : N_{ij} \to M_i\}_{j\in J_i}$ of morphisms of R-modules which is a sum for the family $\{N_{ij}\}_{j\in J_i}$ of R-modules. Show that if $\{g_i : M_i \to M\}$ is a

sum for the family $\{M_i\}_{i\in I}$ and $L = \coprod_{i\in I} J_i$ [that is, L is the subset of $I\times(\bigcup_{i\in I} J_i)$ such that (i, j) is in L if and only if $j\in J_i$], then the family $\{f_ig_i : N_{ij}\to M\}_{(i,j)\in L}$ is a sum for the family of R-modules $\{N_{ij}\}_{(i,j)\in L}$.

(3) Suppose R is an arbitrary ring. Let $\{f: M_i\to M\}_{i\in I}$ be a sum for the family of R-modules $\{M_i\}_{i\in I}$ where only a finite number of the $M_i = (0)$. Show:

(a) M is a noetherian module if and only if I is finite and each of the M_i are noetherian modules.

(b) Show that M is an artinian module if and only if I is finite and each of the M_i is artinian.

(4) Suppose M is a module over an arbitrary ring R. A finite ascending chain of submodules of M

$$0 = M_0\subset M_1\subset M_2\subset\cdots\subset M_{n-1}\subset M_n = M$$

is said to be a **composition series** for M of length n if M_{i+1}/M_i is a simple R-module for each $i = 0, \ldots, n-1$.

(a) Show that an R-module M has a composition series if and only if it is both a noetherian and artinian module.

(b) If $0\to M'\to M\to M''\to 0$ is an exact sequence of R-modules, show that M has a composition series if and only if M' and M'' have a composition series.

(c) Suppose M is a sum for a finite family $\{M_i\}_{i\in I}$ of R-modules. Show that M has a composition series if and only if each M_i has a composition series.

(d) Show that a ring R has a composition series when viewed as a module over itself if and only if every finitely generated R-module has a composition series.

(e) Let R be a PID and $S = R/\mathfrak{a}$ where $\mathfrak{a}$ is a nonzero ideal.

(i) Show that S has a composition series.

(ii) Show that any S-algebra Λ which is a finitely generated S-module also has a composition series.

(5) Let M be a module over an arbitrary ring R. Let M_1 and M_2 be submodules of M, let $M_1\amalg M_2$ be the standard sum of M_1 and M_2, and let $f_i : M_i\to M$ be the inclusion morphisms. If $f_1\amalg f_2 : M_1\amalg M_2\to M$ is the usual morphism given by $f_1\amalg f_2(m_1, m_2) = m_1 + m_2$, then show:

(a) $\mathrm{Im}(f_1\amalg f_2)$ is the submodule of M generated by M_1 and M_2 and

(b) $\mathrm{Ker}(f_1\amalg f_2)\approx M_1\cap M_2$.

(6) Let $\{m_i\}_{i\in I}$ be a family of elements in an R-module M. For each i in I let A_i be the subset of R consisting of all r in R such that $rm_i = 0$. Show:

(a) Each A_i is a left ideal of R.

(b) Let $\{M_i\}_{i\in I}$ be the family of R-modules with the property that $M_i = M$ for all i in I. Show that the map $f: R\to\prod_{i\in I} M_i$ given by $f(r) = (rm_i)_{i\in I}$ is a morphism of R-modules with the property $\mathrm{Ker}\, f = \bigcap_{i\in I} A_i$. Hence:

(c) $R/\bigcap A_i$ is isomorphic to a submodule of $\prod_{i\in I} M_i$.

(d) Show that if R is commutative, then $\bigcap A_i$ is the annihilator of the submodule M' of M generated by the elements $\{m_i\}_{i\in I}$ of M.

(7) Suppose M is a noetherian module over a commutative ring R. Show that if I is the annihilator of M, then R/I is a noetherian ring.

(8) Let R be a commutative ring and F a finitely generated free R-module. Show:
(a) Every basis of F is a finite set.
(b) Any two bases of F have the same number of elements.
(9) Let R be an arbitrary ring and G the group of automorphisms of the free R-module F. Suppose $\{b_i\}_{i\in I}$ is a basis for F and $\{F_i\}_{i\in I}$ is the family of R-modules such that $F_i = F$ for each i in I. Let $X \subset \prod_{i\in I} F_i$ be the subset consisting of all elements $\{x_i\}$ in $\prod_{i\in I} F_i$ such that $\{x_i\}_{i\in I}$ is a basis for F. Show that $\text{card}(G) = \text{card}(X)$.
(10) Let K be a finite field with q elements. Suppose V is a finite-dimensional K-vector space, say $\dim_K V = n$. Find a formula which expresses $\text{card}(\text{Aut}_K(V))$ in terms of q and n.
(11) Suppose we are given a commutative diagram of R-modules

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & A' & & A & & A'' & & \\
 & & \downarrow{\scriptstyle g'} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle g''} & & \\
0 & \longrightarrow & B' & \xrightarrow{\alpha} & B & \xrightarrow{\beta} & B'' & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle f'} & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle f''} & & \\
0 & \longrightarrow & C' & \xrightarrow{\gamma} & C & \xrightarrow{\delta} & C'' & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle h'} & & \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle h''} & & \\
 & & D' & & D & & D'' & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array} \qquad (*)$$

with exact rows and columns.
(a) Show that there are unique R-morphisms

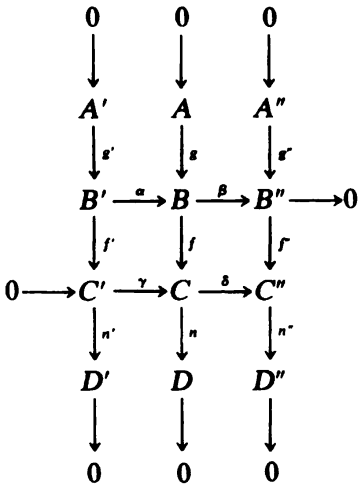
$$\alpha' : A' \to A, \quad \beta' : A \to A'', \quad \gamma' : D' \to D, \quad \delta' : D \to D''$$

such that if they are put in (*), the resulting diagram also commutes.
(b) Show that there is a unique morphism $\epsilon : A'' \to D'$ with the following property: Given b in B and a'' in A'' such that $\beta(b) = g''(a'')$, then an element c' in C' has the property that $b'(c') = \epsilon(a'')$ if and only if $f(b) = \gamma(c')$.
(c) Show that the sequence

$$0 \longrightarrow A' \xrightarrow{\alpha'} A \xrightarrow{\beta'} A'' \xrightarrow{\epsilon} D' \xrightarrow{\gamma'} D \xrightarrow{\delta'} D'' \longrightarrow 0$$

is exact.

(12) Suppose we are given a commutative diagram



(**)

with exact rows and columns. Then prove the following assertions.

(a) There are unique R-morphisms

$$\alpha' : A' \to A, \quad \beta' : A \to A'', \quad \gamma' : D' \to D, \quad \text{and} \quad \delta' : D \to D''$$

such that if they are put in (**), the resulting diagram commutes.

(b) There is a unique morphism $\epsilon : A'' \to D'$ satisfying the same conditions of Exercise 11(b).

(c) The sequence

$$A' \xrightarrow{\alpha'} A \xrightarrow{\beta'} A'' \xrightarrow{\epsilon} D' \xrightarrow{\gamma'} D \xrightarrow{\delta'} D''$$

is exact.

(d) $\alpha' : A' \to A$ is a monomorphism if and only if $\alpha : B' \to B$ is a monomorphism.

(e) $\delta' : D \to D''$ is an epimorphism if and only if $\delta : C \to C''$ is an epimorphism.

(13) An R-module M is said to be **semisimple** if and only if $M \approx \coprod_{i \in I} S_i$ where each S_i is a simple R-module. Suppose M is a semisimple R-module and that $\{f_i : S_i \to M\}_{i \in I}$ is a sum of the family of simple R-modules $\{S_i\}_{i \in I}$. Then show the following:

(a) If N is a proper submodule of M, that is, $N \neq M$, then there is an i in I such that $N \cap \operatorname{Im} f_i = 0$.

(b) Every submodule of M is a summand of M.

(c) M is a module with a composition series if and only if the set I is finite.

(14) Suppose M is an R-module with a composition series. Show that the following are equivalent statements:

(a) M is semisimple.

(b) If N is a proper submodule of M, then there is a nonzero submodule N' of M such that $N \cap N' = 0$.

(c) If N is a submodule of M, then N is a summand of M. Hence, we have:

(d) If M is semisimple, then so is every submodule and factor module of M.

(15) Let R be a ring, S a simple R-module, and $\mathscr{C}_S$ the full subcategory of the category of R-modules consisting of all R-modules M such that there is an integer

n in $\mathbf{N}$ with the property $M \approx \coprod_{i\in[0,n)} S_i$ where each $S_i = S$. We denote the module $\coprod_{i\in[0,n)} S_i$ with each $S_i = S$ by nS. Establish the following:

(a) Each module M in $\mathscr{C}_S$ is a semisimple noetherian and artinian module.

(b) $0\,S = 0$.

(c) $mS \approx nS$ if and only if $m = n$. If M is in $\mathscr{C}_S$, the uniquely determined integer n such that $M \approx nS$ will be denoted by $l_S(M)$ and is called the **S-length** of M.

(d) Suppose $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules with M in $\mathscr{C}_S$. Then:

(i) The exact sequence is splittable.

(ii) M' and M'' are in $\mathscr{C}_S$.

(iii) $l_S(M) = l_S(M') + l_S(M'')$.

(iv) The monomorphism $M' \to M$ is an isomorphism if and only if $l_S(M') = l_S(M)$.

(v) The epimorphism $M \to M''$ is an isomorphism if and only if $l_S(M) = l_S(M'')$.

(vi) If $0 \to M_0 \to M_1 \to M_2 \to \cdots \to M_n \to 0$ is an exact sequence of modules in $\mathscr{C}_S$, then $\Sigma_{i=0}^{n} (-1)^i l_S(M_i) = 0$.

(16) For a ring R we denote by $S(R)$ a set of nonisomorphic simple R-modules such that each simple R-module is isomorphic to one and only one element of $S(R)$. Also, we denote by $\mathscr{D}$ the full subcategory of $\mathrm{Mod}(R)$ consisting of those R-modules which have a composition series.

(a) If $0 \to M_1 \to M_2 \to M_3 \to M_4 \to 0$ is an exact sequence of R-modules, then M_1 and M_4 are in $\mathscr{D}$ if M_2 and M_3 are in $\mathscr{D}$.

(b) If $0 \to M_1 \to M_2 \to M_3 \to 0$ is an exact sequence of R-modules with M_1 and M_3 in $\mathscr{D}$, then M_2 is also in $\mathscr{D}$.

(c) For each M in $\mathscr{D}$ and S in $S(R)$ show there is a unique submodule M_S of M such that:

(i) $M_S \approx nS$ for some n in $\mathbf{N}$.

(ii) If M' is a submodule of M such that $M' \approx mS$ for some m, then $M' \subset M_S$.

(d) If $f: S \to M$ is a morphism of R-modules with S in $S(R)$ and M in $\mathscr{D}$, then $\mathrm{Im}\, f \subset M_S$.

(e) If M is in $\mathscr{D}$, then $M = 0$ if and only if $M_S = 0$ for all S in $S(R)$.

(f) If $f: M' \to M''$ is an R-morphism in $\mathscr{D}$, then $f(M'_S) \subset M''_S$ for each S in $S(R)$.

(g) Show that the following data define a functor $F_S: \mathscr{D} \to \mathscr{C}_S$ for each S in $S(R)$.

(i) $F_S: \mathrm{Ob}\,\mathscr{D} \to \mathrm{Ob}\,\mathscr{C}_S$ is given by $F_S(M) = M_S$.

(ii) If $f: M' \to M''$ is a morphism in $\mathscr{D}$, then $F_S(f): M'_S \to M''_S$ is given by $F_S(f)(m') = f(m')$ for all m' in M'_S.

(h) If $0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M''$ is an exact sequence of modules in $\mathscr{D}$, then

$$0 \longrightarrow F_S(M') \xrightarrow{F_S(f)} F_S(M) \xrightarrow{F_S(g)} F_S(M)$$

is an exact sequence of modules in $\mathscr{C}_S$.

(17) Let R be a ring and let $S(R)$ and $\mathscr{D}$ be as in Exercise 16.

(a) For each module M in $\mathscr{D}$ there is a unique submodule M_1 such that:

(i) M_1 is semisimple.

(ii) If $M' \subset M$ is a semisimple submodule of M, then $M' \subset M_1$.

(b) If A is a semisimple R-module and $f: A \to M$ an R-morphism, then $\mathrm{Im}\, f \subset M_1$.

(c) If $f: M' \to M''$ is a morphism in $\mathcal{D}$, then $f(M_1') \subset M_1''$.

(d) Let $\mathcal{S}$ be the full subcategory of $\mathcal{D}$ whose objects are the semisimple R-modules in $\mathcal{D}$. Show that the following data describe a functor $G_1: \mathcal{D} \to \mathcal{S}$.
  (i) $G_1: \text{Ob } \mathcal{D} \to \text{Ob } \mathcal{S}$ is given by $G(M) = M_1$.
  (ii) If $f: M' \to M''$ is a morphism in $\mathcal{D}$, then $G_1(f): G_1(M') \to G_1(M'')$ is given by $G_1(f)(m') = f(m')$ for all m' in M'.

(e) If M is in $\mathcal{D}$, then $G_1(M) = 0$ if and only if $M = 0$.

(f) Let M be an R-module in $\mathcal{D}$. Define the sequence of submodules

$$M_0 \subset M_1 \subset M_2 \subset \cdots \subset M_n \subset \cdots$$

of M by induction on n. $M_0 = 0$ and M_1 is $G_1(M)$. Suppose we have defined $M_1, \ldots, M_k$. Define M_{k+1} as follows. Let $k: M \to M/M_k$ be the canonical surjective morphism. Then set $M_{k+1} = k^{-1}(G_1(M/M_k))$. This sequence of submodules is known as the **Socle series for** M.
  (i) Show that for some n, we have $M_n = M$. We will denote the smallest value of n such that $M_n = M$ by $L(M)$. Obviously, $L(M) = 0$ if and only if $M = 0$ and $L(M) = 1$ if and only if M is a semisimple nonzero R-module.
  (ii) If $f: M' \to M''$ is a morphism in $\mathcal{D}$, then $f(M_i') \subset M_i''$ for each i in $\mathbf{N}$.
  (iii) For each integer i in $\mathbf{N}$, the following data describe a functor $G_i: \mathcal{D} \to \mathcal{D}_i$ where $\mathcal{D}_i$ is the full subcategory of $\mathcal{D}$ consisting of those M in $\mathcal{D}$ with $L(M) \leq i$.
    (1) $G_i: \text{Ob } \mathcal{D} \to \text{Ob } \mathcal{D}_i$ is given by $G_i(M) = (M_i)$.
    (2) If $f: M' \to M''$ is a morphism in $\mathcal{D}$, then $G_i(f): G_i(M') \to G_i(M'')$ is given by $G_i(f)(m') = f(m')$ for all m' in $G_i(M)$.

(g) If $0 \to M' \to M \to M''$ is an exact sequence of modules in $\mathcal{D}$, then $0 \to G_i(M') \to G_i(M) \to G_i(M'')$ is an exact sequence of R-modules in $\mathcal{D}_i$.

(h) If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules in $\mathcal{D}$, then $L(M') \leq L(M)$ and $L(M'') \leq L(M)$.

(i) If M is an R-module in $\mathcal{D}$, then $G_{i+1}(M)/G_i(M)$ is in $\mathcal{D}_1$ for all i in N and is zero if and only if $i \geq L(M)$. Further, if $f: M \to M'$ is a morphism in $\mathcal{D}$, then there is a unique morphism $F_i(f): G_{i+1}(M)/G_i(M) \to G_{i+1}(M')/G_i(M')$ such that the diagram

$$\begin{array}{ccc} G_{i+1}(M) & \xrightarrow{G_{i+1}(f)} & G_{i+1}(M') \\ \downarrow & & \downarrow \\ G_{i+1}(M)/G_i(M) & \xrightarrow{F_i(f)} & G_{i+1}(M')/G_i(M') \end{array}$$

commutes, where the vertical morphisms are the usual canonical surjective morphisms.

(j) The following data define a functor $F: \mathcal{D} \to \mathcal{D}_1$.
  (i) $F: \text{Ob } \mathcal{D} \to \text{Ob } \mathcal{D}_1$ is given by $F(M) = \coprod_{i \in N} G_{i+1}(M)/G_i(M)$.
  (ii) If $f: M \to M'$ is a morphism in $\mathcal{D}$, then $F(f): F(M) \to F(M')$ is given by $F(f)\{x_i\}_{i \in \mathbf{N}} = \{F_i(f)(x_i)\}_{i \in \mathbf{N}}$ for all $\{x\}_{i \in \mathbf{N}}$ in $\coprod_{i \in \mathbf{N}} G_{i+1}(M)/G_i(M)$.

(18) The notation is the same as in the previous exercises.

(a) For each R-module M in $\mathcal{D}_1$ show that $M = \coprod_{S\in S(R)} M_S$ and $M_S = 0$ for all but a finite number of S in $S(R)$.

(b) Given M and M' in $\mathcal{D}_1$ and morphisms $g_S: M_S \to M'_S$, there is a unique morphism $\coprod_{S\in S(R)} g_S: \coprod_{S\in S(R)} M_S \to \coprod_{S\in S(R)} M'_S$ such that the diagram

$$\begin{array}{ccc} M_{S'} & \xrightarrow{g_{S'}} & M'_{S'} \\ \downarrow & & \downarrow \\ \coprod_{S\in S(R)} M_S & \longrightarrow & \coprod_{S\in S(R)} M'_S \end{array}$$

commutes for all S' in $S(R)$ where the vertical maps are the usual injection morphisms.

(c) Given any morphism $f: M \to M'$ in $\mathcal{D}_1$, then $f = \coprod_{S\in S(R)} f_S$.

(d) Suppose $M \approx \coprod_{S\in S(R)} n_S S$ is in $\mathcal{D}_1$.

(i) All but a finite number of $n_S = 0$.

(ii) $M_S \approx n_S S$ for all S in $S(R)$.

(iii) If $M \approx \coprod_{S\in S(R)} n'_S S$, then $n_S = n'_S$ for all $S \in S(R)$. For each S in $S(R)$, we define $n_S(M)$ to be the uniquely determined integer such that $M \approx \coprod_{S\in S(R)} n_S(M) S$.

(e) If $\{S_i\}_{i\in I}$ and $\{S_j\}_{j\in J}$ are two finite families of simple R-modules, then $\coprod_{i\in I} S_i \approx \coprod_{j\in J} S_j$ if and only if there is a bijective map $\sigma: I \to J$ such that $S_i \approx S_{\sigma(i)}$ for all i in I.

(f) If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of modules in $\mathcal{D}_1$, then $n_S(M) = n_S(M') + n_S(M'')$ for all S in $S(R)$.

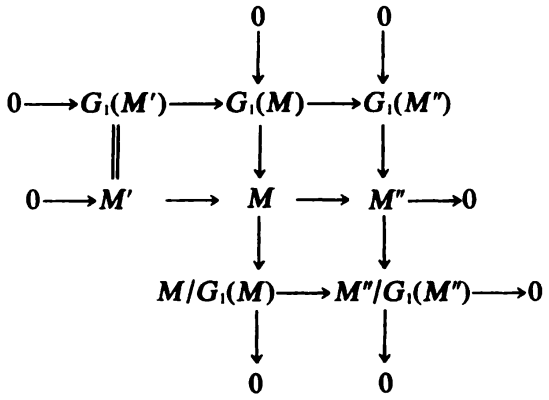
(g) If $0 \to M_0 \to M_1 \to \cdots \to M_t \to 0$ is an exact sequence of modules in $\mathcal{D}_1$, then $\Sigma_{i=0}^{t} (-1)^i n_S(M_i) = 0$ for all S in $S(R)$.

(h) Suppose $0 = M_0 \subset M_1 \subset \cdots \subset M_{n-1} \subset M_n = M$ is a composition series for the R-module M in $\mathcal{D}_1$. Then $\coprod_{i=0}^{n-1} M_{i+1}/M_i \approx \coprod_{S\in S(R)} n_S(M) S$. Hence, if $0 = M_0 \subset M'_1 \subset \cdots \subset M'_{m-1} \subset M'_m = M$ is another composition series, then $m = n = \Sigma_{S\in S(R)} n_S(M)$ and there is a permutation $\sigma: [0, \ldots, m-1] \to [0, \ldots, m-1]$, $S_i \approx T_{\sigma(i)}$ where the $S_i = M_{i+1}/M_i$ and $T_i = M'_{i+1}/M'_i$ for $i = 0, \ldots, n-1$. This common value of the lengths of composition series is called the **length of** M and is usually denoted by $l(M)$. Clearly, $l(M) = \Sigma_{S\in S(R)} n_S(M)$.

(i) If $0 \to M_0 \to M_1 \to \cdots \to M_n \to 0$ is an exact sequence of modules in $\mathcal{D}_1$, then $\Sigma_{i=0}^{n} (-1)^i l(M_i) = 0$. In particular, if $0 \to M_1 \to M_2 \to M_3 \to 0$ is exact, then $l(M_2) = l(M_1) + l(M_2)$.

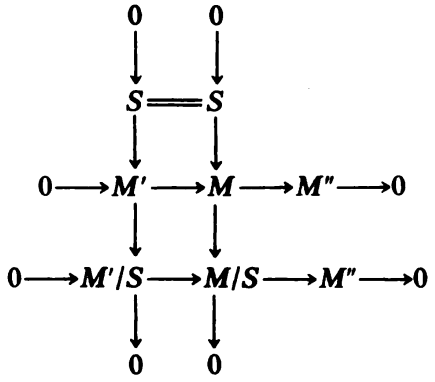
(19) Notation is the same as in Exercise 17(f). We outline a proof that if $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules in $\mathcal{D}$, then $F(M) \approx F(M') \amalg F(M'')$.

(a) First consider the special case that M' is semisimple and proceed by induction on $L(M)$. If $L(M') \leq 1$, we already have the result. Suppose this is true for $L(M) \leq k$ and assume $L(M) = k+1$. Then use the commutative diagram



with exact rows and columns and the fact that $L(M/G_1(M)) = k$ to show that the desired result holds for M.

(b) Proceed to prove the general case by induction on $l(F(M))$. If $l(F(M)) \leq 1$, there is nothing to prove. Suppose the result holds if $l(F(M)) \leq k$ and assume $l(F(M)) = k+1$. Obviously, one can assume that $M' \neq 0$. Let S be a simple submodule of M'. Use the commutative diagram



with exact rows and columns and the fact that $l(F(M/S)) = k$ to prove the desired result for M.

(20) Same notation as above. Let $0 = M_0 \subset M_1 \subset \cdots \subset M_n = M$ be a composition series for M in $\mathcal{D}$. Then show:

(a) $F(M) \approx \coprod_{i=0}^{n-1} M_{i+1}/M_i$.

(b) $n = l(F(M))$.

(c) If $0 = M_0' \subset M_1' \subset \cdots \subset M_m' = M$ is another composition series for M, then $m = n$; and if we let $S_i = M_{i+1}/M_i$ and $S_i' = M_{i+1}'/M_i'$ for $i = 0, \ldots, n-1$, then there is a permutation $\sigma : [0, \ldots, n-1] \to [0, \ldots, n-1]$ such that $S_i \approx S_{\sigma(i)}'$ for all $i = 0, \ldots, n-1$.

(d) Show that $l(F(M))$ is the length of each composition series for M. It is called the **length of** M and is denoted by $l(M)$.

(e) If $0 \to M_0 \to M_1 \to \cdots \to M_n \to 0$ is an exact sequence of modules in $\mathcal{D}$, then $\Sigma_{i=0}^{n} (-1)^i l(M_i) = 0$. In particular, if $0 \to M' \to M \to M'' \to 0$ is exact, then $l(M) = l(M') + l(M'')$.

(21) Same notation as above. Let F be the free R-module generated by $\mathbf{N}$, the set of all nonnegative integers. Let $\mathscr{D}_0$ be the full subcategory of $\mathscr{D}$ consisting of R-modules F/K with K a submodule of F.

(a) Each M in $\mathscr{D}$ is isomorphic to some object F/K in $\mathscr{D}_0$.

(b) $\mathscr{D}_0$ is a small category; that is, the objects in $\mathscr{D}_0$ form a set.

(c) Let $[\mathscr{D}_0]$ be the partition of $\mathscr{D}_0$ given by the equivalence relation $M_1 \mathrm{R} M_2$ if and only if $M_1 \approx M_2$ as R-modules. For each M in $\mathscr{D}_0$, denote by $[M]$ the unique element of $[\mathscr{D}_0]$ containing M. Let A be the free abelian group generated by $[\mathscr{D}_0]$, X the subset of A consisting of all elements of the form $[M]-[M']-[M'']$ if there is an exact sequence of R-modules $0\to M'\to M\to M''\to 0$, and A' the subgroup of A generated by X. Then the abelian group A/A' is called the **Grothendieck group** of $\mathscr{D}$ and is denoted by $\mathfrak{G}(\mathscr{D})$.

  (i) The map of sets $g:\mathscr{D}_0\to\mathfrak{G}(\mathscr{D})$ given by $g(M)=[M]$ for each M in $\mathscr{D}_0$ satisfies $g(M)=g(M')$ if $M\approx M'$ and if $0\to M'\to M\to M''\to 0$ is exact, then $g(M)=g(M')+g(M'')$.

  (ii) Suppose C is an abelian group and $f:\mathscr{D}_0\to C$ is any map of sets satisfying $f(M)=f(M')$ if $M\approx M'$; and if $0\to M'\to M\to M''\to 0$ is an exact sequence of R-modules, then $f(M)=f(M')+f(M'')$. Then there is a unique morphism of abelian groups $h:\mathfrak{G}(\mathscr{D})\to C$ such that $f=hg$.

  (iii) If $h:\mathfrak{G}(\mathscr{D})\to C$ is any morphism of abelian groups, then $hg:\mathscr{D}_0\to C$ has the properties of f cited in part (ii). Hence, for each abelian group C we obtain a map $\psi_C:(\mathfrak{G}(\mathscr{D}), C)\to[\mathscr{D}_0, C]$ given by $\psi_C(h)=hg$ where $(\mathfrak{G}(\mathscr{D}), C)$ is the set of all group morphisms from $\mathfrak{G}(\mathscr{D})$ to C and $[\mathscr{D}_0, C]$ is the set of maps of sets $f:\mathscr{D}_0\to C$ satisfying the conditions of part (ii). Show that ψ_C is an isomorphism of sets for each abelian group C.

(d) Let $S(R)$ be the subset of $\mathscr{D}_0$ consisting of simple R-modules such that each simple R-module is isomorphic to one and only one element of $S(R)$. Let $G(S(R))$ be the free abelian group generated by the set $S(R)$. Define $\alpha:G(S(R))\to\mathfrak{G}(\mathscr{D})$ to be the unique morphism of abelian groups such that $\alpha(S)=[S]$ for each S in $S(R)$. Show that α is a surjective morphism of abelian groups.

(e) For each module M in $\mathscr{D}_0$ there is a unique family $\{n_S(M)\}_{S\in S(R)}$ of elements of M such that $F(M)\approx \coprod_{S\in S(R)} n_S(M)S$. Hence, all but a finite number of the elements $n_S(M)$ are zero. Define $f:\mathscr{D}_0\to G(S(R))$ by $f(M)=\Sigma_{S\in S(R)} n_S(M)S$. Show:

  (i) If $M\approx M'$, then $f(M)=f(M')$.

  (ii) If $0\to M'\to M\to M''\to 0$ is an exact sequence of R-modules in $\mathscr{D}_0$, then $f(M)=f(M')+f(M'')$.

  (iii) $f(S)=S$ for all S in $S(R)$.

(f) The unique group morphism $h:\mathfrak{G}(\mathscr{D}_0)\to G(S(R))$ such that $fg=h$ has the property that the composition

$$G(S(R))\xrightarrow{\alpha}\mathfrak{G}(\mathscr{D})\xrightarrow{h}G(S(R))$$

is the identity of $G(S(R))$.

(g) The morphism $\alpha:G(S(R))\to\mathfrak{G}(\mathscr{D}_0)$ is an isomorphism of abelian groups. Hence:

(h) $\mathfrak{G}(\mathscr{D}_0)$ is a free abelian group with basis $\{[S]\}_{S\in S(R)}$.

(i) If M is in $\mathcal{D}_0$, then $[M]=\Sigma_{S\in S(R)} n_S[S]$ in $\mathcal{G}(\mathcal{D}_0)$ if and only if given any composition series $0=M_0\subset M_1\subset\cdots\subset M_t=M$ for M we have $\coprod_{i=0}^{t-1} M_{i+1}\approx \coprod_{S\in S(R)} n_S S$. In particular, $l(M)=\Sigma_{S\in S(R)} n_S$.

(22) Let $0\to \mathbf{Z}/p\mathbf{Z}\xrightarrow{f} G\xrightarrow{g}\mathbf{Z}/pa\mathbf{Z}\to 0$ be an exact sequence of abelian groups where p is a prime integer and a is any integer. Show that if g is not a splittable epimorphism, then $G\approx \mathbf{Z}/p^2a\mathbf{Z}$ and so G is a cyclic group. [*Hint*: Consider the canonical epimorphism $k:\mathbf{Z}/p^2a\mathbf{Z}\to\mathbf{Z}/pa\mathbf{Z}$.]

(a) Prove that $\operatorname{Ker} k=pa\mathbf{Z}/p^2a\mathbf{Z}$ which is isomorphic to $\mathbf{Z}/p\mathbf{Z}$ by means of the morphism $j:\mathbf{Z}/p\mathbf{Z}\to pa\mathbf{Z}/p^2a\mathbf{Z}$ given by $j(z+p\mathbf{Z})=paz+p^2a\mathbf{Z}$.

(b) Prove that there is a morphism $h:\mathbf{Z}/p^2a\mathbf{Z}\to G$ such that $gh=k$.

(c) Show that there is a commutative diagram of abelian groups

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathbf{Z}/p\mathbf{Z} & \xrightarrow{j} & \mathbf{Z}/p^2a\mathbf{Z} & \xrightarrow{k} & \mathbf{Z}/pa\mathbf{Z} & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle h'} & & \downarrow{\scriptstyle h} & & \| & & \\ 0 & \longrightarrow & \mathbf{Z}/p\mathbf{Z} & \xrightarrow{f} & G & \xrightarrow{g} & \mathbf{Z}/pa\mathbf{Z} & \longrightarrow & 0 \end{array}$$

and that $h'=0$ if and only if the epimorphism g is splittable.

(d) Show that if g is not splittable, then the morphism $h:\mathbf{Z}/p^2a\mathbf{Z}\to G$ is an isomorphism.

(23) Let I be a left ideal in a ring R and let C be the center of R. Consider each R-module M a C-module by means of the inclusion morphism $C\to R$.

(a) If M is an R-module, show that M^I, the subset of M consisting of all m in M such that $Im=0$, is a C-submodule of M.

(b) Show that if $f:R/I\to M$ is an R-morphism, then $f(1+I)$ is in M^I.

(c) Show that the map $\phi_M:\operatorname{Hom}_R(R/I, M)\to M^I$ given by $f\to f(1)$ for all f in $\operatorname{Hom}_R(R/I, M)$ is an isomorphism of C-modules.

(d) Show that if $f:M_1\to M_2$ is a morphism of R-modules, then $f(M_1^I)\subset M_2^I$ and that the map $f^I:M_1^I\to M_2^I$ defined by $f^I(m)=f(m)$ for all m in M_1^I is a C-module morphism.

(e) Show that the following data define a functor $F:\operatorname{Mod}(R)\to\operatorname{Mod}(C)$.

(i) $F:\operatorname{Ob}\operatorname{Mod}(R)\to\operatorname{Ob}\operatorname{Mod}(C)$ is given by $F(M)=M^I$.

(ii) For each pair of R-modules M_1 and M_2 the map $F:\operatorname{Hom}_R(M_1, M_2)\to \operatorname{Hom}_C(F(M_1), F(M_2))$ is defined by $F(f)=f^I$.

(f) Show that the functor $F:\operatorname{Mod}(R)\to\operatorname{Mod}(C)$ also has the property that $F:\operatorname{Hom}_R(M_1, M_2)\to\operatorname{Hom}_C(F(M_1), F(M_2))$ is a morphism of C-modules.

(g) Show that the functor F is isomorphic to the functor $\operatorname{Hom}_R(R/I, \cdot):\operatorname{Mod}(R)\to \operatorname{Mod}(C)$.

(h) Show that the functor F is **left exact**; that is, if $0\to M_1\to M_2\to M_3$ is an exact sequence of R-modules, then $0\to F(M_1)\to F(M_2)\to F(M_3)$ is an exact sequence of C-modules.

(i) Suppose $R=\mathbf{Z}$, the ring of integers, $I=n\mathbf{Z}$, and M is the abelian group $\mathbf{Q}/\mathbf{Z}$. Show that $(\mathbf{Q}/\mathbf{Z})^I$, and hence $\operatorname{Hom}_{\mathbf{Z}}(\mathbf{Z}/n\mathbf{Z}, \mathbf{Q}/\mathbf{Z})$, is isomorphic to $\mathbf{Z}/n\mathbf{Z}$.

(24) Let R and S be rings and let $F:\operatorname{Mod}(R)\to\operatorname{Mod}(S)$ be a functor. F is said to preserve arbitrary sums if whenever the indexed family of R-morphisms $\{M_i\xrightarrow{f_i} M\}_{i\in I}$ is a sum for $\{M_i\}_{i\in I}$, the indexed family of S-morphisms

$\{F(M_i) \xrightarrow{F(f_i)} F(M)\}_{i\in I}$ is a sum for the family of S-modules $\{F(M_i)\}_{i\in I}$. F is said to be an **additive functor** if it preserves finite sums; that is, if I is a finite set and the family of R-morphisms $\{M_i \xrightarrow{f_i} M\}_{i\in I}$ is a sum for $\{M_i\}_{i\in I}$, then the family of S-morphisms $\{F(M_i) \xrightarrow{F(f_i)} F(M)\}_{i\in I}$ is a sum for the family of S-modules $\{F(M_i)\}_{i\in I}$.

(a) Show that if the functor F has the property that for each pair M_1 and M_2 of R-modules the map $F: \mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_S(F(M_1), F(M_2))$ is a morphism of abelian groups, then F is an additive functor.

(b) Show that if F is an additive functor then for each pair M_1 and M_2 of R-modules, the map $F: \mathrm{Hom}_R(M_1, M_2) \to \mathrm{Hom}_S(F(M_1), F(M_2))$ is a morphism of abelian groups. [*Hint*: Use the following description of the addition in $\mathrm{Hom}_R(M_1, M_2)$. For each R-module define the R-morphism $\Delta_M: M \to M \amalg M$ by $\Delta(m) = (m, m)$ and the R-morphism $+_M: M \amalg M \to M$ by $+_M(m_1, m_2) = m_1 + m_2$. Suppose $f_1, f_2: M_1 \to M_2$ are R-morphisms. Show that the composition $M_1 \xrightarrow{\Delta_{M_1}} M_1 \amalg M_1 \xrightarrow{f_1 \amalg f_2} M_1 \amalg M_2 \xrightarrow{+_{M_2}} M_2$ is the R-morphism $f_1 + f_2: M_1 \to M_2$.]

(c) Let R be a ring with center C. Suppose M is an R-module. Show that the functor $\mathrm{Hom}_R(M, \cdot): \mathrm{Mod}(R) \to \mathrm{Mod}(C)$ described by the following data is an additive functor.

(i) $\mathrm{Hom}_R(M, \cdot)(X)$ for each R-module X is the C-module $\mathrm{Hom}_R(M, X)$.

(ii) For each R-morphism $f: M_1 \to M_2$ define $\mathrm{Hom}_R(M, \cdot)(f): \mathrm{Hom}_R(M, M_1) \to \mathrm{Hom}_R(M, M_2)$ to be the C-morphism $\mathrm{Hom}_R(M, f)$.

(d) Show that if the R-module M is a finitely generated R-module, then the functor $\mathrm{Hom}_R(M, \cdot): \mathrm{Mod}(R) \to \mathrm{Mod}(S)$ preserves arbitrary sums. Is this necessarily true if M is not finitely generated?

(25) Suppose $f_1: M_1 \to M$ and $f_2: M_2 \to M$ are morphisms of R-modules. Show that the subset $M_1 \underset{M}{\times} M_2$ of $M_1 \amalg M_2$ consisting of all (m_1, m_2) in $M_1 \amalg M_2$ such that $f_1(m_1) = f_2(m_2)$ has the following properties:

(a) $M_1 \underset{M}{\times} M_2$ is a submodule of $M_1 \amalg M_2$.

(b) The maps $p_1: M_1 \underset{M}{\times} M_2 \to M_1$ and $p_2: M_1 \underset{M}{\times} M_2 \to M_2$ given by $p_1(m_1, m_2) = m_1$ and $p_2(m_1, m_2) = m_2$ for all (m_1, m_2) in $M_1 \underset{M}{\times} M_2$ are R-module morphisms with the property that the diagram

$$\begin{array}{ccc} M_1 \underset{M}{\times} M_2 & \xrightarrow{p_1} & M_1 \\ {\scriptstyle p_2}\downarrow & & \downarrow{\scriptstyle f_1} \\ M_2 & \xrightarrow[f_2]{} & M \end{array}$$

commutes.

(c) Suppose $g: X \to M_1 \underset{M}{\times} M_2$ is a morphism of R-modules. Then the R-morphisms $p_1 g: X \to M_1$ and $p_2 g: X \to M_2$ associated with g have the property $f_1 p_1 g = f_2 p_2 g$.

(d) If X is any R-module and $h_1: X \to M_1$ and $h_2: X \to M_2$ are any two R-morphisms such that $f_1 h_1 = f_2 h_2$, then there is one and only one R-morphism $h: X \to M_1 \underset{M}{\times} M_2$ such that $h_1 = p_1 h$ and $h_2 = p_2 h$.

(e) For each R-module X, define $F(X)$ to be the set of all pairs (h_1, h_2) of R-morphisms $h_1: X \to M_1$ and $h_2: X \to M_2$ which satisfy $f_1 h_1: X \to M$ and $f_2 h_2: X \to M$ are equal. Then $(h_1, h_2) + (h_1', h_2') = (h_1 + h_1', h_2 + h_2')$ is a law of composition on $F(X)$ which makes $F(X)$ an abelian group.

(f) If $h: X \to Y$ is a morphism of R-modules and (h_1, h_2) is in $F(Y)$, then $(h_1 h, h_2 h)$ is in $F(X)$. Moreover, the map $F(h): F(Y) \to F(X)$ given by $F(h) \times (h_1, h_2) = (hh_1, hh_2)$ is a morphism of abelian groups. Finally, there is a contravariant functor $F: \mathrm{Mod}(R) \to \mathscr{Ab}$ with the property $F: \mathrm{Ob}\,\mathrm{Mod}(R) \to \mathrm{Ob}\,\mathscr{Ab}$ is given by $X \mapsto F(X)$ and where $F: \mathrm{Hom}_R(X, Y) \to \mathrm{Hom}_Z(F(Y), F(X))$ is given by $h \mapsto F(h)$ for all h in $\mathrm{Hom}_R(X, Y)$.

(g) The map $\mathrm{Hom}_R(X, M_1 \underset{M}{\times} M_2) \to F(X)$ given by $g \mapsto (p_1 g, p_2 g)$ for each X in $\mathrm{Mod}(R)$ is an isomorphism of functors $\mathrm{Hom}_R(\cdot, M_1 \times M_2) \to F$.

(h) Suppose we are given an R-module N together with R-morphisms $g_1: N \to M_1$ and $g_2: N \to M_2$ satisfying:

(i) $f_2 g_2 = f_1 g_1$ and

(ii) given any commutative diagram of R-morphisms

$$\begin{array}{ccc} X & \xrightarrow{h_1} & M_1 \\ {\scriptstyle h_2}\downarrow & & \downarrow{\scriptstyle f_1} \\ M_2 & \xrightarrow[f_2]{} & M \end{array}$$

there is a unique R-morphism $t: X \to N$ such that $h_1 t = g_1$ and $h_2 t = g_2$. In particular, there is a unique R-morphism $u: N \to M_1 \underset{M}{\times} M_2$ such that $p_1 u = g_1$ and $p_2 u = g_2$, and this uniquely determined R-morphism u is an isomorphism.

Summarizing, we have that given any diagram of R-morphisms

$$\begin{array}{ccc} & & M_1 \\ & & \downarrow{\scriptstyle f_1} \\ M_2 & \xrightarrow{f_2} & M \end{array}$$

there exists a commutative diagram of R-morphisms

$$\begin{array}{ccc} N & \xrightarrow{g_1} & M_1 \\ {\scriptstyle g_2}\downarrow & & \downarrow{\scriptstyle f_1} \\ M_2 & \xrightarrow[f_2]{} & M \end{array}$$

with the property that given any commutative diagram of R-morphisms

$$\begin{array}{ccc} X & \xrightarrow{h_1} & M_1 \\ {\scriptstyle h_2}\downarrow & & \downarrow{\scriptstyle f_1} \\ M_2 & \xrightarrow[f_2]{} & M \end{array}$$

there exists a unique R-morphism $t: X \to N$ such that $g_1 t = h_1$ and $g_2 t = h_2$. Any

such triple (N, g_1, g_2) is called a **pull-back** for the diagram

$$\begin{array}{ccc} & & M_1 \\ & & \downarrow f_1 \\ M_2 & \xrightarrow{f_2} & M \end{array}$$

The triple $(M_1 \underset{M}{\times} M_2, p_1, p_2)$ is called the **standard pull-back** of the diagram

$$\begin{array}{ccc} & & M_1 \\ & & \downarrow f_1 \\ M_2 & \xrightarrow[f_2]{} & M \end{array}$$

The next exercise is devoted to developing some of the basic properties of pull-backs.

(26) Consider the commutative diagram of R-modules with exact rows and columns

$$\begin{array}{ccccccc} & & 0 & & 0 & & \\ & & \downarrow & & \downarrow & & \\ & & \operatorname{Ker} p_2 & & \operatorname{Ker} f_1 & & \\ & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow \operatorname{Ker} p_1 \longrightarrow & M_2 \underset{M}{\times} M_1 & \xrightarrow{p_1} & M_1 & & \\ & & \downarrow{\scriptstyle p_2} & & \downarrow{\scriptstyle f_1} & & \\ 0 & \longrightarrow \operatorname{Ker} f_2 \longrightarrow & M_2 & \xrightarrow{f_2} & M & & \end{array}$$

(a) Show that there are unique R-morphisms $g_1 : \operatorname{Ker} p_1 \to \operatorname{Ker} f_2$ and $g_2 : \operatorname{Ker} p_2 \to \operatorname{Ker} f_1$ which make the diagram commute.

(b) Show that these uniquely determined R-morphisms g_1 and g_2 are isomorphisms of R-modules.

(c) Show that the morphisms $f_2 p_2 : M_2 \underset{M}{\times} M_1 \to M$, $f_1 p_1 : M_2 \underset{M}{\times} M_1 \to M$ are the same and that:

(i) $\operatorname{Ker}(f_1 p_1) \cong \operatorname{Ker} f_1 \amalg \operatorname{Ker} f_2$.

(ii) $\operatorname{Im}(f_1 p_1) = \operatorname{Im} f_1 \cap \operatorname{Im} f_2$.

(d) Show that $\operatorname{Im} p_1 = M_1$ if and only if $\operatorname{Im} f_1 \subset \operatorname{Im} f_2$. Similarly, $\operatorname{Im} p_2 = M_2$ if and only if $\operatorname{Im} f_2 \subset \operatorname{Im} f_1$.

(e) Show that $p_1 : M_1 \underset{M}{\times} M_2 \to M_2$ is a splittable epimorphism if and only if there is a morphism $h : M_1 \to M_2$ such that $f_2 h_1 = f_1$.

(27) Suppose $f_1 : M \to M_1$ and $f_2 : M \to M_2$ are morphisms of R-modules. Show that the subset N of $M_1 \amalg M_2$ consisting of all pairs $(f_1(m), -f_2(m))$ for all m in M is a submodule of $M_1 \amalg M_2$ and that the R-module $(M_1 \amalg M_2)/N$, which we denote by $M_1 \overset{M}{\times} M_2$, has the following properties:

(a) If $\iota_1 : M_1 \to M_1 \overset{M}{\times} M_2$ and $\iota_2 : M_2 \to M_1 \overset{M}{\times} M_2$ are the compositions $M_1 \xrightarrow{\text{inj}_1} M_1 \amalg M_2 \xrightarrow{k} (M_1 \amalg M_2)/N$ and $M_2 \xrightarrow{\text{inj}_2} M_1 \amalg M_2 \xrightarrow{k} (M_1 \amalg M_2)/N$, respectively, then $\iota_1 f_1 = \iota_2 f_2$.

(b) If $g: M_1 \overset{M}{\times} M_2 \to X$ is a morphism of R-modules, then $g\iota_1 f_1 = g\iota_2 f_2$.

(c) If X is an R-module and $g_1: M_1 \to X$ and $g_2: M_2 \to X$ are two R-morphisms such that $g_1 f_1 = g_2 f_2$, then there is one and only one R-morphism $h: M_1 \overset{M}{\times} M_2 \to X$ such that $h\iota_1 = g_1$ and $h\iota_2 = g_2$.

(d) For each R-module X define $G(X)$ to be the set of all pairs (g_1, g_2) of R-morphisms $g_1: M_1 \to X$ and $g_2: M_2 \to X$ such that $g_1 f_1 = g_2 f_2$. Then $(g_1, g_2) + (g_1', g_2') = (g_1 + g_1', g_2 + g_2')$ is a law of composition on $G(X)$ which makes $G(X)$ an abelian group.

(e) If $t: X \to Y$ is a morphism of R-modules and (g_1, g_2) is in $G(X)$, then (tg_1, tg_2) is in $G(Y)$. Moreover, the map $G(t): G(X) \to G(Y)$ given by $G(t)(h_1, h_2) = (th_1, th_2)$ is a morphism of abelian groups. Finally, there is a functor $G: \mathrm{Mod}(R) \to \mathscr{Ab}$ with the property that $G: \mathrm{Ob}\,\mathrm{Mod}(R) \to \mathrm{Ob}\,\mathscr{Ab}$ is given by $X \mapsto G(X)$ and where $G: \mathrm{Hom}_R(X, Y) \to \mathrm{Hom}_Z(G(X), G(Y))$ is given by $t \mapsto G(t)$ for all t in $\mathrm{Hom}_R(X, Y)$.

(f) The maps $\mathrm{Hom}_R(M_1 \overset{M}{\times} M_2, X) \to G(X)$ given by $h \mapsto (h\iota_1, h\iota_2)$ for each X in $\mathrm{Mod}(R)$ is an isomorphism of functors $\mathrm{Hom}_R(M_1 \overset{M}{\times} M_2, \cdot) \to G$.

(g) Suppose we are given an R-module L together with R-morphisms $g_1: M_1 \to L$, $g_2: M_2 \to L$ satisfying:

(i) $g_1 f_1 = g_2 f_2$.

(ii) Given any commutative diagram of R-morphisms

$$\begin{array}{ccc} M & \xrightarrow{f_1} & M_1 \\ \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle h_1} \\ M_2 & \xrightarrow[h_2]{} & X \end{array}$$

there is a unique R-morphism $h: L \to X$ such that $hg_1 = h_1$ and $hg_2 = h_2$. In particular, there is a unique R-morphism $v: M_1 \overset{M}{\times} M_2 \to L$ such that $v\iota_1 = g_1$ and $v\iota_2 = g_2$ and this uniquely determined R-morphism v is an isomorphism.

Summarizing, we have that given any diagram of R-morphisms

$$\begin{array}{ccc} M & \xrightarrow{f_1} & M_1 \\ \downarrow{\scriptstyle f_2} & & \\ M_2 & & \end{array}$$

there exists a commutative diagram of R-morphisms

$$\begin{array}{ccc} M & \xrightarrow{f_1} & M_1 \\ \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle g_1} \\ M_2 & \xrightarrow[g_2]{} & L \end{array}$$

with the property that given any commutative diagram of R-morphisms

$$\begin{array}{ccc} M & \xrightarrow{f_1} & M_1 \\ \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle h_1} \\ M_2 & \xrightarrow[h_2]{} & X \end{array}$$

there is a unique R-morphism $h: L \to X$ such that $hg_1 = h_1$ and $hg_2 = h$. Any such triple (L, g_1, g_2) is called a **push-out** for the diagram

$$\begin{array}{ccc} M & \xrightarrow{f_1} & M_1 \\ \downarrow{\scriptstyle h} & & \\ M & & \end{array}$$

The triple $(M_1 \overset{M}{\times} M_2, \iota_1, \iota_2)$ is called the **standard push-out** of the diagram

$$\begin{array}{ccc} M & \xrightarrow{f_1} & M_2 \\ \downarrow{\scriptstyle h} & & \\ M_2 & & \end{array}$$

The next exercise is devoted to giving some of the basic properties of push-outs.

(28) Consider the commutative diagram

$$\begin{array}{ccccccc} M & \xrightarrow{f_1} & M_1 & \longrightarrow & \operatorname{Coker} f_1 & \longrightarrow & 0 \\ \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle \iota_1} & & & & \\ M_2 & \xrightarrow[\iota_2]{} & M_1 \overset{M}{\times} M_2 & \longrightarrow & \operatorname{Coker} \iota_2 & \longrightarrow & 0 \\ \downarrow & & \downarrow & & & & \\ \operatorname{Coker} f_2 & & \operatorname{Coker} \iota_1 & & & & \\ \downarrow & & \downarrow & & & & \\ 0 & & 0 & & & & \end{array}$$

of R-modules with exact rows and columns.

(a) Show that there are unique R-morphisms $h_1: \operatorname{Coker} f_1 \to \operatorname{Coker} \iota_2$ and $h_2: \operatorname{Coker} f_2 \to \operatorname{Coker} \iota_1$ which make this diagram commute.

(b) Show that these uniquely determined morphisms are isomorphisms.

(c) Show that $\iota_1 f_1 = \iota_2 f_2$ and that:

(i) $\operatorname{Coker}(\iota_1 f_1) \approx \operatorname{Coker} f_1 \amalg \operatorname{Coker} f_2$.

(ii) $\operatorname{Ker}(\iota_1 f_1)$ is the submodule of M generated by $\operatorname{Ker} f_1$ and $\operatorname{Ker} f_2$.

(d) Show that $\operatorname{Ker} \iota_2 = f_2(\operatorname{Ker} f_1)$ from which it follows that ι_2 is a monomorphism if f_1 is a monomorphism.

(e) Show that ι_2 is a splittable monomorphism if and only if there is a morphism $g: M_1 \to M_2$ such that $gf_1 = f_2$.

(29) Let R be an arbitrary ring.

(a) Show that if $\{M_i\}_{i \in I}$ is a family of R-modules, then $\operatorname{ann}(\prod_{i \in I} M_i) = \bigcap_{i \in I} \operatorname{ann}(M_i)$.

(b) Let M be an R-module and suppose $\{m_i\}_{i \in I}$ generates M as a module over the center of R. Let $\{M_i\}_{i \in I}$ be the family of R-modules with each $M_i = M$. Show that the element $(m_i)_{i \in I}$ in $\prod_{i \in I} M_i$ has the property that $\operatorname{ann}((m_i)_{i \in I}) = \operatorname{ann}(M)$ from which it follows that there is an exact sequence of R-modules $0 \to R/\operatorname{ann}(M) \to \prod_{i \in I} M_i$.

(c) Let I be an ideal of R. Let $\operatorname{Mod}(R)_I$ be the full subcategory of $\operatorname{Mod}(R)$ whose

objects are those R-modules M such that $IM = 0$. Show that $\mathrm{Mod}(R)_I$ has the following properties:

(i) If $0 \to M_1 \to M_2 \to M_3 \to 0$ is an exact sequence of R-modules with M_2 in $\mathrm{Mod}(R)_I$, then M_1 and M_3 are also in $\mathrm{Mod}(R)_I$.

(ii) If $\{M_i\}_{i \in I}$ is a family of R-modules in $\mathrm{Mod}(R)_I$, then $\prod_{i \in I} M_i$ is also in $\mathrm{Mod}(R)_I$.

(iii) The category $\mathrm{Mod}(R)_I$ is equivalent to the category $\mathrm{Mod}(R/I)$.

(d) Let $\mathscr{C}$ be a full subcategory of $\mathrm{Mod}(R)$ which satisfies:

(i) If $0 \to M_1 \to M_2 \to M_3 \to 0$ is an exact sequence of R-modules with M_2 in $\mathscr{C}$, then M_1 and M_3 are also in $\mathscr{C}$.

(ii) If $\{M_i\}_{i \in I}$ is a family of R-modules in $\mathscr{C}$, then $\prod_{i \in I} M_i$ is also in $\mathscr{C}$.

Show that there is an ideal I of R such that $\mathscr{C} = \mathrm{Mod}(R)_I$. [*Hint*: Let $I = \bigcap_{M \in \mathscr{C}} \mathrm{ann}(M)$. Show first that $\mathscr{C} \subset \mathrm{Mod}(R)_I$. Next show that $\mathscr{C} \supset \mathrm{Mod}(R)_I$ by showing

(1) R/I is in $\mathscr{C}$ and

(2) all sums $\coprod_{i \in I} M_i$ with $M_i = R/I$ for all i in I are in $\mathscr{C}$.]

(30) Let R be a commutative ring and X a monoid. This exercise is devoted to giving a description of the $R[X]$-modules.

(a) By an **X-module structure** on an R-module M we mean a law of composition $X \times M \to M$ which we denote by $(x, m) \to xm$ which satisfies:

(i) $1m = m$ for all m in M.

(ii) $x_1 x_2(m) = x_1(x_2 m)$ for all x_1, x_2 in X and m in M.

(iii) $x(rm) = r(xm)$ for all x in X, r in R, and m in M.

(b) An R-module M together with an X-module structure on M is called an X-module. If M_1, M_2 are two X-modules, then a map $f: M_1 \to M_2$ is called a **morphism of X-modules** if and only if f is a morphism of R-modules which also satisfies $f(xm) = xf(m)$ for all x in X and m in M. Clearly, if $f: M_1 \to M_2$ and $g: M_2 \to M_3$ are morphisms of X-modules, then the ordinary composition of maps $gf: M_1 \to M_3$ is also a morphism of X-modules. Show that there is a category $\mathscr{C}$ whose objects M are the X-modules, where $\mathscr{C}(M_1, M_2)$ is the set of all X-module morphisms from M_1 to M_2 and where the composition maps $\mathscr{C}(M_1, M_2) \times \mathscr{C}(M_2, M_3) \to \mathscr{C}(M_1, M_3)$ are given by the usual composition of maps.

(c) Show that if M is an X-module, then the map $R[X] \times M \to M$ given by $(\Sigma_{x \in X} r_x x, m) \to \Sigma_{x \in X} r_x(xm)$ is an $R[X]$-module structure on the abelian group M. We denote that $R[X]$-module consisting of underlying abelian groups of the X-module M together with the $R[X]$-module structure we just defined by $F(M)$.

(d) Show that if M_1 and M_2 are two X-modules, then a map $f: M_1 \to M_2$ is a morphism of X-modules if and only if it is a morphism from the $R[X]$-modules $F(M_1)$ to $F(M_2)$ [remember as sets, $F(M) = M$ for all X-modules M]. Hence, $\mathscr{C}(M_1, M_2) = \mathrm{Hom}_{R[X]}(F(M_1), F(M_2))$.

(e) Show that the following data define a fully faithful functor $F: \mathscr{C} \to \mathrm{Mod}(R[X])$:

(i) $F: \mathrm{Ob}\, \mathscr{C} \to \mathrm{Ob}\, \mathrm{Mod}(R[X])$ is given by $M \to F(M)$ for all X-modules M.

(ii) $F: \mathscr{C}(M_1, M_2) \to \mathrm{Hom}_{R[X]}(M_1, M_2)$ is the identity map for all M_1, M_2 in $\mathscr{C}$.

(f) Show that the functor $F: \mathscr{C} \to \mathrm{Mod}(R[X])$ is an isomorphism of categories by showing that given an $R[X]$-module N, there is one and only one X-module M such that $F(M) = N$.

(31) Let $\{M_i\}_{i \in I}$ be a family of submodules of an R-module M.

(a) Define the R-morphism $f: \coprod_{i \in I} M_i \to M$ by $f((m_i)_{i \in I}) = \Sigma_{i \in I} m_i$. Show that $\mathrm{Im}\, f$ is the submodule of M generated by the family $\{M_i\}_{i \in I}$ of submodules of M.

(b) Suppose $I = \{1, 2\}$. Show that the sequence $0 \to M_1 \cap M_2 \xrightarrow{g} M_1 \amalg M_2 \xrightarrow{f} M$ is exact where $g: M_1 \cap M_2 \to M_1 \amalg M_2$ is given by $g(m) = (m, -m)$ for all m in $M_1 \cap M_2$.

(32) Let R be a commutative ring with the property that each prime ideal of R is finitely generated. We outline a proof that R is noetherian, or, what is the same thing, every ideal of R is finitely generated. We assume that not every ideal of R is finitely generated and get a contradiction.

(a) Show that the ordered set $\mathscr{T}$ of all ideals of R which are not finitely generated is an inductive set and thus has a maximal element I.

(b) Because I is not R or a prime ideal, there are elements a, b in R but not in I such that ab is in I. Hence, the ideals (I, a) and (I, b) generated by I and a and I and b respectively are finitely generated R-modules. Hence, there is an exact sequence $0 \to I \cap (b) \to I \amalg (b) \to (I, b) \to 0$ of R-modules where (b) is the ideal generated by b.

(c) Show that the subset $I:b$ of R consisting of all r in R such that $rb \in I$ is an ideal of R containing I properly and is thus a finitely generated R-module.

(d) Show that the R-morphism $g: I:b \to R$ given by $g(x) = bx$ for all x in $I:b$ has the property $\mathrm{Im}\, g = I \cap (b)$ which shows that $I \cap (b)$ is a finitely generated R-module.

(e) Use the fact that both $I \cap (b)$ and (I, b) are finitely generated R-modules to deduce that $I \amalg (b)$ and hence I is finitely generated.

(33) Let I be an ideal in an integral domain R. If K is the field of quotients of R, let I' be the subset of K consisting of all q in K such that qa is in R for all a in I.

(a) Show that I' is a submodule of the R-module K.

(b) Associated with each q in I' is the map $\psi(q): I \to R$ given by $\psi(q)(a) = qa$. Show:

(i) For each q in I', the map $\psi(q): I \to R$ is a morphism of R-modules.

(ii) The map $\psi: I' \mapsto \mathrm{Hom}_R(I, R)$ given by $q \mapsto \psi(q)$ is an isomorphism of R-modules.

(34) Let R be a commutative ring. A sequence of elements $a_1, a_2, \ldots, a_n$ in R is called a **regular R-sequence** if a_1 is regular in R and for each $i = 2, \ldots, n$, the image of a_i under the canonical R-morphism $R \to R/(a_1, \ldots, a_{i-1})$ is a regular element in $R/(a_1, \ldots, a_{i-1})$. Prove that if R is an integral domain and I is an ideal of R which contains a regular R-sequence with at least two elements, then $I' = R$ where I' is the submodule of K, the field of quotients of R, defined in Exercise 33.

(35) Let R be a PID which is not a field and K its field of quotients which we view as an R-module. Let $k: K \to K/R$ be the canonical surjective morphism of R-modules. For each x in K, we denote $k(x)$ by $[x]$.

(a) Show that K/R is a torsion R-module.

(b) Show that $(K/R)_{(p)} \neq 0$ for each prime ideal (p) in $PPD(R)$.

(c) Let (p) be a fixed prime ideal in $PPD(R)$ and let M be the submodule of K/R generated by all the elements of the form $[(1/p)n]$ with n a positive integer.
  (i) Show that $M = (K/R)_{(p)}$.
  (ii) Show that M is not a finitely generated R-module by showing that ann $(M) = 0$ while ann $(M') \neq 0$ for each finitely generated submodule M' of M.

(d) Let M be as in part (c). For each integer n, let M_n be the submodule of M generated by $[(1/p)n]$. Show:
  (i) $M_1 \subset M_2 \subset M_3 \subset \cdots \subset M_n \subset \cdots$.
  (ii) $M_i \neq M_{i+1}$ for all $i = 1, 2, \ldots$.
  (iii) M is the union of all the M_n.
  (iv) M_n is isomorphic to R/p^nR for each positive integer n.

(e) Show that every nonzero element of M can be written as $[m/p^n]$ where p does not divide m.

(f) Show that if p does not divide m, then M_n is the submodule of M generated by $[m/p^n]$. [*Hint*: Proceed by induction on n.]

(g) Show that the M_n are the only nonzero proper submodules of M. Hence, M has the property that every proper submodule of M is generated by a single element even though M is not finitely generated.

(h) Show that M is an artinian module. Thus, M is an example of an artinian module which is not finitely generated and hence not noetherian.

(i) Give an example of a noetherian module which is not artinian.

(36) Let $\mathbf{C}$ be the field of complex numbers. For each z in $\mathbf{C}$ we denote by $\bar{z}$ the complex conjugate of z, that is, if $z = x + iy$, then $\bar{z} = x - iy$.

(a) Show that the set $R = \mathbf{C} \times \mathbf{C}$ together with the following addition and multiplication is a ring:

$$(z_1, z_2) + (u_1, u_2) = (z_1 + u_1, z_2 + u_2)$$
$$(z_1, z_2)(u_1, u_2) = (z_1 u_1, z_1 u_2 + z_2 \bar{u}_1)$$

(b) Describe the center $C(R)$ of R.

(c) Show that the subset I of R consisting of all elements of the form $(0, u)$ is an ideal of R and that the ring R/I is isomorphic to the field $\mathbf{C}$ of complex numbers.

(d) Show that under the canonical ring morphism $R \to R/I$, the image of $C(R)$ is not all of R/I and hence not all of $C(R/I)$.

(37) Let F be a free R-module, R an arbitrary ring, I an ideal of R, and $k : F \to F/IF$ the canonical R-morphism.

(a) Show that if B is a basis for F, then $k(B)$ is a basis for the R/I-module F/IF and the map $B \to k(B)$ given by $b \mapsto k(b)$ is an isomorphism of sets.

(b) Suppose R/I is left noetherian. Show that if F is finitely generated, then any two bases of F have the same number of elements.

(38) Show that if R is an arbitrary commutative ring, then any two bases for a finitely generated free R-module have the same cardinality.

(39) Show that if D is a division ring, then any two bases for the same D-module have the same cardinality, whether the D-module is finitely generated or not.

(40) Show that if R is a commutative ring, then any two bases for a free R-module have the same cardinality, whether the free R-module is finitely generated or not.

Chapter 7 SEMI-SIMPLE RINGS AND MODULES

In Chapter 6 we saw that if every left R-module is free, then R is a division ring. However, when R is a division ring, we know that R is not only a free module but also a simple R-module. Thus, in the case where R is a division ring, to say that every module is free is the same as to say that every module is the sum of copies of a fixed simple free module, namely R. We can therefore ask the following questions:

(1) What rings have the property that every module is the sum of a fixed simple module?

More generally:

(2) What rings have the property that every module is the sum of simple modules (not necessarily a fixed one)?

This chapter is devoted to answering these questions, and naturally we begin with the first one.

As a matter of notational convenience, we shall denote $\mathrm{Hom}_R(X, Y)$ by (X, Y) for all R-modules X and Y.

1. SIMPLE RINGS

Suppose that R is a ring, M_0 is a simple R-module, and every R-module is a sum of copies of M_0. Then this should be almost as good as being a division ring. We

know, by Chapter 6, Corollary 8.5, that $\mathrm{End}_R(M_0)$, the endomorphism ring of M_0, is a division ring. We denote $\mathrm{End}_R(M_0)$ by Λ_0. Can R be related to Λ_0 in some way? If M_0 were R itself, then Λ_0 would be equal to R. In fact, the morphism $\alpha_R : (R, R) \to R$ given by $\alpha_R(f) = f(1)$ is not only an isomorphism of R-modules as shown in Chapter 6, Proposition 4.6 but is an isomorphism of rings, as the reader can readily verify. But what happens if $M_0 \neq R$? Certainly each element $r \in R$ induces an endomorphism of M_0 as an abelian group, namely the group morphism which sends an element x in M_0 to rx. Because R is not necessarily commutative, this endomorphism of M_0 is not necessarily an endomorphism of R-modules. However, Λ_0 is a ring and M_0 is a Λ_0-module by defining $\lambda \cdot x = \lambda(x)$ for $\lambda \in \Lambda_0$ and $x \in M$. The reader should check that this operation does really define M_0 as a Λ-module. We then observe that the group endomorphism of M_0 induced by an element $r \in R$ is an endomorphism of M_0 as a Λ_0-module, because $r(\lambda x) = r\lambda(x) = \lambda(rx)$ for all $r \in R$, $\lambda \in \Lambda_0$, and $x \in M_0$. In fact, we have the following.

Proposition 1.1
Let R be a ring and M any R-module. If Λ denotes the endomorphism ring of M as an R-module, then M is a Λ-module under the operation $\lambda \cdot x = \lambda(x)$ for $\lambda \in \Lambda$, $x \in M$. Moreover, if Ω denotes the endomorphism ring of M now considered as a Λ-module, there is a canonical ring morphism γ of R into Ω which to each element $r \in R$ assigns the endomorphism of M defined by $r(x) = rx$. If $M = R$, then $\Omega = \Lambda = R$ and γ is the identity.

PROOF: The only part of the above statement not already discussed is that which says the map $\gamma : R \to \Omega$ is a ring morphism and that if $M = R$, then γ is the identity. However, this too may be easily checked by the reader.

As a result of Proposition 1.1 we know that we have a canonical ring morphism $\gamma : R \to \Omega_0$ where Ω_0 is the ring of endomorphisms of M_0 as a Λ_0-module, M_0 is a vector space over Λ_0, and Ω_0 is the ring of endomorphisms of the vector space M_0 over the division ring Λ_0. We now ask the following two questions:

(a) Is $\gamma : R \to \Omega_0$ an isomorphism of rings?
(b) Is M_0 a finite-dimensional vector space over Λ_0?

If the answers to (a) and (b) are yes, then this tells us that the ring R is the endomorphism ring of a finite-dimensional vector space over a division ring.

Let us now show that the answers to (a) and (b) are indeed affirmative. Of course, we must somehow use the fact that every R-module is the sum of copies of M_0. In particular, R itself is such a sum and we can prove the following.

Lemma 1.2
Let R be a ring, $\{M_\alpha\}_{\alpha \in A}$ an indexed family of R-modules, and $\{R, i_\alpha : M_\alpha \to R\}$ a sum of the M_α. Then $M_\alpha = 0$ for all but a finite number of $\alpha \in A$.

PROOF: By Chapter 6, Basic Properties 11.5 we know that for each $r \in R$ we have $r = \Sigma i_\alpha p_\alpha(r)$ where the morphisms $p_\alpha : R \to M_\alpha$ are the projection morphisms of this sum. Because p_α is an epimorphism, we may prove that $M_\alpha = 0$ for a certain α by showing that p_α is the zero morphism. Again by Chapter 6, Basic Property

11.5 we know that for each $r \in R$, $p_\alpha(r) \neq 0$ for finitely many α. In particular, let $\alpha_1, \ldots, \alpha_n$ be such that $p_{\alpha_j}(1) \neq 0$. We claim that $M_\alpha = 0$ for $\alpha \neq \alpha_1, \ldots, \alpha_n$. To do this, we must show that $p_\alpha(r) = 0$ for all $r \in R$ if $\alpha \neq \alpha_1, \ldots, \alpha_n$. However, because $\Sigma_{j=1}^n i_{\alpha_j} p_{\alpha_j}(1) = 1$, we have $r = r \cdot 1 = r\Sigma i_{\alpha_j} p_{\alpha_j}(1) = \Sigma i_{\alpha_j} p_{\alpha_j}(r)$. Because this expression for r is unique, we must have $p_\alpha(r) = 0$ for $\alpha \neq \alpha_1, \ldots, \alpha_n$ and we are done.

Corollary 1.3
If R is a sum of copies of a module M, then R is the sum of a finite number of copies of M.

Proposition 1.4
Let R be a ring, M an R-module, and Λ the endomorphism ring of M. If R is the sum of a finite number, n, of copies of M, then M is isomorphic as a Λ-module to the sum of n copies of Λ. Thus, M is a finitely generated free Λ-module.

PROOF: Denote the sum of n copies of M by M^n. Because $M \approx (R, M)$ as an R-module we also have $M \approx (R, M)$ as a Λ-module, where the operation of Λ on (R, M) is defined by $(\lambda \cdot f)(r) = \lambda(f(r))$, that is, the operation of Λ on (R, M) comes through the operation of Λ on M. Using the fact that $R \approx M^n$ as an R-module, we have $(R, M) \approx (M^n, M)$ as Λ modules where again the operation of Λ on (M^n, M) is defined by $(\lambda \cdot f)(x) = \lambda(f(x))$. But $(M^n, M) \approx (M, M)^n$ and this too is an isomorphism of Λ-modules, where (M, M) is regarded as a Λ-module by setting $(\lambda \cdot f)(x) = \lambda(f(x))$. However, $(M, M) = \Lambda$ by definition, and Λ is considered a Λ-module by multiplication, that is, $\lambda \cdot \mu = \lambda\mu$, where $\lambda\mu$ is the endomorphism defined by $\lambda\mu(x) = \lambda(\mu(x))$. Thus, Λ is considered a Λ-module in the same way as (M, M) is defined to be a Λ-module above. Hence, the chain of isomorphisms $M \approx (R, M) \approx (M^n, M) \approx (M, M)^n = \Lambda^n$ is a chain of isomorphisms as Λ-modules and we therefore have $M \approx \Lambda^n$, as asserted.

If the reader would like a more explicit description of this isomorphism, consider $\{R, i_\alpha : M \to R\}$ a given sum of n copies of $M (\alpha = 1, \ldots, n)$. Then $1 = \Sigma_{\alpha=1}^n i_\alpha(m_\alpha)$ with $m_\alpha \in M$. Taking Λ^n to be the set of all n-tuples $(\lambda_1, \ldots, \lambda_n)$, the reader may check that the isomorphism $h : \Lambda^n \to M$ is given by $h(\lambda_1, \ldots, \lambda_n) = \Sigma_{\alpha=1}^n \lambda_\alpha(m_\alpha)$.

Letting $M = M_0$ and $\Lambda = \Lambda_0$ in Proposition 1.4, we see that M_0 is a finite-dimensional vector space over Λ_0. This answers question (b). To answer question (a), consider the following:

Proposition 1.5
Let E be an R-module, Λ' its endomorphism ring as an R-module, and Ω' its endomorphism ring as a Λ'-module, with $\gamma' : R \to \Omega'$ the canonical ring morphism. Let M be an R-module such that E is the sum of n copies of M and let Λ and Ω be the corresponding endomorphism rings of M with $\gamma : R \to \Omega$ the canonical ring morphism. If γ' is injective, then so is γ. If γ' is surjective, then so is γ.

PROOF: For convenience we shall assume that E is the standard sum of n copies of M, that is, E is the module of n-tuples of elements of M. An R-

endomorphism of E is then an n^2-tuple $\{\lambda_{\alpha\beta}\}$ of R-endomorphisms of M because $(E, E) = (M^n, M^n) = \Pi(M, M)$; this latter product involving n^2 factors. Explicitly, if $\{\lambda_{\alpha\beta}\}$ is an endomorphism of E, the element $(x_1, \ldots, x_n)$ of E is sent to $(\Sigma_{j=1}^n \lambda_{1j}(x_j), \Sigma_{j=1}^n \lambda_{2j}(x_j), \ldots, \Sigma_{j=1}^n \lambda_{nj}(x_j))$. Now suppose that $\omega : M \to M$ is a Λ-endomorphism of M. Define the map $\omega^n : E \to E$ by $\omega^n(x_1, \ldots, x_n) = (\omega(x_1), \ldots, \omega(x_n))$. The reader can easily check that this is an endomorphism of E as an abelian group. We claim that this is also an endomorphism of E as a Λ'-module. To see this, we must prove that if $\lambda' \in \Lambda'$, then $\omega^n(\lambda'(x)) = \lambda'(\omega^n(x))$ for all $x \in E$. However, we have already said that $\lambda' = \{\lambda_{\alpha\beta}\}$ so that, if $x = (x_1, \ldots, x_n)$, we have $\omega^n(\lambda'(x)) = \omega^n(\Sigma \lambda_{1j}(x_j), \Sigma \lambda_{2j}(x_j), \ldots, \Sigma \lambda_{nj}(x_j)) = (\omega(\Sigma \lambda_{1j}(x_j)), \ldots, \omega(\Sigma \lambda_{nj}(x_j))), \ldots, \Sigma \omega(\lambda_{nj}(x_j))$. Because $\omega \in \Omega$ and $\lambda_{\alpha\beta} \in \Lambda$, we have $\omega(\lambda_{ij}(x_j)) = \lambda_{ij}(\omega(x_j))$ so $(\Sigma \omega(\lambda_{1j}(x_j)), \ldots, \Sigma \omega(\lambda_{nj}(x_j))) = (\Sigma \lambda_{1j}\omega(x_j), \ldots, \Sigma \lambda_{nj}\omega(x_j))$.

When we compute $\lambda'(\omega^n(x))$, we get $\lambda'(\omega^n(x)) = \lambda'(\omega(x_1), \ldots, \omega(x_n)) = (\Sigma \lambda_{1j}(\omega(x_j)), \ldots, \Sigma \lambda_{nj}(\omega(x_j)))$. Hence, we see that $\omega^n(\lambda'(x)) = \lambda'\omega^n(x))$ and ω^n is indeed an element of Ω'.

Suppose that $\gamma' : R \to \Omega'$ is surjective. We want to show that $\gamma : R \to \Omega$ is surjective. Thus, given $\omega \in \Omega$, we must show that $\omega(x) = rx$ for some $r \in R$ and all $x \in M$. Because $\omega^n : E \to E$ is an element of Ω', and because we are assuming that γ' is surjective, there is an element $r \in R$ such that $\omega^n(x) = rx$ for all $x \in E$. This means that for every n-tuple $(x_1, \ldots, x_n)$ of elements of M we have $\omega^n(x_1, \ldots, x_n) = (\omega(x_1), \ldots, \omega(x_n)) = (rx_1, \ldots, rx_n)$ so that, in particular, we have $\omega(x) = rx$ for all $x \in M$ and γ is surjective.

Now suppose that $\gamma : R \to \Omega'$ is injective. This is just a fancy way of saying that if $r \in R$ and $r(x_1, \ldots, x_n) = 0$ for all $(x_1, \ldots, x_n)$ in E, then $r = 0$. From this we want to deduce that if $r \in R$ is such that $rx = 0$ for all $x \in M$, then $r = 0$. But, if $rx = 0$ for all $x \in M$, then $0 = (rx_1, \ldots, rx_n) = r(x_1, \ldots, x_n)$ for all $(x_1, \ldots, x_n) \in E$ and so r must be 0. Thus, we have shown that if γ' is injective, so is γ.

From this proposition, using the fact that if $E = R$ we have $\gamma' : R \to \Omega'$ is the identity, we obtain the following.

Corollary 1.6
Let R be a ring, M an R-module, Λ the endomorphism ring of the R-module M, and Ω the endomorphism ring of the Λ-module M. If R is the sum of copies of M, then the canonical ring morphism $\gamma : R \to \Omega$ is an isomorphism.

This proposition shows that $\gamma : R \to \Omega_0$ is an isomorphism of rings, which answers question (**a**) in the affirmative. Hence, we see that if a ring R has the property that every R-module or, more particularly, R itself, is the sum of copies of a fixed simple R-module, then R is isomorphic to the endomorphism ring of a finite-dimensional vector space over a division ring.

Of course we must now ask if this is the best we can do. Namely, the above result is true, but we have really only used the fact that R is the sum of copies of a simple module. But our original problem was to describe rings all of whose modules are a sum of a fixed simple module. The following theorem shows that we have already solved this problem.

Theorem 1.7
Let R be a ring. Then the following statements are equivalent:

(a) R is isomorphic to the endomorphism ring of a finite-dimensional vector space over a division ring.
(b) R is the sum of copies of a fixed simple R-module.
(c) Every R-module is the sum of copies of a fixed simple R-module.

PROOF: The proof of this theorem is lengthy, and involves proving various lemmas and propositions in a slightly more general setting than we seem to need. This is because we want ultimately to study rings which have the property that every module is the sum of simple modules, without specifying that all these simple modules be isomorphic to one fixed one.

We begin by showing that (a) implies (b). Let D be a division ring, V a finite-dimensional (say n-dimensional) vector space over D, and let R be the endomorphism ring of V. We shall show that R is the sum of copies of a fixed simple R-module. From our discussion we have the clue that V must be the simple R-module we are looking for. Hence, let us first show that V is really a simple R-module, and that R is the sum of n copies of V.

To show that V is a simple R-module, it suffices to show that if $v \in V$ and $v \neq 0$, then the R-submodule generated by v is all of V. In other words, we must show that if v' is any element of V, then there is an endomorphism f of V such that $v' = f(v)$. But this is obvious. Because $v \neq 0$, we know that v may be extended to a basis for V over D. Then an endomorphism f of V is determined by assigning arbitrary values in V to the elements of this basis. Assigning the value v' to v and, say, zero to all other basis elements, we get an endomorphism $f: V \to V$ such that $f(v) = v'$. This proves that V is simple as an R-module.

We now want to show that R is a sum of n copies of V. Consider R as the ring $M_n(D)$ of $n \times n$ matrices over D by choosing a basis $x_1, \ldots, x_n$ of V. Under this identification of R with $M_n(D)$, the R-module V becomes the $M_n(D)$-module given by the operation $(d_{ij})(\Sigma_{i=1}^n c_i x_i) = \Sigma_{i=1}^n c_i' x_i$ where $c_i' = \Sigma_{j=1}^n c_j d_{ji}$. We have already seen in Chapter 6, Example 11.10 that $M_n(D) = \coprod_{k \in [1,n]} C_k$ where C_k is the left ideal of $M_n(D)$ generated by the matrix e_{k1} all of whose entries are zero except for the entry in the first column and kth row which is 1. We also saw that $C_k \approx C_l$ for all $k = 1, \ldots, n$ and $l = 1, \ldots, n$. Therefore, if we show that $C_1 \approx V$ as $M_n(D)$-modules, we will have shown that $M_n(D)$ and hence R is isomorphic to the sum of n copies of the fixed simple module V. That $C_1 \approx V$ follows from the fact that the map $f: C_1 \to V$ given by $f(d_{ij}) = \Sigma_{j=1}^n d_{ij} x_j$ for all (d_{ij}) in C_1 is an $M_n(D)$-isomorphism. This finishes the proof that (a) implies (b).

We have already shown that (c) implies (a). Thus, we must show that (b) implies (c) to finish the proof of Theorem 1.7. Here is where we start to set things up a bit more generally.

2. SEMISIMPLE MODULES

Definition
An R-module M is said to be a **semisimple module** if it is the sum of simple modules. A ring R is said to be **semisimple** if it is semisimple as an R-module.

Basic Properties 2.1

(a) A simple module is semisimple.
(b) A sum of semisimple modules is semisimple.
(c) If a ring R is semisimple, then every free R-module is semisimple.
(d) If a ring R is semisimple, then every R-module contains a simple module.

PROOF: (a) and (b) are completely trivial [see exercises of Chapter 6 for (b)].

(c) is an immediate consequence of (b), because every free R-module is a sum of copies of R.

(d) is also quite easy. For, if M is a nonzero R-module, we have $0 \neq M = (R, M) = (\coprod_{i \in I} M_i, M)$, where M_i are simple modules (we actually know that there are only finitely many m_i by Lemma 1.2). Because $(\coprod_{i \in I} M_i, M) \approx \prod_{i \in I} (M_i, M)$, and because $(\coprod_{i \in I} M_i, M) \neq 0$, we have $\prod_{i \in I} (M_i, M) \neq 0$. But then (M_i, M) must be different from zero for at least one index i, and so for some i we have $f: M_i \to M$ and $f \neq 0$. Because M_i is a simple module, we know that f is injective. Hence, the image of f, being isomorphic to M_i, is a simple submodule of M.

Example 2.2 Every vector space over a division ring is semisimple.

Example 2.3 Every division ring and every endomorphism ring of a finite-dimensional vector space over a division ring is semisimple.

In fact, these rings, in addition to being semisimple, are also examples of what are called simple rings. Since we have mentioned the term, we shall define it here but we will not discuss simple rings until later.

Definition
A ring R is called **simple** if it has no ideals other than (0) and (1).

Although the rings we have mentioned so far are both semisimple and simple, in general, it is not the case that a simple ring is semisimple. Examples will be given later to illustrate this fact.

Now let us return to semisimple modules.

Proposition 2.4
An R-module M is semisimple if and only if every submodule M' of M is a summand of M.

PROOF: Recall that a submodule M' of M is a summand if and only if there is a submodule M'' of M such that $\{M, i': M' \to M, i'': M'' \to M\}$ is a sum of M' and M'' where i' and i'' are the inclusion morphisms. As we saw in Chapter 6,

Proposition 11.6, this means that $M' \cap M'' = 0$ and $M' + M''$ (the submodule generated by M' and M'') is equal to all of M. Equivalently, M' is a summand of M if there is a morphism $p : M \to M'$ such that $pi = \mathrm{id}_{M'}$ where $i : M' \to M$ is the inclusion morphism. Ker p is then the module M'' of the preceding description of summands.

Suppose that M is a semisimple module, M' is a submodule of M, and that $M = \amalg M_i$, where M_i are simple modules which we may assume to be submodules of M. Consider the set $\mathcal{N}$ of pairs (N, f_N) where N is a submodule of M containing M' and $f_N : N \to M'$ is a morphism such that $f_N i_N = \mathrm{id}_{M'}$ where $i_N : M' \to N$ is the inclusion. We order the set $\mathcal{N}$ by setting $(N_1, f_{N_1}) \leq (N_2, f_{N_2})$ if N_1 is a submodule of N_2 and $f_{N_2}|N_1 = f_{N_1}$. This clearly defines an order relation on $\mathcal{N}$. $\mathcal{N}$ is not empty because $(M', \mathrm{id}_{M'})$ is in $\mathcal{N}$. If $\{(N_\alpha, f_{N_\alpha})\}$ is a totally ordered subset of $\mathcal{N}$, let $N = \cup N_\alpha$ and define $f_N : N \to M'$ by $f_N(x) = f_{N_\alpha}(x)$ if $x \in N_\alpha$. Clearly, N is a submodule of M containing M', $f_N : N \to M'$ is a morphism, and $f_N i_N = \mathrm{id}_{M'}$. Equally clearly, the pair (N, f_N) is the l.u.b. of the set $\{(N_\alpha, f_{N_\alpha})\}$. Thus, we may apply Zorn's lemma and conclude that $\mathcal{N}$ contains a maximal element (N, f_N). We shall show that $N = M$ and this will then tell us that M' is a summand of M because it is a summand of N. Suppose that $N \neq M$. Then, for some i, the submodule M_i is not contained in N. For, if $M_i \subset N$ for all i, then $\amalg M_i \subset N$ and thus $M = N$ contrary to our assumption. If M_i is not contained in N, then $M_i \cap N = 0$ because the only other alternative is that $M_i \cap N = M_i$ which implies that $M_i \subset N$ (do not forget that M_i is a simple module). Because $M_i \cap N = 0$, the submodule $N + M_i$ generated by N and M_i in M is a sum of N and M_i so that we may define a morphism $f : N + M_i \to M'$ by sending N to M' by f_N and sending M_i to zero in M'. The inclusion $i : M' \to N + M_i$ is the composition $M' \to N \to N + M_i$ so that it is clear that $fi = \mathrm{id}_{M'}$ and $f|N = f_N$. Therefore, the pair $(N + M_i, f)$ is in $\mathcal{N}$ and is properly larger than (N, f_N). This contradiction of the maximality of (N, f_N) is $\mathcal{N}$ shows that $N = M$ and thus M' is a summand of M.

Now let us suppose that every submodule of M is a summand of M, and show that M is semisimple. First, we show that every nonzero submodule M' of M contains a simple submodule. Let $x \in M'$ with $x \neq 0$, and let $\mathcal{S}$ be the set of submodules of M' not containing x. $\mathcal{S}$ is not empty, because it contains the zero module. Order the set $\mathcal{S}$ by inclusion and suppose $\{N_\alpha\}$ is a totally ordered subset of $\mathcal{S}$. Then $\cup N_\alpha$ is also in $\mathcal{S}$ (why?) so that, by Zorn's lemma, $\mathcal{S}$ contains a maximal element, say N. Now N is a summand of M'. In fact, if M'' is any submodule of M', then M'' is a summand of M'. For, if M'' is a submodule of M', it is also a submodule of M and hence, by our assumption on M, M'' is a summand of M. This means we have a morphism $p : M \to M''$ such that $pi'' = \mathrm{id}_{M''}$ where $i'' : M'' \to M$ is the inclusion. If $j : M'' \to M'$ is the inclusion of M'' in M' and $i' : M' \to M$ is the inclusion of M' in M, we have $i'' = i'j$. Then $pi'' = pi'j$ so that $pi' : M' \to M''$ is a morphism such that $(pi')j = \mathrm{id}_{M''}$ and thus M'' is also a summand of M'.

In particular, then, our module N is a summand of M' and thus we may find a submodule N' of M' such that $N \cap N' = 0$ and $N + N' = M'$, that is, M' is generated by N and N'. We claim that N' is a simple module. For, if N' contains a nontrivial submodule N'', we know that N'' is a summand of N' (same argument as above for M'), and hence there is a submodule N''' of N' such that $N''' \cap N'' = 0$ and $N''' + N'' = N'$. Thus, $M' = N + N' = N + N''' + N''$. Now consider our element

$x \in M'$ especially for which our module N was found. We claim that either $x \notin N + N''$ or $x \notin N + N'''$. For, suppose $x \in N + N''$ and $x \in N + N'''$. Then $x = n_1 + n''$ and $x = n_2 + n'''$ where $n_1, n_2 \in N$, $n'' \in N''$, and $n''' \in N'''$. Because $n_1 + n'' = n_2 + n'''$, we have $n_1 - n_2 = n''' - n''$. However, $n_1 - n_2 \in N$ and $n''' - n'' \in N'$, and, because $N \cap N' = 0$, we must have $n_1 - n_2 = 0 = n''' - n''$. If $n''' - n'' = 0$, then $n''' = 0 = n''$ because $N'' \cap N''' = 0$. Thus, $x = n_1 = n_2$ and hence x is in N which is impossible because N does not contain x. Consequently, either $x \notin N + N''$ or $x \notin N + N'''$ and this contradicts the maximality of N as a submodule of M' not containing x. Hence, N' must be a simple module.

We now use the fact that every submodule of M contains a simple module to show that M is semisimple. The idea is to take as large a semisimple submodule of M as we can find and show that it is M. We might naively start by considering the set of all semisimple submodules of M and ordering this set by inclusion. However, we would have difficulty then in showing that the union of a completely ordered subset is semisimple. Thus, we have to be a little careful about how we define our ordered set of submodules. Let $\mathscr{L}_1$ be the set of simple submodules of M and let $2^{\mathscr{L}_1}$ be the set of subsets of $\mathscr{L}_1$. An element of $2^{\mathscr{L}_1}$ is then a set $S = \{M_\alpha\}$ of simple submodules of M. We define a subset $\mathscr{L}$ of $2^{\mathscr{L}_1}$ by saying that an element $S = \{M_\alpha\}$ of $2^{\mathscr{L}_1}$ is in $\mathscr{L}$ if and only if the submodule generated by the M_α, denoted by (S), is a sum of the M_α; that is, if $(S) = \coprod_\alpha M_\alpha$. Because $2^{\mathscr{L}_1}$ is an ordered set (the usual ordering), the subset $\mathscr{L}$ is also ordered. Obviously, $\mathscr{L}$ is not empty because M contains simple modules. To apply Zorn's lemma to $\mathscr{L}$, we want to show that if $\{S_k\}$ is a totally ordered subset of $\mathscr{L}$, then $S = \cup S_k$ is again in $\mathscr{L}$. Because each S_k is a set of simple submodules of M, $S = \cup S_k$ is a set of simple submodules of M, say $S = \{M_\alpha\}$. What we must show is that $((S), i_\alpha : M_\alpha \to (S))$ is a sum where i_α is the inclusion morphism. The reader should check that, because the set $\{S_k\}$ is totally ordered, the corresponding set of modules $\{(S_k)\}$ is totally ordered and $(S) = \cup(S_k)$.

Now suppose that we have a family of morphisms $f_\alpha : M_\alpha \to N$ for all $M_\alpha \in S$. We want to define a morphism $f : (S) \to N$ such that $f|M_\alpha = f_\alpha$ for all α, and show that such a morphism f is unique. This will show that $(S) = \coprod M_\alpha$. If such an f exists, it must be unique because the M_α generate (S). That the M_α generate (S) can be seen from the fact that each (S_k) is generated by some of the $M_\alpha \in S$ and $(S) = \cup(S_k)$.

We now show that an $f : (S) \to N$ such that $f|M_\alpha = f_\alpha$ for all α exists. For each k let $g_k : (S_k) \to N$ be the unique morphism such that $g_k|M_\alpha = f_\alpha$ for every M_α in S_k. Such a g_k exists, because each $S_k = \coprod_{M_\alpha \in S_k} M_\alpha$. Clearly, if $S_{k_1} \subset S_{k_2}$, then $g_{k_2}|(S_{k_1}) = g_{k_1}$. Thus, we define $f : (S) \to N$ by $f(x) = g_k(x)$ if $x \in (S_k)$ and, as we have seen several times already, this defines a map f which is a morphism. Also, $f|M_\alpha = (f|(S_k))|M_\alpha$ if $M_\alpha \in S_k$, and $f|(S_k) = g_k$ by definition, so $f|M_\alpha = g_k|M_\alpha = f_\alpha$. This shows that S is in $\mathscr{L}$ and thus, by Zorn's lemma, we may assume that $\mathscr{L}$ contains some maximal element which we shall again denote by $S = \{M_\alpha\}$.

We claim that $(S) = M$. But suppose not. Then $M = (S) \coprod M'$ because every submodule of M is a summand. Because M' is a nonzero submodule of M, M' contains a simple submodule N and, because $M' \cap (S) = 0$, we have $N \cap (S) = 0$. Hence, the submodule of M generated by N and (S) is a sum, and thus the set $S \cup \{N\}$ is an element of $\mathscr{L}$. This, however, contradicts the maximality of S in $\mathscr{L}$

and so we must have $(S) = M$. But (S), being a sum of simple modules, is semisimple and so we have shown that M is semisimple.

Corollary 2.5
If M is a semisimple module, then so is every submodule and factor module of M.

PROOF: If M is semisimple, then every submodule of M is a summand of M. But we have seen that if M' is a submodule of M, then every submodule of M' is also a summand of M' so that, by Proposition 2.4, M' is also semisimple.

If M'' is a factor module of M, we have an epimorphism $f: M \to M''$ and we let $M' = \text{Ker } f$. We thus obtain an exact sequence $0 \to M' \to M \to M'' \to 0$ which splits because M', being a submodule of M, is a summand of M. But then M'' is isomorphic to a submodule of M and hence is semisimple, because we have already shown that every submodule of M is semisimple.

Corollary 2.6
If R is a semisimple ring, then every R-module is semisimple.

PROOF: If R is semisimple, then every free R-module is semisimple by Basic Property 2.1 of semisimple modules. However, because every module is a factor module of a free R-module (Chapter 6, Proposition 7.9), we have, by Corollary 2.5, that every module is semisimple.

Corollary 2.7
If R is a sum of copies of a fixed simple module M_0, then every R-module is the sum of copies of M_0.

PROOF: The hypothesis on R implies that R is semisimple so that every R-module is semisimple, by Corollary 2.6. By the proof of Basic Property 2.1 of semisimple modules, we know that every R-module contains a simple submodule isomorphic to M_0. Thus, if M is itself a simple module, it must be isomorphic to M_0 and hence every simple R-module is isomorphic to M_0. Consequently, because every R-module is a sum of simple modules, we have that every R-module is a sum of copies of M_0.

We have therefore proved Theorem 1.7, because the only remaining step to prove was the implication (**b**) implies (**c**), which we have just completed.

Earlier in this section we mentioned simple rings. We are now ready to consider the following.

Example 2.8 Let D be a division ring, let V be a finite-dimensional vector space over D, and let $R = \text{End}_D(V)$, the ring of endomorphisms of V. We shall show that R is a simple ring, that is, it has no ideals other than (0) and (1). We shall show this by considering a correspondence between left ideals of R and subvector spaces of V as follows: For each subspace W of V, let $I(W) = \{f \in R \mid f(x) = 0 \text{ for all } x \in W\}$, and for each left ideal I of R, let $W(I) = \{x \in V \mid f(x) = 0 \text{ for all } f \in I\}$. The reader can easily check that $I(W)$ is a left ideal of R and that $W(I)$ is a vector subspace of V. What we will prove is that for every

subspace W of V, we have $W(I(W)) = W$ and that for every left ideal I of R we have $I(W(I)) = I$.

To prove that $W(I(W)) = W$ is easy. For clearly, $W(I(W)) \supset W$ for any subspace W of V. It is equally obvious that $I(W(I)) \supset I$ for any left ideal I of R. To see that $W \supset W(I(W))$, we shall prove that if $x \notin W$, then $x \notin W(I(W))$. To say that $x \notin W(I(W))$ means that there is some $f \in I(W)$ such that $f(x) \neq 0$. But $I(W)$ consists of those $f: V \to V$ such that $W \subset \text{Ker } f$. If we produce an $f: V \to V$ whose kernel is precisely W, then f will be in $I(W)$ and $f(x) \neq 0$ if $x \notin W$. However, because V is a vector space, every subspace of V is a summand, so, in particular, there is a subspace W' such that $V = W \amalg W'$. The projection morphism $p': V \to W'$ has as its kernel the subspace W. The morphism $f = i'p': V \to V$, where $i': W' \to V$ is the inclusion, also has kernel W. This proves that $W(I(W)) = W$. The reader should observe that the finite dimensionality of V was not used in this argument.

To prove that $I(W(I)) = I$ requires more power. We have already observed that $I \subset I(W(I))$ so we must show that $I \supset I(W(I))$. Because I is a submodule of R and R is semisimple, we know that I is a summand of R. Thus, there is an epimorphism $p: R \to I$ such that $pi: I \to I$ is the identity, where $i: I \to R$ is the inclusion. This tells us that I is generated by $e_0 = p(1)$, because 1 generates R as an R-module. Because p is a morphism of R-modules, we have $e_0^2 = e_0 e_0 = e_0 p(1) = p(e_0 \cdot 1) = p(e_0) = p(i(e_0)) = pi(e_0) = e_0$.

Using this information, we are ready to look at $I(W(I))$. First of all, $W(I) = \text{Ker } e_0$. For $W(I) = \{x \in V | f(x) = 0 \text{ for all } f \in I\}$. But then, because $e_0 \in I$, $W(I) \subset \text{Ker } e_0$. However, if $x \in \text{Ker } e_0$ and $f \in I$, then $f = f' e_0$ for some $f' \in R$, because e_0 generates I as a left ideal, and $f(x) = f' e_0(x) = 0$. Hence, if $x \in \text{Ker } e_0$, $x \in W(I)$ and so we do have $W(I) = \text{Ker } e_0$. Now let g be an element of $I(W(I))$. We have $W(I) \subset \text{Ker } g$, so that $\text{Ker } e_0 \subset \text{Ker } g$ [do not forget that $W(I) = \text{Ker } e_0$]. From this, together with the fact that $e_0^2 = e_0$, we may conclude that $g = ge_0$, that is, $g(v) = ge_0(v)$ for all $v \in V$. Hence, $(v - e_0(v)) \in \text{Ker } e_0$ for all v. If $\text{Ker } g$ contains $\text{Ker } e_0$, we then have $g(v - e_0(v)) = 0$ for all $v \in V$, or $g(v) = ge_0(v)$ for all v, and so $g = ge_0$ which is in I.

Knowing that $I(W(I)) = I$ and $W(I(W)) = W$, the map from the set $\mathcal{I}$ of left ideals of R to the set $\mathcal{W}$ of vector subspaces of V defined by $I \mapsto W(I)$ is easily seen to be bijective. For, if $W(I_1) = W(I_2)$, we have $I_1 = I(W(I_1)) = I(W(I_2)) = I_2$ so that the map is injective, while, if $W \in \mathcal{W}$, we have $W = W(I(W))$ so that the map is surjective. We now use this isomorphism between $\mathcal{I}$ and $\mathcal{W}$ to show that R is a simple ring.

Suppose that I is an ideal. Then $W(I)$ is not only a subvector space of V but is actually an R-submodule of V. To see this we want to show that, if $x \in W(I)$ and $g \in R$, then $g \cdot x = g(x)$ is again in $W(I)$. That is, if $f(x) = 0$ for all $f \in I$, we want to show that $f(g(x)) = 0$ for all $f \in I$. However, $f(g(x)) = (fg)(x)$ and, since I is assumed to be an ideal, fg is again in I, so that $fg(x) = 0$. Thus, $W(I)$ is an R-module. Because V is a simple R-module, we must have $W(I) = 0$ or $W(I) = V$. From the isomorphism between $\mathcal{I}$ and $\mathcal{W}$, we then conclude that $I = R$ or $I = (0)$. Hence, the only ideals of R are (0) and R, and R is thus a simple ring.

Besides showing that R is simple, we can use this isomorphism between left ideals and subvector spaces to establish that the intersection of all maximal left

ideals of R is zero. First of all, if W_1 and W_2 are subspaces of V, we have $I(W_1) \cap I(W_2) \subset I(W_1 + W_2)$, where $W_1 + W_2$ is the subspace of V generated by W_1 and W_2. This is clear, because if $f \in I(W_1) \cap I(W_2)$, then $f(x) = 0$ for all $x \in W_1$ and $f(x) = 0$ for all $x \in W_2$, so $f(x) = 0$ for all $x \in W_1 + W_2$. If W is a one-dimensional subspace of V, then W contains no subspace other than (0), so that, by our correspondence between W's and I's, we know that $I(W)$ is contained in no ideal other than R. Hence, $I(W)$ is a maximal left ideal of R. If $\{x_1, \ldots, x_n\}$ is a basis for V, and W_i is the subspace generated by x_i, then $W_1 + \cdots + W_n = V$. Thus, $I(V) = I(W_1 + \cdots + W_n) \supset I(W_1) \cap \cdots \cap I(W_n)$. But $I(V) = 0$ so that $I(W_1) \cap \cdots \cap I(W_n) = 0$. Because the intersection of the maximal left ideals $I(W_i)$ is zero, it follows that the intersection of all the maximal left ideals in R is zero.

As a consequence of this discussion we have the following.

Proposition 2.9
Let V be a finite-dimensional vector space over a division ring D. Then the ring $R = \mathrm{End}_D(V)$ has the following properties:

(a) R is a simple ring.
(b) The intersection of the maximal left ideals of R is zero.
(c) R is left artinian and left noetherian.

PROOF: (a) and (b) have already been demonstrated.

(c) It is not difficult to see that, if I_1 and I_2 are left ideals of R, then $I_1 \subset I_2$ if and only if $W(I_1) \supset W(I_2)$. Thus, if $I_1 \supset I_2 \supset \cdots \supset I_n \supset \cdots$ is a descending chain of left ideals of R, then $W(I_1) \subset W(I_2) \subset \cdots \subset W(I_n) \subset \cdots$ is an ascending chain of subvector spaces of V. The vector space V is a noetherian D-module, because it is a finitely generated module over the noetherian ring D. Thus, for an integer m, we have $W(I_{m+k}) = W(I_m)$ for all $k \geq 0$. This implies that $I_{m+k} = I_m$ for all $k \geq 0$. Hence, R is a left artinian ring.

The proof that R is left noetherian proceeds similarly.

3. PROJECTIVE MODULES

We have seen in our proof of Theorem 1.7 that if every module over a ring R is a sum of copies of a simple module M_0, then R is a pretty special sort of ring. Is M_0, besides being simple, any special sort of module? We know that M_0 cannot, in general, be free, for then that would imply that R is a division ring. It turns out, however, that M_0 does share some properties with free modules. For instance, because R itself is a sum of copies of M_0, we have morphisms $i : M_0 \to R$ and $p : R \to M_0$ such that $pi = \mathrm{id}_{M_0}$. That is, M_0 is a summand of a free module, namely, R. Since every free module is a summand of a free module, namely, itself, this is a property shared by M_0 and free modules. Proposition 7.10, Chapter 6, tells us that if $0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0$ is an exact sequence of R-modules and if F is any free R-module, then $0 \to (F, M') \to (F, M) \to (F, M'') \to 0$ is exact. We claim that M_0, being a summand of R, also has the property that $0 \to (M_0, M') \to (M_0, M) \to (M_0, M'') \to 0$ is exact. In fact, if all we knew were that M_0 was

a summand of a free R-module, not necessarily of R itself, the same would be true.

To see this, suppose that M_0 is a summand of the free module F and that we therefore have morphisms $i: M_0 \to F$ and $p: F \to M_0$ such that $pi = \mathrm{id}_{M_0}$. Given an h in (M_0, M''), we establish the exactness of the sequence $(M_0, M) \xrightarrow{(M_0, g)} (M_0, M'') \longrightarrow 0$ by producing an element q in (M_0, M) such that $gq = h$. To produce the morphism q, we first consider the morphism $hp: F \to M''$. Because F is free, there is a morphism $h': F \to M$ such that $gh' = hp$. Define $q: M_0 \to M$ by $q = h'i$. Then $gq = g(h'i) = (gh')i = (hp)i = h(pi) = h$, and so we have the result.

This property of M_0 is so important that modules having this property are given a special name.

Definition

A module P is called **projective** if for every exact sequence of R-modules, $0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0$, the sequence $0 \to (P, M') \to (P, M) \to (P, M'') \to 0$ is exact.

Basic Properties 3.1

(a) A module is projective if and only if it is a summand of a free module.
(b) A module P is projective if and only if every exact sequence $0 \to M' \to M \to P \to 0$ of R-modules splits.
(c) A sum of modules is projective if and only if each summand is projective.

PROOF: (a) We have already seen that if P is a summand of a free module, then P is projective. We know there is an epimorphism $F \xrightarrow{g} P$ where F is a free module. If P is projective, the sequence $(P, F) \to (P, P) \to 0$ is exact, so, in particular, there is a morphism $h: P \to F$ such that $gh = \mathrm{id}_P$. Thus, P is a summand of the free module F.

(b) If P is projective and $0 \to M' \to M \to P \to 0$ is exact, then so is $0 \to (P, M') \to (P, M) \to (P, P) \to 0$. Thus, there is a morphism $h: P \to M$ such that the composition $P \to M \to P$ is id_P and hence the original sequence splits. Conversely, if every exact sequence $0 \to M' \to M \to P \to 0$ splits, then in particular the sequence $0 \longrightarrow M' \longrightarrow F \xrightarrow{g} P \longrightarrow 0$ splits, where $g: F \to P$ is a surjective morphism of a free module F onto P and $M' = \mathrm{Ker}(F \xrightarrow{g} P)$. Therefore, P is a direct summand of F and, by (a), is projective.

(c) Let $\{P, i_\alpha: P_\alpha \to P\}$ be a sum of the modules P_α and let $p_\alpha: P \to P_\alpha$ be the corresponding projections. If P is a projective, then clearly each P_α is projective. One may use the same proof that was used in showing that a summand of a free module is projective. If we now assume that each P_α is projective, we show that P is projective as follows. Let $0 \longrightarrow M' \longrightarrow M \xrightarrow{g} P \longrightarrow 0$ be exact. Because each P_α is projective, there is a morphism $h_\alpha: P_\alpha \to M$ such that $gh_\alpha = i_\alpha$ for each α. Thus, there is a unique morphism $h: P \to M$ such that $hi_\alpha = h_\alpha$ for each α. To show that h is a splitting for g, that is, $gh = \mathrm{id}_P$, we need only show that $(gh)i_\alpha = i_\alpha$ for each α. But $(gh)i_\alpha = g(hi_\alpha) = gh_\alpha = i_\alpha$ and we are done.

Having these basic properties of projective modules at our disposal, we can get a slight refinement of Theorem 1.7.

Theorem 3.2
Let R be a ring. Then the following statements are equivalent:

(a) R is the endomorphism ring of a finite-dimensional vector space over a division ring.
(b) R is the sum of copies of a simple module M_0.
(c) Every R-module is the sum of copies of a projective R-module P_0.
(d) Every R-module is the sum of copies of a simple R-module M_0.

PROOF: Because we have already shown that (a) implies (b) and (d) implies (a), we need only show that (b) implies (c) and (c) implies (d). We have already seen that M_0 is projective if R is the sum of copies of M_0. Because (b) implies that every R-module is the sum of copies of M_0 (Theorem 1.7), every R-module is the sum of copies of a projective module $P_0 = M_0$. Hence, (b) implies (c). In order to prove that (c) implies (d), we first prove Proposition 3.3.

Proposition 3.3
Let R be a ring. Then every R-module is projective if and only if every R-module is semisimple.

PROOF: If every R-module is projective, then every exact sequence $0 \to M' \to M \to M'' \to 0$ splits. Thus, if M is any R-module and M' is a submodule of M, then M' is a summand of M, so that M is semisimple. Conversely, if every R-module is semisimple, let P be an R-module and we shall show that P is projective. Consider the exact sequence $0 \to M' \to F \to P \to 0$, where F is a free R-module. Because M' is a submodule of F and F is semisimple, the exact sequence splits. Hence, P is a summand of the free module F and thus is projective.

Getting back to our proof that (c) implies (d), suppose we know that every R-module is a sum of copies of a projective module P_0. Then, by Basic Property 3.1 of projective modules, we have that every module is projective and hence, by Proposition 3.3, every module is semisimple. Thus, every module is a sum of simple modules. But, if M_0 is a simple module, $M_0 = P_0$, because M_0, being a sum of copies of P_0, must contain a submodule isomorphic to P_0. Because M_0 is simple, M_0 is equal to this submodule so that M_0 and P_0 are isomorphic. Thus, (c) implies (d), and the proof of Theorem 3.2 is complete.

Condition (c) of Theorem 3.2 gives another description of endomorphism rings of vector spaces, which is very closely analogous to the description of division rings. Namely, for division rings, every module is the sum of copies of the fixed *free* module R, while for endomorphism rings of vector spaces every module is the sum of copies of a fixed *projective* module P_0.

We now turn to question (2) asked at the beginning of this chapter; namely, what rings have the property that all modules are sums of simple modules? In view of our results, this is the same as asking what rings R have the property that all R-modules are semisimple? From Proposition 3.3 we know that such rings are precisely those for which all modules are projective. But this still does not tell us much about the explicit structure of the ring itself. We shall devote the rest of this section to proving the following theorem.

Theorem 3.4
Let R be a ring. Then the following statements are equivalent:

(a) R is the product of a finite number of endomorphism rings of finite-dimensional vector spaces over division rings.
(b) R is semisimple.
(c) Every R-module is projective.
(d) Every R-module is semisimple.

PROOF: (a) implies (b). Let $D_1, \ldots, D_s$ be division rings, not necessarily distinct, and let $V_1, \ldots, V_s$ be finite-dimensional vector spaces over $D_1, \ldots, D_s$, respectively. Let R_i be the endomorphism ring of V_i over D_i for $i = 1, \ldots, s$ and let $R = R_1 \times \cdots \times R_s$. We show that R is the sum of simple R-modules.

Recall that R is a ring in which addition and multiplication are defined component-wise, that is, $(r_1, \ldots, r_s) + (r'_1, \ldots, r'_s) = (r_1 + r'_1, \ldots, r_s + r'_s)$ and $(r_1, \ldots, r_s)(r'_1, \ldots, r'_s) = (r_1 r'_1, \ldots, r_s r'_s)$. The map $p_i : R \to R_i$ defined by $(r_1, \ldots, r_s) \to r_i$ is a ring surjection. Because we have already seen that V_i is a simple R_i-module, V_i becomes an R-module via the morphism p_i, and, because p_i is surjective, V_i is a simple R-module. We also know that if V_i is a vector space of dimension n_i over D_i, then R_i is a sum of n_i copies of V_i as an R_i-module. Now, because R is a sum of R_i as R-modules, and because each R_i is a sum of copies of V_i as R_i-modules and hence also as R-modules, it follows that R is the sum of copies of the V_i and thus R is the sum of simple modules. This shows that (a) implies (b).

Before showing that (b) implies (a) we observe that, even though we have not assumed that all the D_i are distinct, nor that all the V_i are distinct, V_i and V_j are *not* isomorphic as R-modules if $i \neq j$. In fact, if $f : V_i \to V_j$ is an R-morphism, then for all $v_i \in V_i$ we have $f(v_i) = f(1_i v_i)$ where $1_i = (0, 0, \ldots, 1, 0, \ldots, 0)$ with 1 in the ith spot, and $f(1_i v_i) = 1_i f(v_i) = 0$ because $f(v_i) \in V_j$ and $1_i v_j = 0$ for all $v_j \in V_j$ if $i \neq j$.

The proof that (b) implies (a) is based on the following generalization of Proposition 1.5.

Proposition 3.5
Suppose E is an R-module. Let $\Lambda = \mathrm{End}_R(E)$, $\Omega = \mathrm{End}_\Lambda(E)$, and $\gamma' : R \to \Omega$, the canonical morphism. Let $M_1, \ldots, M_s$ be left R-modules such that $E = M_1^{n_1} \amalg \cdots \amalg M_s^{n_s}$ and such that $(M_i, M_j) = 0$ for $i \neq j$. In addition, let $\Lambda_j = \mathrm{End}_R(M_j)$ and $\Omega_j = \mathrm{End}_{\Lambda_j}(M_j)$, and let $\gamma_j : R \to \Omega_j$ be the corresponding morphisms. Finally, let $\gamma : R \to \Omega_1 \times \cdots \times \Omega_s$ be the obvious morphism. If γ' is injective, then so is γ. If γ' is surjective, then so is γ.

PROOF: We shall only sketch the proof because the details are pretty much as they were in the proof of Proposition 1.5. If $\omega_i \in \Omega_i$ are given for $i = 1, \ldots, s$, we define $\omega_i^{n_i} : M_i^{n_i} \to M_i^{n_i}$ by $\omega_i^{n_i}(m_{i1}, \ldots, m_{in_i}) = (\omega_i(m_{i1}), \ldots, \omega_i(m_{in_i}))$. We then define $\omega^n : E \to E$ by $\omega^n(e_1, \ldots, e_s) = (\omega_1^{n_1}(e_1), \ldots, \omega_s^{n_s}(e_s))$ where $e_i \in M_i^{n_i}$. Using the fact that $(M_i, M_j) = 0$ for $i \neq j$ and hence that $(M_i^{n_i}, M_j^{n_j}) = 0$ for $i \neq j$, we prove that ω^n is a Λ-endomorphism of E; that is, $\omega^n \in \Omega$. To do this, we first use the fact that $\Lambda = (E, E) = (\coprod_i M_i^{n_i}, \coprod_j M_j^{n_j}) = \coprod_{i,j} (M_i^{n_i}, M_j^{n_j}) = \coprod_i (M_i^{n_i}, M_i^{n_i})$ [because $(M_i^{n_i}, M_j^{n_j}) = 0$

for $i \neq j$] and then the fact, as in Proposition 1.5, that an element of $(M_i^{n_i}, M_i^{n_i})$ is a collection of morphisms $\lambda_i = (\lambda_{i\alpha\beta})$ where $\lambda_{i\alpha\beta} : M_i \to M_i$ is in Λ_i. Thus, if $(e_1, \ldots, e_s)$ is in E and $\lambda : E \to E$, then $\lambda = (\lambda_1, \ldots, \lambda_s)$, where $\lambda_i = (\lambda_{i\alpha\beta})$ and $\lambda(e_1, \ldots, e_s) = (\lambda_1(e_1), \ldots, \lambda_s(e_s))$. The operation of $(\lambda_{i\alpha\beta})$ on e_i is then the same as in 1.5. With this setup, the proof now proceeds as did the proof of 1.5.

To show that (b) implies (a) we assume that R is semisimple. Then $R = M_1^{n_1} \amalg \cdots \amalg M_s^{n_s}$ where each M_i is a simple R-module and M_i is not isomorphic to M_j if $i \neq j$. Because the M_i are nonisomorphic simple modules, $(M_i, M_j) = 0$ for $i \neq j$. Letting $E = R$ in Proposition 3.5, we see that the ring morphism $\gamma : R \to \amalg \Omega_i$ is an isomorphism where each $\Omega_i = \mathrm{End}_{\Lambda_i}(M_i)$ and each Λ_i is the division ring $\mathrm{End}_R(M_i)$. Since each M_i is a finite-dimensional vector space over Λ_i (Proposition 1.4), we have that (b) implies (a).

The remaining equivalences of Theorem 3.4 have already been established, so the proof of the theorem is complete.

Corollary 3.6
Every semisimple ring is a left artinian ring and left noetherian ring.

PROOF: Let R be a semisimple ring. Then $R = \coprod_{i=1}^{n} M_i$ where each M_i is a simple R-module. Obviously, every simple module over any ring is artinian as well as noetherian. Hence, $\coprod_{i=1}^{n} M_i$ is artinian (noetherian), because a finite sum of artinian (noetherian) modules is artinian (noetherian). Therefore, R is both a left artinian and a left noetherian ring.

4. THE OPPOSITE RING

Associated with every ring R is the ring R^0, called the **opposite ring**. The underlying set of R^0 is the same as that of R. The addition in R^0 is also the same as that of R. It is only in the multiplication that they differ. For r_1 and r_2 in R^0 we define the product $r_1 r_2$ to be the element $r_2 r_1$ in R. In short, the multiplication in R^0 is opposite to that in R. The reader is invited to check that R^0 is a ring. Obviously, $R^0 = R$ if and only if R is a commutative ring. It is equally obvious that $(R^0)^0 = R$.

Definition
An R^0-module M is called a **right R-module**. A morphism f of R^0-modules is called a **morphism of right R-modules.**

Suppose M is a right R-module. Since M is an R^0-module, we know that for each m in M and r_1, r_2 in R^0 we have $r_1(r_2 m) = (r_2 r_1)m$. If we write mr for rm, then the formula $r_1(r_2 m) = (r_2 r_1)m$ becomes $(mr_2)r_1 = m(r_2 r_1)$. Because this latter formula is easier to handle, we always write the operation of the elements of R on the right when dealing with right R-modules. In fact, this is the reason for this terminology.

In the light of this discussion we see that a right R-module M is the same thing as an abelian group M together with a map $M \times R \to M$ which we denote by $(m, r) \mapsto mr$ satisfying:

(a) $(mr_1)r_2 = m(r_1r_2)$.
(b) $(m_1 + m_2)r = m_1r + m_2r$.
(c) $m(r_1 + r_2) = mr_1 + mr_2$.
(d) $m1 = m$.

In this notation we see that a morphism $f: M \to N$ of right R-modules is the same thing as a morphism of abelian groups satisfying $f(mr) = f(m)r$.

In view of these remarks, the reader should have no difficulty expressing our results for modules for right modules. In particular, the ring R can be viewed as a right R-module. Moreover, the right submodules of R are precisely what we called, in Chapters 5 and 6, the right ideals of R.

Linguistic symmetry suggests that because we have right modules we should also have something called **left modules**, as we do for left and right ideals. However, the reader will immediately see, if he writes down the obvious axioms for a left module, that they are the same as the axioms already given for a module. In fact, most authors refer to modules as left modules, as we may also do occasionally for the sake of clarity.

We have defined a ring R to be semisimple if it is semisimple as a left module. We did not say that R is left semisimple because, as we shall presently show, a ring R is semisimple if and only if R^0 is also semisimple.

In dealing with this question we have to introduce a little more terminology.

Definitions
Let R and S be rings. A map $f: R \to S$ is called an **antimorphism** if it is a morphism of abelian groups satisfying $f(r_1r_2) = f(r_2)f(r_1)$ and $f(1) = 1$.

An antimorphism $f: R \to S$ is called an **anti-isomorphism** if it is a bijective map.

Basic Properties 4.1
Let R be a ring.

(a) The identity map of sets $\mathrm{id}_R: R \to R^0$ is an anti-isomorphism.
(b) The composition of two antimorphisms of rings is a ring morphism.
(c) A ring S is anti-isomorphic to R if it is isomorphic to R^0.
(d) If $f: R \to S$ is an antimorphism, then the subgroup $\mathrm{Ker}\, f$ of R is an ideal of R.

We say that any ring anti-isomorphic to R is an **opposite ring** of R.

Now, the left (right) module theory of a ring R is identical to the right (left) module theory of an opposite ring of R. For, if $f: R \to S$ is an anti-isomorphism with inverse $g: S \to R$, and if M is any left (right) R-module, then M becomes a right (left) S-module by setting $ms = g(s)m (sm = mg(s))$ for all $s \in S$ and all $m \in M$. If M is a simple left R-module, then M is a simple right S-module, and conversely. Also, if R is the sum of copies of a simple left R-module M (or if R is left semisimple), then S is clearly the sum of copies of the right S-module M (or S is right semisimple), and conversely. Every left R-module is R-projective if and only if every right S-module is projective, etc.

Suppose that a ring R is the sum of copies of one simple right R-module. Denoting by R^0 its opposite ring, we know that R^0 is the sum of copies of a simple left R^0-module. Hence, we know that R^0 is the endomorphism ring of a finite-

dimensional vector space over a division ring D. But what about R? If we could show that R, too, is the endomorphism ring of a (left or right) vector space over a division ring, that would imply that R is a sum of copies of a simple left R-module. Hence, R would be left semisimple as well as right semisimple and, in fact, this would imply that every left semisimple ring is also right semisimple, and conversely. This is so because every left semisimple ring is the product of a finite number of endomorphism rings of vector spaces over division rings. Because the opposite of a product of rings is the product of the opposite rings, we would have the implications

$$\begin{aligned} R \text{ left semisimple} &\Longrightarrow R^0 \text{ right semisimple} \Longrightarrow R^0 \text{ left semisimple} \\ &\Longrightarrow R \text{ right semisimple} \end{aligned}$$

On the other hand, we would have

$$\begin{aligned} R \text{ right semisimple} &\Longrightarrow R^0 \text{ left semisimple} \\ &\Longrightarrow R^0 \text{ right semisimple (by the above)} \\ &\Longrightarrow R \text{ is left semisimple} \end{aligned}$$

All this distinction between left and right semisimplicity will therefore disappear if we prove the following.

Proposition 4.2
Let R be the endomorphism ring of a finite-dimensional left vector space V over a division ring D. Let V^* be the group of D-morphisms of V into D. Then V^* can be considered a right vector space over D in such a way that R^0 is isomorphic to the endomorphism ring of V^* over D. Hence, R^0 is also left semisimple.

PROOF: We make V^* into a right D-module by setting v^*d to be the morphism of V into D defined by $v^*d(x) = v^*(dx)$ for all $x \in V$. The reader may check that V^* is a right D-module under this operation. If V is n-dimensional with basis $\{x_1, \ldots, x_n\}$, define $x_i^* \in V^*$ by $x_i^*(x_j) = \delta_{ij}$ (the Kronecker delta, that is, $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$). It can be easily verified that $\{x_1^*, \ldots, x_n^*\}$ is a basis for V^* over D, and thus V^* too is an n-dimensional right vector space over D. Let $S = \mathrm{End}_D(V^*)$. Because D is a division ring, its opposite ring D^0 is also a division ring and V^* is a left n-dimensional vector space over D^0. Obviously, $\mathrm{End}_D(V^*) = \mathrm{End}_{D^0}(V^*)$ so that S is the endomorphism ring of a finite-dimensional left vector space over a division ring and is therefore left semisimple. Define the map $h : R \to S$ as follows: If $f \in R$ (that is, $f : V \to V$) and $v^* \in V^*$, $h(f)$ is that element of S which takes the element v^* to $v^*f \in V^*$. Thus, $h(f)(v^*) = v^*f$. The map h is a group morphism which carries 1 into 1. Further, we have that $h(f_1f_2)(v^*) = v^*f_1f_2 = (v^*f_1)f_2 = h(f_2)(v^*f_1) = h(f_2)(h(f_1)(v^*)) = (h(f_2)h(f_1))(v^*)$ for all $v^* \in V^*$ so that $h(f_1f_2) = h(f_2)h(f_1)$. Hence, $h : R \to S$ is an antimorphism. Being an antimorphism, the kernel of h is a two-sided ideal of R. However, we know that R is a simple ring, so that Ker h is either (0) or R. Because $h(1) = 1 \neq 0$, Ker $h \neq R$ so Ker $h = (0)$, from which it follows that h is injective. If we can show that h is surjective, we will have shown that h is an anti-isomorphism and that $S \approx R^0$. Therefore, we must try to prove that h is surjective.

Let $g: V^* \to V^*$ be an element of S. We want to show that there is an $f: V \to V$ such that $g(v^*) = v^* f$ for all $v^* \in V^*$. Letting $\{x_1, \ldots, x_n\}$ be a basis for V, we have the corrresponding basis $\{x^\dagger, \ldots, x_n^\dagger\}$ of V^* as described at the beginning of our proof. It then suffices to find $f: V \to V$ such that $g(x_i^\dagger) = x_i^\dagger f$ for $i = 1, \ldots, n$ because from this it follows that $g(v^*) = v^* f$ for all $v^* \in V^*$. Suppose that $g(x_i^\dagger)(x_j) = c_{ij}$ and define $f: V \to V$ by $f(x_j) = \Sigma_k a_{kj} x_k$. Then $(x_i^\dagger f)(x_j) = x_i^\dagger(f(x_j)) = x_i^\dagger(\Sigma_k c_{kj}\delta_{ik}) = c_{ij} = g(x_i^\dagger)(x_j)$ for each i and j so that $g(x_i^\dagger) = x_i^\dagger f$ for all i, and h is therefore surjective. Hence, $S \approx R^0$ which completes the proof of the proposition, because $S = \mathrm{End}_D(V^*)$.

As any easy consequence of the fact that left semisimplicity implies right semisimplicity, we have the following.

Proposition 4.3
Every semisimple ring is left and right artinian as well as left and right noetherian.

EXERCISES

(1) Let R be a ring and M an R-module. We call the R-module M a **generator** if for every nonzero R-morphism $f: X \to Y$, the morphism $\mathrm{Hom}_R(M, f): \mathrm{Hom}_R(M, X) \to \mathrm{Hom}_R(M, Y)$ is not zero.

(a) Prove that if M is any R-module, then $M \amalg R$ is a generator.

(b) If X is an R-module, let I be the set of all R-morphisms from M to X. For each $i \in I$, let M_i be the module M and define the morphism $\epsilon: \coprod_{i \in I} M_i \to X$ by the property that $\epsilon(m_i) = i(m_i)$ for every $m_i \in M_i$. Prove that if M is a generator, then ϵ is an epimorphism. [*Hint*: Let $X' = \mathrm{Im}\, \epsilon$ and show that the canonical surjective morphism $X \to X/X'$ must be zero.]

(c) Prove that the R-module M is a generator if and only if for every R-module X there is a set I and an epimorphism $\coprod_{i \in I} M_i \to X$ where $M_i \approx M$ for each $i \in I$.

(d) Prove that the R-module M is a generator if and only if there is an epimorphism $M^n \to R$ for some positive integer n, where M^n denotes the sum of n copies of M.

(e) Consequently, prove that M is a generator if and only if $M^n \approx M' \amalg R$ for some R-module M' and some positive integer n.

(2) Let R be a ring and M an R-module. Recall that M is a balanced R-module if the canonical morphism $k: R \to \Omega$ is an isomorphism where $\Omega = \mathrm{End}_\Lambda(M)$ and $\Lambda = \mathrm{End}_R(M)$.

(a) Prove that if M is any R-module, then $M \amalg R$ is balanced.

(b) Use **(a)** to prove that if the R-module M is a generator, then M is balanced.

(c) Prove that every left ideal in a simple ring R is balanced.

(3) Let $R_1, \ldots, R_n$ be rings and let $R = R_1 \times \cdots \times R_n$.

(a) Prove that $C(R) = C(R_1) \times \cdots \times C(R_n)$ where $C(R)$ and $C(R_i)$ denote the centers of the rings R and R_i.

(b) Let k be a field and let $\phi: k \to R$ be a nontrivial ring morphism such that $\mathrm{Im}\, \phi$ is contained in $C(R)$. If $\pi_i: R \to R_i$ denotes the projection morphism, prove

that $\pi_i\phi : k \to R_i$ is a nontrivial ring morphism and that Im $\pi_i\phi$ is contained in $C(R_i)$ for $i = 1, \ldots, n$.

(c) Using the same notation as in (b), suppose that R, considered as a vector space over k via the morphism ϕ, is a finite-dimensional vector space over k. Prove that each ring R_i is a finite-dimensional vector space over k via the morphism $\pi_i\phi$.

(d) Prove that the center of a simple ring is always a field.

(e) If $R_i = \mathrm{End}_{D_i}(V_i)$ where D_i is a division ring and V_i is a finite-dimensional vector space over D_i, prove that $C(R_i) = C(D_i)$.

(f) Let R be a semisimple ring and suppose that $R \approx \prod_{i=1}^{n} \mathrm{End}_{D_i}(V_i)$ where D_i are division rings and V_i are finite-dimensional vector spaces over D_i. If R is a finite-dimensional vector space over a field k contained in $C(R)$, prove that D_i is also a finite-dimensional vector space over k and that k is contained in $C(D_i)$. [Really we should say that the projection $\pi_i : R \to \mathrm{End}_{D_i}(V_i)$ maps k monomorphically into the center of D_i.]

(g) Prove that if R is a commutative semisimple ring, then R is isomorphic to a finite direct product of fields.

(4) Let D be a division ring, k a field contained in $C(D)$, and suppose that D is a finite-dimensional vector space over k. Suppose that every polynomial $f(X)$ in $k[X]$ has a root in k. Prove that $D = k$. [*Hint*: If D is an n-dimensional vector space over k, and a is any element of D, then the elements $1, a, \ldots, a^n$ are linearly dependent over k. We therefore have elements $c_0, \ldots, c_n \in k$ such that $\sum_{i=0}^{n} c_i a^i = 0$ and not all $c_i = 0$. Let $f(X) = \sum c_i X^i$. Then $f(X)$ is a polynomial in $k[X]$ and thus has the root, a, in K where K is the subring of D generated by k and the element a. Let $g(X) \in k[X]$ be a polynomial of lowest possible degree which has a as a root. Then $g(X)$ also has a root b in k. Prove that $a = b$ and thus $a \in k$.]

(5) Let G be a finite group and R a commutative ring. If M and N are modules over the group ring $R(G)$ (see Exercise 1 of Chapter 4 for the definition of group ring and monoid ring) and if $f : M \to N$ is an R-module morphism, define the map $\tilde{f} : M \to N$ by $\tilde{f}(m) = \sum_{x \in G} x^{-1} f(xm)$ for all m in M.

(a) Prove that $\tilde{f} : M \to N$ is an $R(G)$-morphism.

(b) Prove that if $r \in R$, then $r\tilde{f} : M \to N$ defined by $(r\tilde{f})(m) = (\tilde{f}(m))$ for all m in M is an $R(G)$-morphism.

(6) Let G be a finite group of order n and let K be a field whose characteristic does not divide n. We will denote the element $n \cdot 1$ in K by n, and its inverse in K by $1/n$. Let $g : M \to N$ be a $K(G)$-morphism and let $f : N \to M$ be a K-morphism such that $gf = \mathrm{id}_N$.

(a) Prove that the composition $g((1/n)\tilde{f}) = \mathrm{id}_N$.

(b) From (a) deduce that $K(G)$ is a semisimple ring. [*Hint*: Prove that every $K(G)$-epimorphism $g : M \to N$ is splittable.

This latter fact is known as Maschke's theorem, that is, if G is a finite group and K is a field whose characteristic does not divide the order of G, then $K(G)$ is semisimple.]

(7) (a) Let G be a finite group and let $\mathbf{C}(G)$ be the group ring of G over the complex numbers. Prove that $\mathbf{C}(G)$ is isomorphic to the direct product of a finite number of matrix rings over the complex numbers. [*Hint*: Notice that $\mathbf{C} \subset C(\mathbf{C}(G))$

and that $\mathbf{C}(G)$ is a finite-dimensional vector space over $\mathbf{C}$. Then use Exercises 3, 4, and 7.]

(b) Prove that there are only finitely many nonisomorphic simple $\mathbf{C}(G)$-modules $V_1, \ldots, V_t$ and that each V_i is a finite-dimensional vector space over $\mathbf{C}$.

(c) Prove that the center of $\mathbf{C}(G)$ is isomorphic to the product of t copies of $\mathbf{C}$ where t is the number of nonisomorphic simple $\mathbf{C}(G)$-modules. Hence, the number of nonisomorphic simple $\mathbf{C}(G)$-modules is equal to the dimension of the center of $\mathbf{C}(G)$ as a vector space over $\mathbf{C}$.

(d) Assume, further, that G is an abelian group, so that $\mathbf{C}(G)$ is a commutative ring. Prove that $\mathbf{C}(G)$ has only a finite number of nonisomorphic simple modules and that these are all one-dimensional vector spaces over $\mathbf{C}$. Prove that the number of nonisomorphic simple modules is equal to the order of the abelian group G.

In the language of group representations, this last fact is stated as follows: All the irreducible complex representations of an abelian group are one-dimensional.

(8) (a) Let G be a finite group and R a commutative ring. Prove that an element $\Sigma_{g\in G} r_g g$ of $R(G)$ is in the center of $R(G)$ if and only if $r_g = r_{g'}$ whenever g and g' are conjugate elements of G. (Recall that two elements g and g' of a group G are said to be conjugate if there is an element h in G such that $g' = hgh^{-1}$. Conjugacy is an equivalence relation on the set G.)

(b) Let G be a finite group. Prove that the center of $\mathbf{C}(G)$ is a d-dimensional vector space over $\mathbf{C}$ where d is the number of conjugacy classes of G.

(c) Hence, prove that the number of nonisomorphic simple $\mathbf{C}(G)$-modules is equal to the number of conjugacy classes of G.

(9) (a) Let R be a ring and M an R-module. Suppose that the sequences $0 \to K_1 \xrightarrow{f_1} P_1 \xrightarrow{g_1} M \to 0$ and $0 \to K_2 \xrightarrow{f_2} P_2 \xrightarrow{g_2} M \to 0$ are exact and that P_1 and P_2 are projective R-modules. Prove that $P_1 \amalg K_2$ and $P_2 \amalg K_1$ are isomorphic. [*Hint*: Use the "pull-back" of

$$\begin{array}{ccc} & & P_2 \\ & & \downarrow{\scriptstyle g_2} \\ P_1 & \xrightarrow{g_1} & M \end{array}$$

(b) Show that the isomorphism $\theta : P_1 \amalg K_2 \to P_2 \amalg K_1$ can be so chosen that the morphism $u : P_1 \to P_2$ defined as the composition $P_1 \xrightarrow{i} P_1 \amalg K_2 \xrightarrow{\theta} P_2 \amalg K_1 \xrightarrow{p} P_2$ has the property that $g_2 u = g_1$ where i is the injection of P_1 into $P_1 \amalg K_2$ and p is the projection of $P_2 \amalg K_1$ onto P_2.

(10) Let R be a ring and let M and N be R-modules. If $0 \to N \xrightarrow{f_1} X_1 \xrightarrow{g_1} M \to 0$ and $0 \to N \xrightarrow{f_2} X_2 \xrightarrow{g_2} M \to 0$ are exact sequences, we say they are **equivalent** if there is an R-morphism $h : X_1 \to X_2$ such that the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N & \xrightarrow{f_1} & X_1 & \xrightarrow{g_1} & M & \longrightarrow & 0 \\ & & \| {\scriptstyle \mathrm{id}_N} & & \downarrow {\scriptstyle h} & & \| {\scriptstyle \mathrm{id}_M} & & \\ 0 & \longrightarrow & N & \xrightarrow{f_2} & X_2 & \xrightarrow{g_2} & M & \longrightarrow & 0 \end{array}$$

commutes.

(a) Prove that if the morphism h exists, it is an isomorphism.

(b) Prove that the equivalence we have just defined is reflexive, symmetric, and transitive.

(c) Let $0\to K\xrightarrow{\alpha}P\xrightarrow{\beta}M\to 0$ be an exact sequence, with P a projective R-module. Prove that if $0\to N\xrightarrow{f}X\xrightarrow{g}M\to 0$ is an exact sequence, then there are R-morphisms $h:P\to X$ and $h':K\to N$ such that the diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & K & \xrightarrow{\alpha} & P & \xrightarrow{\beta} & M & \longrightarrow 0 \\ & & \downarrow{\scriptstyle h'} & & \downarrow{\scriptstyle h} & & \| & \\ 0 & \longrightarrow & N & \xrightarrow{f} & X & \xrightarrow{g} & M & \longrightarrow 0 \end{array}$$

commutes.

(d) Using the notation of part (c), let $0\to N\xrightarrow{f'}Y\xrightarrow{g'}M\to 0$ be the exact sequence obtained from $0\to K\xrightarrow{\alpha}P\xrightarrow{\beta}M\to 0$ and the morphism h' as in Chapter 6, Exercise 28. Prove that the exact sequences $0\to N\xrightarrow{f}X\xrightarrow{g}M\to 0$ and $0\to N\xrightarrow{f'}Y\xrightarrow{g'}M\to 0$ are equivalent. Hence, show that the cardinality of the set of equivalence classes of exact sequences of the form $0\to N\xrightarrow{f}X\xrightarrow{g}M\to 0$ cannot exceed the cardinality of $\mathrm{Hom}_R(K, N)$.

(11) We retain the notation and terminology of the preceding exercise and let $\mathrm{Ext}^1_R(M, N)$ denote the set of equivalence classes of exact sequences of the form: $0\to N\xrightarrow{f}X\xrightarrow{g}M\to 0$.

(a) Define a morphism $\phi:\mathrm{Hom}_R(K, N)\to\mathrm{Ext}^1_R(M, N)$ as follows. For each R-morphism $h':K\to N$, let $0\to N\xrightarrow{f}Y\xrightarrow{g}M\to 0$ be the exact sequence obtained from the exact sequence $0\to K\xrightarrow{\alpha}P\xrightarrow{\beta}M\to 0$ and the R-morphism $h':K\to N$. Let $\phi(h')$ be equal to the equivalence class in $\mathrm{Ext}^1_R(M, N)$ of the exact sequence $0\to N\xrightarrow{f}Y\xrightarrow{g}M\to 0$. Prove that ϕ is a surjective map.

(b) Prove that if $h'_1:K\to N$ and $h'_2:K\to N$ are two R-morphisms from K to N, then $\phi(h'_1)=\phi(h'_2)$ if and only if $h'_1=h'_2+t\alpha$ for some R-morphism $t:P\to N$.

(c) Prove that there is a unique group structure on $\mathrm{Ext}^1_R(M, N)$ such that the map $\phi:\mathrm{Hom}_R(K, N)\to\mathrm{Ext}^1_R(M, N)$ is a morphism of groups and that the sequence of abelian groups

$$\mathrm{Hom}_R(P, N)\xrightarrow{\mathrm{Hom}_R(\alpha, N)}\mathrm{Hom}_R(K, N)\longrightarrow\mathrm{Ext}^1_R(M, N)\longrightarrow 0$$

is exact.

(d) Let $0\to K'\xrightarrow{\alpha'}P'\xrightarrow{\beta'}M\to 0$ be an exact sequence with P' a projective R-module. Let $\phi':\mathrm{Hom}_R(K', N)\to\mathrm{Ext}^1_R(M, N)$ be the map analogous to the map ϕ defined in part (a). Prove that ϕ' is a morphism of abelian groups where the group structure on $\mathrm{Ext}^1_R(M, N)$ is the one defined using the map ϕ in part (c). [*Hint*: Use Exercise 9.]

(e) Let E_1 and E_2 be elements of $\mathrm{Ext}^1_R(M, N)$ represented by the exact sequences $0\to N\xrightarrow{f_1}X_1\xrightarrow{g_1}M\to 0$ and $0\to N\xrightarrow{f_2}X_2\xrightarrow{g_2}M\to 0$, respectively.

(i) Prove that the sequence

$$0\longrightarrow N\amalg N\xrightarrow{f_1\amalg f_2}X_1\amalg X_2\xrightarrow{g_1\amalg g_2}M\amalg M\longrightarrow 0 \qquad (*)$$

is exact.

(ii) Let $\Delta:M\to M\amalg M$ be the morphism defined by $m\mapsto(m, m)$ for all m in M, and let $\nabla:N\amalg N\to N$ be the morphism defined by $(n_1, n_2)\mapsto n_1+n_2$.

The exact sequence (*) together with the morphism $\Delta: M \to M \amalg M$ gives rise to the exact sequence

$$0 \longrightarrow N \amalg N \xrightarrow{s} Y \xrightarrow{t} M \longrightarrow 0 \qquad (**)$$

and the exact sequence (**) together with the morphism $\nabla: N \amalg N \to N$ gives rise to an exact sequence

$$0 \longrightarrow N \xrightarrow{f} Z \xrightarrow{g} M \longrightarrow 0 \qquad (E)$$

Define $E_1 + E_2$ to be the equivalence class of the exact sequence (E) in $\mathrm{Ext}^1_R(M, N)$. Prove that $E_1 + E_2$ is the sum of E_1 and E_2 in the group structure of $\mathrm{Ext}^1_R(M, N)$ which we defined in part (c) using the map ϕ.

(12) **(a)** Let $u: M_1 \to M_2$ be a morphism of R-modules and let N be any R-module. Consider exact sequences

$$0 \longrightarrow K_1 \xrightarrow{\alpha_1} P_1 \xrightarrow{\beta_1} M_1 \longrightarrow 0$$

and

$$0 \longrightarrow K_2 \xrightarrow{\alpha_2} P_2 \xrightarrow{\beta_2} M_2 \longrightarrow 0$$

with P_1 and P_2 projective R-modules. Prove that there exist R-morphisms $h: P_1 \to P_2$ and $h': K_1 \to K_2$ such that $\beta_2 h = \beta_1$ and $h\alpha_1 = \alpha_2 h'$. Using the morphisms $\phi_1: \mathrm{Hom}_R(K_1, N) \to \mathrm{Ext}^1_R(M_1, N)$ and $\phi_2: \mathrm{Hom}_R(K_2, N) \to \mathrm{Ext}^1_R(M_2, N)$ together with the morphisms h and h', define a morphism $E^*(u): \mathrm{Ext}^1_R(M_2, N) \to \mathrm{Ext}^1_R(M_1, N)$.

(b) Let E be an element of $\mathrm{Ext}^1_R(M_2, N)$ represented by the exact sequence

$$0 \longrightarrow N \xrightarrow{f} X \xrightarrow{g} M_2 \longrightarrow 0 \qquad (*)$$

Let $0 \to N \xrightarrow{f'} Y \xrightarrow{g'} M_1 \to 0$ be the exact sequence obtained from (*) and the morphism $u: M_1 \to M_2$ of part **(a)**. Prove that the equivalence class of the exact sequence $0 \to N \xrightarrow{f'} Y \xrightarrow{g'} M_1 \to 0$ in $\mathrm{Ext}^1_R(M_1, N)$ is independent of the choice of the representative exact sequence (*) for E. Define the map $\mathrm{Ext}^1_R(u, N): \mathrm{Ext}^1_R(M_2, N) \to \mathrm{Ext}^1_R(M_1, N)$ by sending the element E in $\mathrm{Ext}^1_R(M_2, N)$ to the equivalence class of $0 \to N \xrightarrow{f'} Y \xrightarrow{g'} M_1 \to 0$ in $\mathrm{Ext}^1_R(M_1, N)$. Prove that $\mathrm{Ext}^1_R(u, N)$ is a morphism of abelian groups.

(c) Using the notation of the preceding two parts, prove that the morphisms $E^*(u)$ and $\mathrm{Ext}^1_R(u, N)$ are the same.

(d) Let $v: N_1 \to N_2$ be an R-morphism and M an R-module. Carry out similar procedures for defining morphisms $E_*(v)$ and $\mathrm{Ext}^1_R(M, v)$ from $\mathrm{Ext}^1_R(M, N_1)$ to $\mathrm{Ext}^1_R(M, N_2)$, and show that they yield the same morphism.

(e) Finally, show that $\mathrm{Ext}^1_R(M, N)$ is an additive functor.

(13) Let $0 \to M' \xrightarrow{u} M \xrightarrow{v} M'' \to 0$ and $0 \to N' \xrightarrow{u'} N \xrightarrow{v'} N'' \to 0$ be exact sequences of R-modules.

(a) Let $0 \to K \xrightarrow{\alpha} P \xrightarrow{\beta} M \to 0$ be an exact sequence, with P a projective R-module. Define a map $\delta: \mathrm{Hom}_R(M, N'') \to \mathrm{Ext}^1_R(M, N')$ as follows: First show that if s is in $\mathrm{Hom}_R(M, N'')$, we obtain a commutative diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & K & \xrightarrow{\alpha} & P & \xrightarrow{\beta} & M \longrightarrow 0 \\ & & \downarrow{\scriptstyle h'} & & \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle s} \\ 0 & \longrightarrow & M' & \xrightarrow{u} & M & \xrightarrow{v} & M'' \longrightarrow 0 \end{array}$$

The morphism h' in $\mathrm{Hom}_R(K, M')$ goes, via the appropriate morphism $\phi : \mathrm{Hom}_R(K, M') \to \mathrm{Ext}^1_R(M, M')$, to $\phi(h')$. Show that $\phi(h')$ depends only on the morphism s and define $\delta(s) = \phi(h')$. Prove that δ is a morphism of abelian groups and that the sequence

$$0 \longrightarrow \mathrm{Hom}_R(M, N') \longrightarrow \mathrm{Hom}_R(M, N) \longrightarrow \mathrm{Hom}_R(M, N'') \xrightarrow{\delta} \mathrm{Ext}^1_R(M, N')$$
$$\longrightarrow \mathrm{Ext}^1_R(M, N) \longrightarrow \mathrm{Ext}^1_R(M, N'')$$

is exact.

(b) Define a map $\bar{\delta} : \mathrm{Hom}_R(M, N'') \to \mathrm{Ext}^1_R(M, N')$ as follows: Let s be an R-morphism from M to N''. Then the exact sequence $0 \to N' \to N \to N'' \to 0$ together with the morphism s yields an exact sequence $0 \to N' \to X \to M \to 0$. Define $\bar{\delta}(s)$ to be the equivalence class of $0 \to N' \to X \to M \to 0$ in $\mathrm{Ext}^1_R(M, N')$. Prove that $\bar{\delta}$ is a morphism and is, in fact, the morphism described in part (a).

(c) Carry out similar discussions to define morphisms ∂ and $\bar{\partial}$ from $\mathrm{Hom}_R(M', N)$ to $\mathrm{Ext}^1_R(M'', N)$. Show that these morphisms are equal and prove that the sequence

$$0 \longrightarrow \mathrm{Hom}_R(M'', N) \longrightarrow \mathrm{Hom}_R(M, N) \longrightarrow \mathrm{Hom}_R(M', N)$$
$$\xrightarrow{\partial} \mathrm{Ext}^1_R(M'', N) \longrightarrow \mathrm{Ext}^1_R(M, N) \longrightarrow \mathrm{Ext}^1_R(M', N)$$

is exact.

(14) (a) Prove that an R-module P is projective if and only if $\mathrm{Ext}^1_R(P, N) = 0$ for all R-modules N.

(b) Prove that R is a semisimple ring if and only if $\mathrm{Ext}^1_R(M, N) = 0$ for all R-modules M and N.

(c) Let R be the ring of integers, $\mathbf{Z}$. What is $\mathrm{Ext}^1_{\mathbf{Z}}(\mathbf{Z}/(m), \mathbf{Z}/(n))$?

Chapter 8 ARTINIAN RINGS

In the last chapter we described semisimple rings by means of their module theory. It is our aim in this chapter to give an ideal theoretic description of semisimple rings.

Our starting point is the observation made in Chapter 7, Proposition 4.3 that all semisimple rings are left and right artinian. This naturally suggests the question: Which left and right artinian rings are semisimple? In answering this question many classical notions, such as the radical of a ring, idempotents, etc., are introduced and discussed.

1. IDEMPOTENTS IN LEFT ARTINIAN RINGS

We begin by giving an ideal theoretic characterization of when a ring is isomorphic to the endomorphism ring of a finite-dimensional vector space over a division algebra.

Theorem 1.1
A ring R is the endomorphism ring of a finite-dimensional vector space over a division ring if and only if it is a simple left (or right) artinian ring.

PROOF: Having already proved half of this theorem (see Chapter 7, Proposition 2.9), we shall prove only that if R is left artinian and simple, then R is what we have claimed it to be. (If R is right artinian and simple, then R^0 is left artinian and

simple, etc.) Because R is left artinian, we know that the set of nonzero left ideals of R contains a minimal element, say I_0. Obviously, I_0 is a simple R-module. If our theorem is true, then all simple R-modules should be isomorphic to I_0 and R should be a sum of copies of I_0. We now proceed to prove this.

Suppose that R is not the sum of a finite number of copies of I_0. Consider the set $\{J\}$ of all nonzero left ideals of R which are not the sum of a finite number of copies of I_0. Because R is in this set, the set is nonempty. Therefore, there is a minimal element, say J_0, in the set. We want to show that J_0 contains I_0 (or an isomorphic copy of I_0) as a summand. For then we would have $J_0 = I_0 \amalg J_1$ where J_1 is a nonzero submodule of J_0 unless $J_0 \approx I_0$. Because $J_0 \in \{J\}$, J_0 cannot be isomorphic to I_0, so $J_1 \neq (0)$. If $J_1 \notin \{J\}$, then J_1 is itself a sum of copies of I_0, so that because $J_0 = I_0 \amalg J_1$, we would have $J_0 \notin \{J\}$, which is a contradiction. Thus, $J_1 \in \{J\}$ and $J_1 \subset J_0$, contradicting the minimality of $J_0 \in \{J\}$. This contradiction shows that R is a sum of copies of I_0 and thus we would be done.

We now show that I_0 is contained in J_0. To do this, let I_1 and I_2 be any two nonzero left ideals. Because R is a simple ring and $I_2 \neq 0$, the ideal generated by I_2 in R is all of R. Thus, we can find elements $b_1, \ldots, b_s \in I_2$ and $r_1, \ldots, r_s \in R$ such that $1 = \Sigma b_i r_i$. Then for every $a \in I_1$ we have $a = a \cdot 1 = a\Sigma b_i r_i = \Sigma(ab_i)r_i$ and $ab_i \in I_2$ for all i. Because $I_1 \neq (0)$, there is an a in I_1 which is not zero. Hence, $ab_i \neq 0$ for some i. Thus, there are elements $a_1 \in I_1$ and $a_2 \in I_2$ with $a_1 a_2 \neq 0$. In particular, then, we can find an element $a_1 \in I_0$ and an element $a_2 \in J_0$ such that $a_1 a_2 \neq 0$. Define the morphism $f: I_0 \to J_0$ by $f(x) = xa_2 \in J_0$. Because $f(a_1) = a_1 a_2 \neq 0$, $\operatorname{Ker} f \neq I_0$. Thus, because I_0 is simple, $\operatorname{Ker} f = 0$ and f is a monomorphism. The image of f is then a left ideal isomorphic to I_0 contained in J_0, and we may assume that this ideal is actually equal to I_0. This shows that J_0 contains a copy of I_0.

Next we show that I_0 is a summand of J_0. We can do this if we show that I_0 is a summand of R. By what has already been said, we know that we can find elements $a_1, a_2 \in I_0$ such that $a_1 a_2 \neq 0$. Because I_0 is simple, the left ideal I_0' consisting of all elements of the form aa_2 with a running through all elements of I_0 must be I_0 itself or (0). Since $a_1 a_2 \neq 0$, we have $I_0' = I_0$. In particular, $a_2 = aa_2$ for some $a \in I_0$, and therefore $aa_2 = a^2 a_2$ or $(a - a^2)a_2 = 0$. We claim that $a = a^2$. For, if not, then $a - a^2 \neq 0$ and, because I_0 is simple and $a - a^2$ is in I_0, we have that $a - a^2$ generates I_0. Then $a_1 = r(a - a^2)$ for some $r \in R$ and hence $a_1 a_2 = r(a - a^2)a_2 = 0$, which is a contradiction. Thus, $a = a^2$ and, because $a \neq 0$, I_0 is generated by a. Let $p: R \to I_0$ be defined by $p(r) = ra$. If $i: I_0 \to R$ is the inclusion morphism, then for each $x \in I_0$ we have $pi(x) = x$ because $x = ra$ for some $r \in R$ and thus $pi(x) = pi(ra) = p(ra) = ra^2 = ra = x$. Hence, $pi = \mathrm{id}_{I_0}$ which shows that I_0 is a summand of R. This implies that I_0 is a summand of any left ideal of R in which it is contained. In particular, I_0 is a summand of J_0 and our previous discussion then tells us that R is a sum of copies of I_0. Applying Theorem 1.7, Chapter 7, we now know that R is the endomorphism ring of a finite-dimensional vector space over a division ring.

Before isolating and generalizing some of the crucial arguments in this proof to obtain results about semisimple rings in general, let us look at an example of a ring which is simple but not artinian and hence not semisimple.

Example 1.2 Let D be a division ring and V an infinite-dimensional vector space over D with a denumerable basis Z, that is, $\text{card}(Z) = \text{card}(\mathbf{N})$. We shall use the fact, stated in an exercise in Chapter 6, that a nonfinite-dimensional subspace of a vector space with a denumerable basis also has a denumerable basis. Denote by R the endomorphism ring $\text{End}_D(V)$ of V over D. Let I be the set of elements $f \in R$ such that $\text{Im}\, f$ is a finite-dimensional subspace of V. Then the reader can verify easily that I is a two-sided ideal of R. Because the identity of R is not in I, and because I is clearly not zero, we see that R is not a simple ring. We now show that R/I is a simple ring.

Let $\bar{f}$ be an element of R/I, and let $f \in R$ be such that $k(f) = \bar{f}$ where $k: R \to R/I$ is the canonical morphism. If $\bar{f} \neq 0$, then $f \notin I$ so that $\text{Im}\, f$ is not finite-dimensional. Therefore, because $\text{Im}\, f$ is a subspace of V, the dimension of $\text{Im}\, f$ is also denumerable. Let X' be a basis for $\text{Ker}\, f$ and let X be a linearly independent subset of V such that $X' \cap X = \emptyset$ and $X' \cup X$ is a basis for V. Then $f(x)$ is a basis for $\text{Im}\, f$ and, because f restricted to the subspace generated by X is injective, we see that X is denumerable. Let X_1 and X_2 be denumerable subsets of X such that $X = X_1 \cup X_2$ and $X_1 \cap X_2 = \emptyset$. If X' is finite, choose X_1 to have the same number of elements as X'. Let $h': X' \to X_1$ be a bijective map and $h: X \to X_2$ a bijective map. Define $\tilde{h}: V \to V$ to be the morphism corresponding to the map $X' \cup X \to V$ which sends an element $x' \in X'$ to $h'(x')$ and an element $x \in X$ to $h(x)$. Then, because the restriction of $\tilde{h}$ to $X' \cup X$ is injective, $\tilde{h}$ itself is injective. The image of $\tilde{h}$ is, clearly, the subspace generated by X.

Now extend $f(X)$ to a basis Y of V so that $Y = f(X) \cup Y'$ and $f(X) \cap Y' = \emptyset$. Let $g: Y \to V$ be a map such that $g(y') = 0$ for $y' \in Y'$ and $g|f(X)$ be a bijective map of $f(X)$ onto Y. Because $f(X)$ and Y are denumerable, this is possible. Now let $\tilde{g}: V \to V$ be the morphism corresponding to the map g. Clearly $\tilde{g}$ is an epimorphism because the image contains Y and $\text{Ker}\, \tilde{g} \cap \text{Im}\, f = 0$. We claim that $\tilde{g}\tilde{f}\tilde{h}$ is an automorphism of V. For, if $\tilde{g}\tilde{f}\tilde{h}(v) = 0$, then $\tilde{f}\tilde{h}(v) \in \text{Ker}\, \tilde{g}$ so that, because $\text{Ker}\, \tilde{g} \cap \text{Im}\, f = 0$, we have $\tilde{f}\tilde{h}(v) = 0$. But $\tilde{h}(v)$ is in the subspace generated by X and f is injective on that subspace, so this tells us $\tilde{h}(v) = 0$. Because $\tilde{h}$ is injective, we have $v = 0$ and thus $\tilde{g}\tilde{f}\tilde{h}$ is injective.

To see that $\tilde{g}\tilde{f}\tilde{h}$ is surjective, we observe that $\tilde{g}$ is surjective and, in fact, any element of V is the image under $\tilde{g}$ of an element in the subspace generated by $f(X)$. Because the image of $\tilde{h}$ is the space generated by X, it is clear that $\tilde{g}\tilde{f}\tilde{h}$ is surjective and thus the morphism is an automorphism. This tells us that the ideal generated in R by f is all of R, and so the same can be said for $\bar{f}$ in R/I. As a result, we see that R/I is a simple ring.

Because R/I is simple, we know that R/I is semisimple if and only if it is left or right artinian. Let us show that R/I is not right artinian. Since V has a denumerable basis X, we may write X as the union of a denumerable number of disjoint denumerable subsets: $X = \bigcup_{n \geq 1} X_n$ where $X_n \cap X_m = \emptyset$ for $n \neq m$ and each X_n is denumerable. If I_k' is the set of $f \in R$ such that $\text{Im}\, f$ is contained in the subspace generated by $\bigcup_{n \geq k} X_n$, then I_k' is a right ideal of R. Let $I_k = I_k' + I$, and let $\bar{I}_k = I_k/I$. Because $I_1' \supset I_2' \supset \cdots$, we have $I_1 \supset I_2 \supset \cdots$ and so $\bar{I}_1 \supset \bar{I}_2 \supset \cdots$ is a decreasing chain of right ideals in R/I. We claim that the ideals in this chain of ideals do not satisfy the descending chain condition, and thus R/I is not right

artinian. If it did, we would have $I_n = I_{n+k}$ for some integer n and all $k \geq 0$. But we can show that for every integer n, $I_n \not\subset I_{n+1}$. To see this, take an $f: V \to V$ such that Im f is the whole subspace generated by $\bigcup_{k \geq n} X_k$. This can be done by choosing a surjective map of X onto $\bigcup_{k \geq n} X_k$ and then extending. If $f \in I_{n+1}$, then $f = g + h$ where $g \in I'_{n+1}$ and $h \in I$, that is, Im h is finite-dimensional. Because X_n is denumerable and Im g is contained in the subspace generated by $\bigcup_{k \geq n+1} X_k$, we clearly cannot have $\text{Im}(g + h)$ equal to the subspace generated by $\bigcup_{k \geq n} X_k$. Thus, $f \notin I_{n+1}$ and the proof is complete.

If we analyze the proof of Theorem 1.1, we see that if all we had wanted to show was that R is left semisimple we would not have had to bother showing that I_0 is contained in J_0 but simply that if $\{J\}$ is the set of nonsemisimple submodules of R and J_0 is a minimal element of this set, then any minimal ideal contained in J_0 is a summand of J_0. Because R is assumed artinian, every left ideal of R contains a minimal left ideal of R, so then, if each left ideal I were a summand of J_0, we would have $J_0 = I \oplus J_1$, and if J_1 were not semisimple, the minimality of $J_0 \in \{J\}$ would be contradicted. If J_1 were semisimple, so too would be J_0 (which is absurd), and so we would have to conclude that $\{J\}$ is empty and that therefore R is left semisimple. Thus, the crucial step in such a proof is showing that if I is a minimal left ideal of R, then I is a summand of R (and hence of any left ideal that contains I). The proof of Theorem 1.1 shows us that we may deduce that I is a summand of R from the fact that $I^2 \neq (0)$ where I^2 is the left ideal consisting of all finite sums $\Sigma a_i b_i$ with a_i and $b_i \in I$. This observation immediately leads us to the following.

Proposition 1.3
A ring R is left semisimple if and only if it is left artinian and for every minimal left ideal I of R, $I^2 \neq 0$.

PROOF: If R is artinian and $I^2 \neq 0$ for every minimal left ideal I of R, then our foregoing discussion shows that R is left semisimple. The converse is straightforward, and the proof goes as follows.

We have already seen that if R is left semisimple, then R is left artinian (see Proposition 4.3, Chapter 7). If I is any nonzero ideal of R, I is a summand of R since R is semisimple, and therefore we have a morphism $R \xrightarrow{p} I$ such that $pi = \text{id}_I$ where $i: I \to R$ is the inclusion. Letting $e = p(1)$, we have $e^2 = ep(1) = p(e) = p(i(e)) = pi(e) = e$ so that, because $e^2 = e \neq 0$ we have $I^2 \neq 0$. Because $I^2 \neq 0$ for every left ideal of R, this is certainly true of minimal left ideals, and we are through.

We see from this proof and the proof of Theorem 1.1 that we actually have the following auxiliary result:

Proposition 1.4
Let R be any ring and I a nonzero left ideal of R. Then I is a summand of R if and only if I can be generated by an element e such that $e^2 = e$. If I is a minimal left ideal of R, then I is a summand of R if and only if $I^2 \neq 0$.

The reader is urged to supply the proof which may be obtained by reading the proofs of Theorem 1.1 and Proposition 1.3 carefully.

We have been talking about elements e such that $e^2 = e$, and have also rather cavalierly introduced a new term I^2. Let us stop to make two definitions.

Definitions

(a) An element e in a ring R is called an **idempotent** if $e = e^2$.
(b) If I is a left ideal of R and M is a left R-module, we denote by IM the submodule of M consisting of all finite sums $\Sigma\, a_i m_i$ with $a_i \in I$ and $m_i \in M$. We define I^nM inductively by setting $I^nM = I(I^{n-1}M)$. In particular, if M is a left ideal of R, we obtain left ideals IM, I^nM. If $M = I$, we get I^n defined for every n.

Basic Properties 1.5

(a) If e is an idempotent, then so is $1-e$ and $e(1-e) = (1-e)e = 0$.
(b) If I is a left ideal of R and M is a left R-module, then $I^{n_1}(I^{n_2}M) = I^{n_2}(I^{n_1}M) = I^{n_1+n_2}M$.

We leave the proofs of these properties to the reader.

Example 1.6 In the ring of two-by-two matrices over a field, the idempotents other than $\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$ and $\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$ are all matrices of the form $\begin{pmatrix}a & b\\ c & 1-a\end{pmatrix}$ with determinant zero. The reader can check easily that every such matrix is idempotent. To see that these are all the idempotents, suppose we have an idempotent $\begin{pmatrix}a & b\\ c & d\end{pmatrix}$. Then, because

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}a & b\\ c & d\end{pmatrix} = \begin{pmatrix}a^2+bc & b(a+d)\\ c(a+d) & cb+d^2\end{pmatrix},$$

we must have $a = a^2 + bc$, $b = b(a+d)$, $c = c(a+d)$, and $d = cb + d^2$. If either b or c is not zero, this forces the condition $a + d = 1$, and the condition $bc = a - a^2 = a(1-a)$ forces the determinant to be zero if $a + d = 1$. If $b = c = 0$, and if we do not take $a + d = 1$, then $a = d = 0$ or $a = d = 1$.

Example 1.7 If R is the ring of two-by-two triangular matrices over a field K, that is, R consists of all matrices of the form $\begin{pmatrix}a & b\\ 0 & c\end{pmatrix}$, then a similar computation shows that the idempotents are either $\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$, $\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$, $\begin{pmatrix}0 & b\\ 0 & 1\end{pmatrix}$, or $\begin{pmatrix}1 & b\\ 0 & 0\end{pmatrix}$. R is a three-dimensional vector space over the field K, so that R is both left and right artinian. This follows from the fact that a left (right) ideal of R is, in particular, a subvector space of R and thus, because finite-dimensional vector spaces are artinian, any decreasing sequence of left (right) ideals of R must satisfy the descending chain condition. R is not simple, nor is it semisimple. If it were semisimple, every left ideal would be a summand and hence, by Proposition 1.4, would be generated by an idempotent other that $\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$. However, in R we have

the left (actually two-sided) ideal I consisting of all elements $\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}$. Because this ideal contains no idempotent, it cannot be generated by an idempotent and so R is not semisimple. The elements $\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}$ have the property that $\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. This leads us to the following.

Definition
Let R be a ring and x an element of R. If $x^n = 0$ for some positive integer n, then x is said to be a **nilpotent** element. If I is a left ideal of R all of whose elements are nilpotent, then I is called a **nil left ideal** of R. If I is a left ideal of R such that $I^n = 0$ for some positive integer n, then I is called a **nilpotent left ideal**.

The example above shows that the ideal I of all elements $\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}$ is a nil ideal. In fact, because the product of any two elements in I is zero, I is nilpotent. The reader can show easily that every nilpotent element of R is in I, so that I consists of all the nilpotent elements of R. Finally, it should be observed that the map $h: R \to K \times K$ defined by $h\begin{pmatrix} x & y \\ 0 & z \end{pmatrix} = (x, z)$, is a surjective ring morphism whose kernel is I. Here, by $K \times K$ we mean the product of the field K with itself. As a consequence, because $K \times K$ is a semisimple ring and R/I is isomorphic to $K \times K$, we see that R/I is a semisimple ring.

2. THE RADICAL OF A LEFT ARTINIAN RING

The considerations of Section 1 lead us to the following questions. Suppose R is a left artinian ring. Is there always some ideal of R such that R/I is semisimple? Moreover, is it possible to describe such ideals explicitly? For example, we showed in Example 1.7 that dividing out by all nilpotent elements gave a semisimple ring.

Let us take these questions one at a time. Can we find an ideal I in the left artinian ring R such that R/I is semisimple? If I were a maximal ideal (that is, maximal among the set of ideals), then R/I would not only be a simple but also semisimple, because R and hence R/I is left artinian. If we wanted to use Zorn's lemma, we could easily prove that any ring R has a maximal ideal, and we would then have answered our first question for left artinian rings. However, because we are assuming that our ring is left artinian, perhaps we can find another proof of existence of maximal two-sided ideals without resorting to Zorn's lemma, a procedure which is always welcomed by mathematicians who prefer to have as many things independent of the Axiom of Choice as possible.

Because our ring R is left artinian, we know without using Zorn's lemma that R has simple modules, for instance, a minimal left ideal. Let M be a simple R-module and let $I = \text{ann}(M)$ be the annihilator of M. Recall that $\text{ann}(M) = \{x \in R \mid xm = 0 \text{ for all } m \in M\}$, and that $\text{ann}(M)$ is an ideal of R. Then M is an R/I-module which is faithful and simple also as an R/I-module. Recall that an R-module of M is faithful if $\text{ann}(M) = 0$. Thus, R/I is a left artinian ring having a

simple faithful module. If we show that the left artinian ring R/I is therefore a simple ring, this will show that I is a maximal ideal of R. We do this in the following.

Proposition 2.1
Let R be a left artinian ring. Then R is simple if and only if R has a simple faithful R-module.

PROOF: If R is left artinian, it has a simple module M. If, further, R is simple, we must have $\text{ann}(M) = 0$ because $\text{ann}(M)$ is an ideal and hence must be (0) or R. But $M \neq 0$ so $\text{ann}(M) \neq R$ and thus M is faithful.

Conversely, suppose R is artinian and has a simple faithful module M. If we show that R is semisimple and that all minimal left ideals are isomorphic to M, we will have that R is simple, because, in that case, R is a sum of minimal left ideals all isomorphic to a fixed module M. To show that a left artinian ring is semisimple, it suffices by Proposition 1.3 to show that if I is any minimal left ideal of R, then $I^2 \neq 0$. So, let I be any minimal left ideal of R. We observe that $IM \neq 0$, because otherwise we would have $I \subset \text{ann}(M)$ and because $I \neq 0$ and M is faithful [that is, $\text{ann}(M) = 0$] this is impossible. Therefore, there exists an element $x \in M$ such that $Ix \neq 0$, where $Ix = \{ax\}$ with $a \in I$. But then $Ix = M$ because M is simple and Ix is a nonzero submodule of M. We thus obtain an epimorphism $f: I \to M$ defined by $f(a) = ax$. Because $\text{Ker } f \neq I$ and I is minimal (and therefore simple), we have $\text{Ker } f = 0$ and f is an isomorphism. Therefore, we have shown that every minimal left ideal of R is isomorphic to M. Furthermore, $I^2 \neq 0$. Because, if $Ix = M$, there is an $a \in I$ such that $ax = x$. Thus, $a^2x = a(ax) = ax = x \neq 0$ and so $a^2 \neq 0$. Because $a^2 \in I^2$, we have $I^2 \neq 0$. This completes the proof of the proposition.

Theorem 2.2
Let R be a left artinian ring and let M be a simple R-module. Then $\text{ann}(M)$ is a maximal ideal of R. Furthermore, the correspondence $M \to \text{ann}(M)$ induces an isomorphism between the set of isomorphism classes of simple R-modules and the set of maximal ideals of R. Moreover, the set of maximal ideals of R is finite. If $\{I_1, \ldots, I_s\}$ is this set of maximal ideals of R, we have the ring $R/I_1 \cap \cdots \cap I_s$ is isomorphic to the ring $R/I_1 \times \cdots \times R/I_s$ defined by $k(r) = (k_1(r), \ldots, k_s(r))$ where $k_j: R \to R/I_j$ is the canonical surjection. Thus, $R/I_1 \cap \cdots \cap I_s$ is semisimple.

PROOF: If M is a simple R-module, then M is a simple faithful module over $R/\text{ann}(M)$, so that $\text{ann}(M)$ is a maximal ideal of R because, by Proposition 2.1, $R/\text{ann}(M)$ is simple. Clearly, if M_1 and M_2 are isomorphic, $\text{ann}(M_1) = \text{ann}(M_2)$. Therefore, the map sending a simple R-module M to $\text{ann}(M)$ induces a map φ from the set of isomorphism classes of simple R-modules to the set of maximal ideals of R. If I is a maximal ideal of R, then R/I is a left artinian simple ring and so R/I has a simple faithful module M which is also a simple R-module. Therefore, $\text{ann}(M) = I$ and so the map φ is surjective. It is injective because, if $\text{ann}(M_1) = \text{ann}(M_2) = I$, then M_1 and M_2 are both simple R/I-modules and M_1 and M_2 must be isomorphic, because R/I is a simple left artinian ring. Thus, the map φ is an isomorphism of sets.

If the set of maximal ideals of R were infinite, we could find a set $\{I_1, I_2, \ldots\}$

of distinct maximal ideals. Then, setting $J_k = I_1 \cap \cdots \cap I_k$, we would have a decreasing sequence of ideals: $J_1 \supset J_2 \supset \ldots$. Because R is left artinian, we get $J_n = J_{n+1} = \cdots$ for some integer n. But then, because $J_{n+1} = J_n \cap I_{n+1} = J_n$, we have $J_n \subset I_{n+1}$. This implies that $I_j \subset I_{n+1}$ for some $j = 1, \ldots, n$ and hence $I_j = I_{n+1}$.

To see this, consider first any two ideals I and J. Then the left ideal $I + J$ generalted by I and J (namely, the set of all elements $a + b$ with $a \in I$ and $b \in J$) is also an ideal. Hence, if I and J are distinct maximal ideals, we must have $I + J = R$. Suppose that $I_j \neq I_{n+1}$ for $j = 1, \ldots, n$. Then, for each $j = 1, \ldots, n$, we can find $a_j \in I_j$ and $b_j \in I_{n+1}$ such that $1 = a_j + b_j$. Hence, $1 = \Pi 1 = \Pi(a_j + b_j)$. However, $\Pi(a_j + b_j) = a_1 \cdots a_n +$ sums of products each of which contains at least some b as a factor. Because I_{n+1} is an ideal, each such product is in I_{n+1}, so that the sum of those products is again in I_{n+1}. Also, because each I_k is an ideal, the term $a_1 \cdots a_n$ is in $I_1 \cap \cdots \cap I_n$ and thus in I_{n+1}. Hence, $1 = \Pi(a_j + b_j)$ is in I_{n+1}, which is a contradiction. Therefore, we have $I_j = I_{n+1}$ for some j, contradicting the fact that all the I_n were distinct. Thus, the set of maximal ideals of R is finite.

Because we have shown that for any two ideals I and J we have $I + J = R$, we may apply the Chinese Remainder Theorem (see the exercises in Chapter 6) to the set of all maximal ideals $I_1, \ldots, I_s$ and obtain the isomorphism of rings $k : R/I_1 \cap \cdots \cap I_s \to R/I_1 \times \cdots \times R/I_s$ induced by the surjective ring morphism $\bar{k} : R \to R/I_1 \times \cdots \times R/I_s$.

We now know that if R is left artinian and $\{I_1, \ldots, I_s\}$ is the set of all maximal ideals of R, then $R/I_1 \cap \cdots \cap I_s$ is semisimple. Thus, if $I_1 \cap \cdots \cap I_s = 0$, R itself is semisimple. Now we may ask if $I_1 \cap \cdots \cap I_s$ is the smallest ideal J such that R/J is semisimple. As a preliminary step in answering this question, we show that if R is a semisimple ring, then $\cap I_k = 0$, where $\{I_k\}$ is the set of all maximal ideals of R.

If R is semisimple, we know that $J^2 \neq 0$ for every minimal left ideal J of R. If $\{I_1, \ldots, I_s\}$ is the set of maximal ideals of R and if $I_1 \cap \cdots \cap I_s \neq 0$, then $I_1 \cap \cdots \cap I_s$ contains some minimal left ideal J. Because J is a simple R-module, ann(J) is one of the ideals I_k. Hence, $I_k J = 0$. However, because $J \subset I_1 \cap \cdots \cap I_s$, J is in I_k, so $J^2 \subset I_k J = 0$ and $J^2 = 0$. This contradicts Proposition 1.3, and so $I_1 \cap \cdots \cap I_s = 0$.

Definition
Let R be an artinian ring and $I_1, \ldots, I_s$ the set of all maximal ideals of R. The **radical of R** [written rad(R)] is the ideal $I_1 \cap \cdots \cap I_s$ of R.

In this terminology our preceding discussion gives the following.

Theorem 2.3
A ring R is semisimple if and only if it is left artinian and rad$(R) = 0$.

Corollary 2.4
If R is a left artinian ring and J is an ideal of R such that R/J is semisimple, then $J = I_1 \cap \cdots \cap I_t$ where $\{I_k\}$ is the set of all maximal ideals of R which contain J. Thus, rad(R) is the smallest ideal J such that R/J is semisimple.

PROOF: Because there is a bijective map between ideals of R/J and ideals of R containing J, the maximal ideals of R/J correspond to the maximal ideals $I_1, \ldots, I_t$ of R which contain J. If R/J is semisimple, then rad$(R/J) = 0$ so that

$(I_1/J) \cap \cdots \cap (I_t/J) = 0$. Because $(I_1/J) \cap \cdots \cap (I_t/J) = (I_1 \cap \cdots \cap I_t)/J$, we have $I_1 \cap \cdots \cap I_t = J$.

We have now answered the questions we posed by giving a description of the smallest ideal J of R such that R/J is semisimple, namely, the radical. We saw in Example 1.7 that the radical might have something to do with nilpotent elements, nilpotent left ideals, or nil left ideals. We make this connection more specific in the following.

Proposition 2.5
If R is a left artinian ring, then $\text{rad}(R)$ is nilpotent.

The proof of this proposition uses the following description of the radical of a left artinian ring.

Proposition 2.6
The radical of a left artinian ring R is the intersection of all the maximal left ideals of R.

PROOF: If I is a maximal ideal of R, then R/I is a simple left artinian ring. As we saw in Example 2.8 in Chapter 7, the intersection of all maximal left ideals of R/I is zero. Hence, I is the intersection of all maximal left ideals containing I. Consequently, the radical of R, being the intersection of all maximal ideals of R, equals the intersection of all those maximal left ideals of R which contain some maximal ideal of R. However, every maximal left ideal J of R does contain some maximal ideal I. For R/J is a simple R-module and thus $\text{ann}(R/J)$ is a maximal ideal I which is contained in J. Thus, the intersection of the maximal ideals of R, which is the radical of R, is contained in the intersection of the maximal left ideals of R. Hence, $\text{rad}(R)$ is the intersection of the maximal left ideals of R.

We now return to the proof of Proposition 2.5.

PROOF: Let $I = \text{rad}(R)$ and consider the decreasing chain of ideals $I \supset I^2 \supset I^3 \supset \cdots$. Because R is left artinian, there is a positive integer n such that $I^n = I^{n+k}$ for all $k \geq 0$. We want to show that $I^n = 0$. Let $J = I^n$ and suppose $J \neq 0$. Then $J^2 = I^{n+n} = I^n = J \neq 0$. Thus, $J^2 \neq 0$ and $J^2 = J$. Among the nonempty set of all left ideals $J' \subset J$ such that $JJ' \neq 0$ there is a minimal one, J_0'. Then, because $JJ_0' \neq 0$, there is an element $b \in J_0'$ such that $Jb \neq 0$. Jb is clearly a left ideal of R contained in J_0', and $J(Jb) = J^2b = Jb \neq 0$. Thus, $Jb = J_0'$. Because $Jb \subset Rb \subset J_0' = Jb$, we have $J_0' = Rb$, that is, J_0' is the left ideal generated by the element b. Also, from the fact that $J(Jb) = Jb$, we see that $JJ_0' = J_0'$, because $J_0' = Jb = Rb$. Let us conclude from this that $J_0' = 0$.

The element b is in $J_0' = Jb$ so that $b = ab$ where $a \in J$. Therefore, $(1-a)b = 0$. If we show that the left ideal generated by $(1-a)$ is R, then $v(1-a) = 1$ for some $v \in R$ so that $0 = v \cdot 0 = v((1-a)b) = (v(1-a))b = 1 \cdot b = b$. Hence, b and J_0 are both zero.

Why then does $(1-a)$ generate all of R as a left ideal? If it did not, then $(1-a)$ would be contained in some maximal left ideal. But a is contained in every maximal left ideal of R because $a \in J \subset \text{rad}(R)$ and $\text{rad}(R)$ is the intersection of

all the maximal left ideals of R. Thus, $(1-a)$ cannot be contained in any maximal left ideal, because otherwise, 1 would be contained in that ideal. Hence, $(1-a)$ generates R as a left ideal, and our proof is complete.

From the fact that the radical of a left artinian ring is nilpotent, we get the following.

Proposition 2.7
A left artinian ring is left noetherian.

PROOF: We know that $R/\text{rad}(R)$ is semisimple and therefore left noetherian. Hence, if we show that $\text{rad}(R)$ is a left noetherian R-module, we will be done, because $0\to\text{rad}(R)\to R\to R/\text{rad}(R)\to 0$ is an exact sequence of left R-modules with both $\text{rad}(R)$ and $R/\text{rad}(R)$ noetherian modules.

Let $I=\text{rad}(R)$, and consider the chain $I\supset I^2\supset\cdots\supset I^n\supset I^{n+1}=0$. Each of the ideals I^k is artinian and, therefore, so is each factor module I^k/I^{k+1}. In addition, each of the R-modules I^k/I^{k+1} is annihilated by I, so that I^k/I^{k+1} is an R/I-module and artinian. If we can show that I^k/I^{k+1} is a noetherian R/I-module, then it will also be a noetherian R-module. (Why?) Finally, if we show that I^k/I^{k+1} is noetherian for each k, we can show that I itself is noetherian. This is so because, first of all, I^{n+1} is noetherian (because it is zero). Now suppose I^k is noetherian, and let us show that I^{k-1} is noetherian. We have the exact sequence $0\to I^k\to I^{k-1}\to I^{k-1}/I^k\to 0$. Because I^k and I^{k-1}/I^k are noetherian, I^{k-1} is noetherian, and so, by induction, we finally get I is noetherian.

How, then, do we see that I^k/I^{k+1} is noetherian? Because we have seen that I^k/I^{k+1} is an artinian R/I-module, and because all R/I-modules are semisimple, we have that I^k/I^{k+1} is a semisimple artinian module. If we prove that all such modules are noetherian, we will be done.

Lemma 2.8
Let M be an artinian semisimple module over an arbitrary ring. Then M is a noetherian module.

PROOF: Because M is a semisimple module, $M=\coprod_{\alpha\in A} M_\alpha$ where each M_α is a simple R-module. We leave it to the reader to show that A must be a finite set because M is assumed to be an artinian module. However, we know that simple modules are noetherian, so that M, being a finite sum of noetherian modules, is noetherian.

This lemma shows that I^k/I^{k+1} is noetherian for each k and so the proof of Proposition 2.7 is complete.

3. THE RADICAL OF AN ARBITRARY RING

We have discussed the radical of an artinian ring but not of an arbitrary ring. We could define the radical of an arbitrary ring to be the intersection of its maximal ideals. However, we know in the artinian case that this is the same as the

intersection of its maximal left ideals and it was this latter property of the radical that played a crucial role in the proof of Proposition 2.5.

Definition
Let R be a ring. The **radical of R** [written rad(R)] is the intersection of the maximal left ideals of R.

Proposition 3.1
Let R be any ring and let I be a left ideal of R. Then the following statements are equivalent:

(a) $1+a$ has a left inverse for all a in I; that is, there is a b in R such that $b(1+a)=1$.
(b) If M is a finitely generated left R-module and $IM=M$, then $M=0$.
(c) I is contained in the radical of R.

PROOF: (a) implies (b). Assume that M is finitely generated and that $IM=M$. Let $m_1,\ldots,m_n$ be a set of generators of M. Then, because $IM=M$, we have $m_i=\sum_{j=1}^{s} a_{ij}m'_j$ with $a_{ij}\in I$ and m'_j in M. For any $m\in M$, we have $m=\sum_{i=1}^{n} b_i m_i$ with $b_i\in R$, and so for any m in M we have $m=\sum_{ij} b_i a_{ij} m'_j=\sum_j(\sum_i b_i a_{ij})m'_j$. Notice that $\sum_i b_i a_{ij}\in I$ for each j so that M is finitely generated over I in the sense that there is a finite set $\{m'_1,\ldots,m'_s\}$ of elements of M such that every element of M may be written as a linear combination of the $m'_1,\ldots,m'_s$ with coefficients in I.

Let $\{m'_1,\ldots,m'_s\}$ be a minimal set of generators of M over I (that is, no subset of $\{m'_1,\ldots,m'_s\}$ generates M over I). If M were not zero, then s would have to be greater than zero. Assume $M\neq 0$. Because $IM=M$, $m'_1=\sum_{i=1}^{s} a_i m'_i$ with $a_i\in I$ so that we get $(1-a_1)m'_1=\sum_{i=2}^{s} a_i m'_i$. But because $1-a_1$ has a left inverse, we have, multiplying through by this left inverse, that $m'_1=\sum_{i=2}^{s} b_i m'_i$ with $b_i\in I$. Thus, $\{m'_2,\ldots,m'_s\}$ generates M over I and this contradiction proves that M must be zero.

(b) implies (c). Let J be a maximal left ideal of R, and suppose that I is not contained in J. Letting $M=R/J$, we then have $IM\neq 0$. However, because J is maximal, M is simple and so $IM=M$. But M is generated by one element and so by (b) we have $M=0$, which is a contradiction. Hence, $I\subset J$ for every maximal left ideal, and thus $I\subset\text{rad}(R)$.

(c) implies (a). Because $I\subset\text{rad}(R)$, we have, for all $a\in I$, that $a\in\text{rad}(R)$ so that $1+a$ is not in any maximal left ideal of R. Thus, the left ideal generated by $1+a$ is all of R and so there is a $b\in R$ such that $b(1+a)=1$.

As a result of this proposition we see that if I is a left ideal contained in rad(R), then $IM=0$ for every simple R-module M. In fact, rad(R) is the intersection of the annihilators of all simple R-modules. Hence, rad(R) is an ideal, not just a left ideal, of R.

The reader should also observe that if $1+a$ has a left inverse for all $a\in I$, then $1+a$ also has a right inverse for all $a\in I$. To see this, let $v(1+a)=1$. Because $va\in I$, $-va$ is also in I and hence $1-va$ has a left inverse, say $w(1-va)=1$. But, because $v(1+a)=1$, we have $v+va=1$ or $v=1-va$. Thus, $wv=1$ and therefore $1+a=(wv)(1+a)=w(v(1+a))=w$ from which we get $(1+a)v=wv=1$ and $1+a$

has v as a right inverse. It is important to note that we made use of the fact that *every* element b of I (or at least every left multiple of a) had the property that $1+b$ has a left inverse in order to prove this.

On the basis of these remarks, we leave it to the reader to prove the following.

Proposition 3.2
Let R be a ring. Then rad(R) is an ideal which is equal to the intersection of all maximal right ideals as well as the intersection of all maximal left ideals of R.

Proposition 3.3
If R is a ring and I is a left (right) nil ideal of R, then I is contained in rad(R).

PROOF: If a is any nilpotent element of R, say $a^n = 0$, then the identity $1 - a^n = (1 - a + a^2 \pm \cdots \pm a^{n-1})(1 + a)$ shows that $1 + a$ has a left inverse. Thus, if I is nil ideal, we have that $1 + a$ has a left inverse for every $a \in I$ and so, by Proposition 3.1, $I \subset \text{rad}(R)$.

The reader might suspect, as a result of Proposition 3.3, that the radical of every ring is a nil ideal. For artinian rings we know it is even nilpotent. Let us look at the following.

Example 3.4 Let R be the subring of the rational numbers consisting of those fractions a/b for which b is not divisible by 2. Obviously, this is the ring of quotients $\mathbf{Z}_S$ where S is the multiplicative subset of $\mathbf{Z}$ consisting of all odd integers. Because R is a commutative ring, we need not worry about distinctions between left and right ideals. Let us try to find all the maximal ideals of R. We leave it to the reader to check that there is only one, namely the subset I or R consisting of those elements a/b such that a is divisible by 2. Hence, the radical of R is just I itself. Because R is an integral domain, I contains no nilpotent element other than zero, so I is certainly not a nil ideal, let alone nilpotent.

Lest the reader think that all nil ideals are nilpotent (they are, of course, in an artinian ring), we suggest that he work out the following.

Example 3.5 Let $S = k[X_1, X_2, \ldots]$ be the polynomial ring over a field k in a denumerable number of indeterminates $X_1, X_2, \ldots$. For each n, because $k[X_1, \ldots, X_n] \subset k[X_1, \ldots, X_{n+1}]$, we have $S = \bigcup_n k[X_1, \ldots, X_n]$. Let X be the ideal in S generated by $\{X_n^{n+1}\}_{n \in \mathbf{N} - \{0\}}$, and let $R = S/X$. Denote by $\bar{X}_i$ the residue class of X_i modulo X, and let I be the ideal in R generated by all the $\bar{X}_i$. I is a nil ideal. To see this, the reader should show, using the generalization of the binomial theorem, that if $a_1, \ldots, a_n$ are nilpotent elements in a commutative ring, then any linear combination of them is nilpotent. Because R is commutative and $\bar{X}_i^{i+1} = 0$, I is a nil ideal. But I is not a nilpotent ideal. The idea behind this is that if $I^n = 0$ for some n, then $\bar{X}_n^n = 0$. But all we are given is that $\bar{X}_n^{n+1} = 0$. The reader should carry out these arguments in sufficient detail to convince himself that this nil ideal is not nilpotent.

The radical of a ring has some very useful properties. We list some of these in the following.

Basic Properties 3.6

Let R be a ring and let $J = \text{rad}(R)$. Then:

(a) If M is a finitely generated R-module such that $M/JM = (0)$, then $M = (0)$.

(b) If M is a finitely generated R-module and $m_1, \ldots, m_s$ are elements of M, then $\{m_1, \ldots, m_s\}$ generates M if and only if $\{\bar{m}_1, \ldots, \bar{m}_s\}$ generates M/JM as an R/J-module, where $\bar{m}_i$ denotes the residue class of m_i in M/JM. Thus, if R/J is a division ring, any two minimal sets of generators of M have the same number of elements, and a minimal set of generators of M may be selected from any given set of generators. Hence:

(c) If M is a finitely generated R-module and $f: X \to M$ is a morphism of R-modules such that the composition $X \xrightarrow{f} M \xrightarrow{k} M/JM$ is surjective, then f is surjective. In particular, if M' is a submodule of M such that $M' + JM = M$, then $M' = M$.

(d) If M and N are finitely generated projective R-modules such that M/JM and N/JN are isomorphic as R/J-modules, then M and N are isomorphic. In particular, if R/J is a division ring, every finitely generated projective R-module is free.

PROOF: (a) Left to the reader.

(b) Clearly, if $\{m_1, \ldots, m_s\}$ generates M, then $\{\bar{m}_1, \ldots, \bar{m}_s\}$ generates M/JM as an R-module and hence as an R/J-module.

Suppose that $\{\bar{m}_1, \ldots, \bar{m}_s\}$ generates M/JM as an R/J-module and hence as an R-module. Let M' be the submodule of M generated by $\{m_1, \ldots, m_s\}$ and let $M'' = M/M'$. Then $JM'' = (JM + M')/M'$ where $JM + M'$ is the submodule of M generated by JM and M'. Because $\{\bar{m}_1, \ldots, \bar{m}_s\}$ generates M/JM, it follows that $M = M' + JM$. Thus, $M/(M' + JM) = 0$. But $M/(M' + JM) \approx M/M'/(JM + M')/M' = M''/JM''$. Therefore, $M''/JM'' = 0$. The fact that M is a finitely generated R-module implies that M'' is also finitely generated. Therefore, by (a), $M'' = 0$, and thus, $M = M'$.

(c) Follows readily from (b).

(d) Let $f: M/JM \to N/JN$ be an isomorphism, where M and N are finitely generated projective R-modules, and let $k_1: M \to M/JM$ and $k_2: N \to N/JN$ be the canonical surjective morphisms. Because M is R-projective, there is a morphism $g: M \to N$ such that $k_2 g = f k_1$. The composition $f k_1$ is surjective because f and k_1 are surjective. Hence, by (c), the morphism $g: M \to N$ is surjective.

Because N is projective, the morphism $g: M \to N$ is a splittable epimorphism. Let $t: N \to M$ be a splitting for g, that is, $gt = \text{id}_N$. Because t is injective, it suffices to show that t is surjective in order to conclude that t is an isomorphism.

Because $t(JN) \subset JM$, we know that $t: N \to M$ induces a unique morphism $\bar{t}: N/JN \to M/JM$ such that $\bar{t} k_2 = k_1 t$. It follows that $f\bar{t} = \text{id}_{N/JN}$ because $f\bar{t} k_2 = f k_1 t = k_2 g t = k_2$ and k_2 is surjective. The fact that f is an isomorphism implies that $\bar{t} = f^{-1}$. Hence, $\bar{t}$ is, in particular, surjective. Thus, $\bar{t} k_2$ is surjective. Because $\bar{t} k_2 = k_1 t$, it follows from (c) that t is surjective. Hence, $t: N \to M$ is an isomorphism, which finishes the first part of (d).

To see the second part of (d), observe that if R/J is a division ring, then any two finite-dimensional vector spaces over R/J are isomorphic if and only if they have the same dimension. Now suppose P is a finitely generated projective

R-module. Then P/JP is a finitely generated R/J-module and hence is an R/J-vector space of dimension n. Let F be a free R-module having a basis X with $\text{card}(X) = n$. It is easy to see that $k(X)$ is a basis for F/JF as an R/J-module, where $k : F \to F/JF$ is the canonical morphism. Hence, F/JF is an n-dimensional vector space over R/J and is therefore isomorphic to P/JP. Thus, P and F are isomorphic, and this completes the proof of (d).

EXERCISES

(1) Show that the subring of $M_2(\mathbf{Q})$, where $\mathbf{Q}$ is the field of rational numbers, consisting of all $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ with a an integer, is right noetherian but not left noetherian.

(2) Show that the subring of $M_2(\mathbf{R})$, where $\mathbf{R}$ is the field of real numbers, consisting of all $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ with a in $\mathbf{Q}$, is right artinian but not left artinian.

(3) Let J be the radical of the left artin ring R. Suppose $0 \to K \to P \to M \to 0$ is an exact sequence of finitely generated R-modules with P a projective R-module. Then M is projective if and only if $0 \to K/JK \to P/JP \to M/JM \to 0$ is exact. [*Hint*:

(a) Show that $0 \to K/JK \to P/JP$ is exact if and only if $\text{Hom}_R(P, S) \to \text{Hom}_R(K, S) \to 0$ is exact for all finitely generated semisimple modules S.

(b) Show that $\text{Hom}_R(P, X) \to \text{Hom}_R(K, X) \to 0$ is exact for all finitely generated R-modules X, if and only if it is exact for all X which are semisimple R-modules.]

(4) Let J be the radical of the left artin ring R, M a finitely generated R-module, and $k : M \to M/JM$ the canonical surjective map. Suppose N is a submodule of M such that $k|N : N \to M/JM$ is not the zero morphism, or equivalently, $N \not\subset JM$. Then given any submodule L of $k(N)$, there is a submodule N' of N such that the induced morphism $N'/JN' \to M/JM$ is a monomorphism whose image is precisely L. [*Hint*: Show that a minimal element N' of the nonempty set of all submodules X of N with the property $k(X) = L$ has the property $N'/JN' \to M/JM$ is a monomorphism.]

(5) Let J be the radical of a left artin ring R and P a finitely generated projective R-module. Let K be a submodule of P with $K \not\subset JP$ and L an arbitrary submodule of the image of K in P/JP. Show that there is a submodule K' of K having the following properties:

(a) The image of K' in P/JP is L.

(b) K' is a summand of P and hence of K.

(c) K' is a projective R-module.

(6) Let J be the radical of a left artin ring R. For an epimorphism $A \xrightarrow{f} B$ of finitely generated R-modules, the following statements are equivalent:

(a) A morphism $g : X \to A$ is an epimorphism if the composition $X \xrightarrow{g} A \xrightarrow{f} B$ is an epimorphism.

(b) $\text{Ker } f \subset JA$, or equivalently, the epimorphism $A/JA \to B/JB$ induced by $f : A \to B$ is an isomorphism.

(7) Let R be an arbitrary ring. An epimorphism $f : Y \to Z$ or arbitrary R-modules is called an **essential epimorphism** if it has the property that a morphism $g : X \to Y$ is

an epimorphism if the composition $X \xrightarrow{g} Y \xrightarrow{f} Z$ is an epimorphism. Show that if $f: A \to B$ and $g: B \to C$ are epimorphisms of R-modules, then gf is an essential epimorphism if and only if both f and g are essential epimorphisms.

(8) Let J be the radical of a left artin ring R and M a finitely generated R-module.

(a) Show that there is an essential epimorphism $f: P \to M$ with P a finitely generated projective R-module. Such an essential epimorphism is called a **projective cover** of M.

(b) Suppose $f: P \to M$ and $f': P' \to M$ are two projective covers of M. Then there is a morphism $h: P \to P'$ such that $f'h = f$ and any such h is an isomorphism. Hence, the projective covers of finitely generated R-modules are essentially unique.

(c) If P is a finitely generated R-module and $f: P \to M$ is an epimorphism, then f is a projective cover for M if and only if the composition $P \to M \to M/JM$ is a projective cover for M/JM.

(9) Let J be the radical of a left artin ring R.

(a) If P is a finitely generated projective R-module, then the canonical epimorphism $k: P \to P/JP$ is a projective cover for P/JP.

(b) Two finitely generated projective R-modules P and P' are isomorphic if and only if P/JP and P'/JP' are isomorphic.

(c) Let $\{f_i: A_i \to B_i\}_{i \in I}$ be a finite family of essential epimorphisms of finitely generated R-modules. Then the morphism $\coprod_{i \in I} f_i : \coprod_{i \in I} A_i \to \coprod_{i \in I} B_i$ given by $\coprod f_i : \{a_i\}_{i \in I} = \{f_i(a_i)\}_{i \in I}$ is an essential epimorphism.

(d) Let P be a finitely generated projective R-module. Then $P/JP \approx \coprod_{i \in I} S_i$ where the $\{S_i\}_{i \in I}$ are a finite family of simple R-modules. Suppose $f_i: P_i \to S_i$ are projective covers for the S_i. Then $P \approx \coprod_{i \in I} P_i$ where each projective R-module P_i has the property that P_i/JP_i is simple.

(e) If $P \approx \coprod_{i \in I} P_i$ and $P \approx \coprod_{k \in K} P'_k$ where the P_i, P'_j are projective R-modules such that all the P_i/JP_i and P'_k/JP'_k are simple R-modules, then there is an isomorphism of sets $\sigma: I \to K$ such that $P_i \approx P'_{\sigma(i)}$ for all i in I.

Definition

Let R be an arbitrary ring. An R-module M is said to be **indecomposable** if (0) and M are the only submodules of M which have complements in M.

(10) Show that every noetherian M is the finite sum of indecomposable submodules of M.

(11) Let J be the radical of a left artin ring R.

(a) Show that every finitely generated R-module is isomorphic to a finite sum of indecomposable R-modules.

(b) Show that a finitely generated projective R-module P is indecomposable if and only if P/JP is a simple R-module.

(c) Let $\mathcal{P}(R)$ be the collection of finitely generated indecomposable projective modules and let $\mathcal{S}(R)$ be the collection of simple R-modules. Show that the map $\mathcal{P}(R) \to \mathcal{S}(R)$ given by $P \to P/JP$ for all P in $\mathcal{P}(R)$ induces an isomorphism between the isomorphism classes of indecomposable projective

R-modules and the isomorphism classes of simple R-modules. Hence, the number of nonisomorphic indecomposable finitely generated projective R-modules is the same as the number of nonisomorphic simple R-modules, which is finite.

(d) If P is a finitely generated indecomposable projective module, then there is an idempotent element e in R such that P is isomorphic to the left ideal Re of R.

(12) Let K be a field and $T_n(K)$ the ring of $n \times n$ triangular matrices over K, that is, $T_n(K)$ is the subring of $M_n(K)$ consisting of all matrices (a_{ij}) with $a_{ij} = 0$ if $j > i$.

(a) Show that $T_n(K)$ is an artin ring.

(b) Show that the radical J of $T_n(K)$ consists of all matrices (a_{ij}) in $T_n(K)$ with $a_{ii} = 0$ for all $i = 1, \ldots, n$. Also, $J^n = 0$ and $J^{n-1} \neq (0)$.

(c) Show that the ring $T_n(K)/J \approx \prod_{i=1}^{n} K_i$ where each $K_i = K$ and hence $T_n(K)$ has precisely n nonisomorphic simple modules.

(d) For each $k = 1, \ldots, n$ define I_k to be the subset of $T_n(K)$ consisting of all (a_{ij}) with $a_{ij} = 0$ if $j \neq k$.

(i) Each I_k is a left ideal of $T_n(K)$.

(ii) Each I_k is an indecomposable projective $T_n(K)$-module.

(iii) If $I_k \approx I_{k'}$, then $k = k'$.

(iv) If P is an indecomposable projective $T_n(K)$-module, then $P \approx I_k$ for some k. Further, each nonzero submodule of P is an indecomposable projective module.

(e) Every submodule of a finitely generated projective $T_n(K)$-module is a projective module.

(f) Show that the subset I of $T_n(K)$ consisting of all (a_{ij}) with $a_{ij} = 0$ if $i < n$ is an ideal in $T_n(K)$.

(i) Show that as a left module the ideal I is the sum of n copies of a projective simple $T_n(K)$-module.

(ii) As a right module, I is an indecomposable projective $T_n(K)$-module.

(iii) $T_n(K)/I \approx T_{n-1}(K)$ for each n.

(g) Show that each indecomposable finitely generated projective $T_n(K)$-module has precisely one composition series.

(h) Show that $l(I_k) = (n-k)+1$.

(i) Because $T_n(K)$ is a subring of $M_n(K)$, we can view $M_n(K)$ as a $T_n(K)$-module. Show that $M_n(K)$ is a projective $T_n(K)$-module. Also, find the indecomposable projective modules $P_1, \ldots, P_t$ such that $M_n(K) \approx \coprod_{i=1}^{t} P_i$.

(j) Show that the map $\phi: T_n(K) \to T_n(K)^{op}$ given by $\phi((a_{ij})) = (a_{ji})$ is an isomorphism of rings.

(k) Show that the categories $\mathrm{Mod}(T_n(K))$ and $\mathrm{Mod}(T_n(K)^{op})$ are equivalent categories.

(13) Let K be a field. Let $L_n(K)$ be the subring of $T_n(K)$ consisting of all matrices (a_{ij}) in $T_n(K)$ such that $a_{11} = a_{22} = \cdots = a_{nn}$.

(a) The radical J of $L_n(K)$ consists of all matrices (a_{ij}) with $a_{ii} = 0$ for all $i = 1, \ldots, n$.

(b) Show that the rings K and $L_n(K)/J$ are isomorphic.

(c) All simple $L_n(K)$-modules are isomorphic.

(d) $L_n(K)$ is the only indecomposable, finitely generated projective $L_n(K)$-module.
(e) All finitely generated projective $L_n(K)$-modules are free modules.
(f) If K is a submodule of a free $L_n(K)$-module F and $K \subset JF$, then K has no projective submodules.
(g) Because $L_n(K)$ is a subring of $T_n(K)$, we can view $T_n(K)$ as an $L_n(K)$-module. Is $T_n(K)$ a projective $L_n(K)$-module?
(h) Because $L_n(K)$ is a subring of $M_n(K)$, we can view $M_n(K)$ as an $L_n(K)$-module. Is $M_n(K)$ a projective $L_n(K)$-module?
(i) Show that $L_n(K)$ and $L_n(K)^{op}$ are isomorphic rings.

(14) Let K be a field and $V_n(K)$ the subset of $T_n(K)$ consisting of all (a_{ij}) such that $a_{ij} = 0$ unless either $i = j$ or $j = 1$.
(a) Show that $V_n(K)$ is a subring of $T_n(K)$.
(b) Show that the radical J of $V_n(K)$ consists of all (a_{ij}) satisfying $a_{ii} = 0$ for all $i = 1, \ldots, n$.
(c) Show that $J^2 = 0$.
(d) Show that the rings $V_n(K)/J$ and $\prod_{i=1}^{n} K_i$ (where each $K_i = K$) are isomorphic.
(e) Show that $V_n(K)$ has precisely n nonisomorphic simple modules.
(f) Find n idempotent elements $e_1, \ldots, e_n$ such that the projective $V_n(K)$-modules $V_n(K)e_i$ are indecomposable with no two distinct ones isomorphic.
(g) Show that every submodule of a finitely generated indecomposable $V_n(K)$-module is a projective $V_n(K)$-module.
(h) Show that J is a projective $V_n(K)$-module and find the indecomposable projective $V_n(K)$-modules whose sum is J.
(i) Show that every submodule of a finitely generated $V_n(K)$-module is a projective $V_n(K)$-module.
(j) Because $V_n(K)$ is a subring of $T_n(K)$ and $M_n(K)$, we can view $T_n(K)$ and $M_n(K)$ as $V_n(K)$-modules. Is either of them a projective $V_n(K)$-module?

(15) Same hypotheses and notation as in Exercise 14.
(a) Show that $V_n(K)$ has precisely n nonisomorphic simple right modules.
(b) Let $e_1, \ldots, e_n$ be the idempotents you obtained in Exercise 14(f). Show that the right ideals $e_iV_n(K)$ are indecomposable projective $V_n(K)$-modules such that $e_iV_n(K) \approx e_jV_n(K)$ if and only if $i = j$.
(c) Show that the submodules of the finitely generated indecomposable right $V_n(K)$-modules are projective right $V_n(K)$-modules.
(d) Every submodule of a finitely generated projective right $V_n(K)$-module is a projective $V_n(K)$-module.
(e) Show that J is a projective right $V_n(K)$-module and find the indecomposable projective right $V_n(K)$-modules whose sum is J.
(f) Are $T_n(K)$ and $M_n(K)$ projective as right $V_n(K)$-modules?
(g) Are $V_n(K)$ and $V_n(K)^{op}$ isomorphic rings?

(16) Let R be a commutative ring and M a monoid. Define the map $\epsilon : R[M] \to R$ by $\epsilon(\Sigma r_i m_i) = \Sigma r_i$.
(a) Prove that ϵ is an R-algebra morphism which enables us to view R as an $R[M]$-module.
(b) Prove that $I = \operatorname{Ker} \epsilon$ is generated as an R-module by $\{1 - m\}_{m \in M}$.

(c) For each $R[M]$-module A we denote by A^M the $R[M]$-submodule of A consisting of all a in A such that $ma = a$ for all m in M. Show that the map $\mathrm{Hom}_{R[M]}(R, A) \to A$ given by $f \to f(1)$ is an injective R-morphism whose image is A^M. Thus, we obtain an isomorphism $\mathrm{Hom}_{R[M]}(R, A) \to A^M$ which we usually view as an identification of R-modules. Consequently, we have:

(d) If $0 \to A_1 \to A_2 \to A_3$ is an exact sequence of $R[M]$-modules, then $0 \to A_1^M \to A_2^M \to A_3^M$ is an exact sequence of R-modules.

(e) R is a projective $R[M]$-module if and only if given any exact sequence $0 \to A_1 \to A_2 \to A_3 \to 0$ of $R[M]$-modules, the sequence of R-modules $0 \to A_1^M \to A_2^M \to A_3^M \to 0$ is exact.

(17) Let M be a monoid. An M-module is an abelian group A together with a map $M \times A \to A$ which we denote by $(m, a) \to ma$ satisfying

(i) $m(a_1 + a_2) = ma_1 + ma_2$,

(ii) $(m_1 m_2)a = m_1(m_2 a)$,

(iii) $1a = a$,

for all m, m_1, m_2 in M and a, a_1, a_2 in A. Show that the following data define a category called the category of M-modules and denoted by $\mathrm{Mod}(M)$.

(a) $\mathrm{Ob}\,\mathrm{Mod}(M)$ are the M-modules.

(b) (A_1, A_2) is the set of all morphisms of abelian groups $f: A_1 \to A_2$ such that $f(ma) = mf(a)$ for all m in M.

(c) $(A_1, A_2) \times (A_2, A_3) \to (A_1, A_3)$ is given by ordinary composition of morphisms of abelian groups.

(18) Let R be a commutative ring and M a monoid. Show that the category of $R[M]$-modules is isomorphic to the category $\mathscr{C}$ defined by the following data:

(a) An object of $\mathscr{C}$ is a pair (A, f) where A is an R-module and $f: M \times A \to A$ is an M-module structure on A such that $m(ra) = r(ma)$ for all m in M and r in R.

(b) A morphism $(A, f) \to (A', f')$ is an R-module map $g: A \to A'$ such that $g(ma) = mg(a)$ for all m in M.

(c) Composition of morphisms is the ordinary composition of R-module morphisms.

[*Hint*: Show that there is a natural functor $F: \mathrm{Mod}(R[M]) \to \mathscr{C}$ with the property $F(A) = (A, f)$ where A is the R-module given by the fact that R is a subring of $R[M]$ and $f: M \times A \to A$ is given by $f(m, a) = ma$. Show that there is a natural functor $G: \mathscr{C} \to \mathrm{Mod}(R)$ with the property $G(A, f)$ is the $R[M]$-module whose underlying abelian group is that of A and where the operation $R[M] \times A \to A$ is given by $(\Sigma_{i \in I} r_i m_i, a) = \Sigma_{i \in I} r_i(m_i a)$. Then GF and FG are identity functors on $\mathrm{Mod}(R[M])$ and $\mathscr{C}$, respectively. The categories $\mathrm{Mod}(R[M])$ and $\mathscr{C}$ are usually identified by means of the functors F and G.]

(19) Let G be a group and R a commutative ring. Let A and B be $R[G]$-modules.

(a) If $f: A \to B$ is an R-morphism and g is in G, then the map $(gf): A \to B$ defined by $(gf)(a) = g(f(g^{-1}(a))$ is also an R-morphism.

(b) The map $G \times \mathrm{Hom}_R(A, B) \to \mathrm{Hom}_R(A, B)$ given by $(g, f) \to gf$ makes $\mathrm{Hom}_R(A, B)$ a G-module.

(c) Show that $g(rf) = r(gf)$ for all g in G, r in R, and f in $\mathrm{Hom}_R(A, B)$. Hence, $\mathrm{Hom}_R(A, B)$ is an $R[G]$-module.

(d) $\mathrm{Hom}_R(A, B)^G = \mathrm{Hom}_{R[G]}(A, B)$.

(20) Let K be a field and G a group. Then:

(a) $K[G]$ is a semisimple ring if and only if K is a projective $K[G]$-module. [*Hint*: Show that all $K[G]$-modules are projective if and only if K is a projective $K[G]$-module.]
(b) Show that $(K[G])^G$ is the set of all $\Sigma r_i g_i$ in $K[G]$ such that $r_i = r_j$ for all i and j.
(c) Show that K is a projective $K[G]$-module if and only if G is a finite group and the characteristic of K does not divide the order of G.
(d) $K[G]$ is a semisimple ring if and only if G is a finite group and the characteristic of K does not divide the order of G.

(21) Suppose J is the radical of an artin ring R. Show that if e is an idempotent element of R/J, there is an idempotent element e' in R such that $k(e') = e$ where $k: R \to R/J$ is the canonical surjective ring morphism. [*Hint*: Use the fact that every finitely generated module has a projective cover.]

Definition

Let $\mathscr{C}$ and $\mathscr{D}$ be categories. A duality between $\mathscr{C}$ and $\mathscr{D}$ is a contravariant functor $F: \mathscr{C} \to \mathscr{D}$ satisfying:

(a) If D is an object of $\mathscr{D}$, then $D \approx F(C)$ for some C in $\mathscr{C}$.
(b) For each pair of objects C_1 and C_2 in $\mathscr{C}$, the maps $F: (C_1, C_2) \to (F(C_2), F(C_1))$ are isomorphisms.

(22) Let $F: \mathscr{C} \to \mathscr{D}$ be a contravariant functor. F is a duality if and only if there is a contravariant functor $G: \mathscr{D} \to \mathscr{C}$ such that $GF \approx \mathrm{id}_{\mathscr{C}}$ and $FG \approx \mathrm{id}_{\mathscr{D}}$.

(23) Let K be a field and $\mathscr{C}$ the category of finite-dimensional vector spaces over K. Show that the functor $(\ , K): \mathscr{C} \to \mathscr{C}$ given by $X \to \mathrm{Hom}_K(X, K)$ is a duality. [*Hint*: Show that the usual K-morphisms $\psi_X: X \to \mathrm{Hom}_K(\mathrm{Hom}_K(X, K), K)$ given by $\psi_X(x)(f) = f(x)$ for all x in X and f in $\mathrm{Hom}_K(X, K)$ define an isomorphism of functors $I_{\mathscr{C}} \to (\ , K)(\ , K)$.

(24) Let K be a field and $f: K \to R$ a K-algebra such that R is a finite-dimensional algebra over K.
(a) Show that the map $g: K \to R^{\mathrm{op}}$ given by $g(k) = f(k)$ for all k in K makes R^{op} a K-algebra.
(b) Show that an R-module M is a finitely generated R-module if and only if viewed as a vector space over K by means of the ring morphism $K \to R$ it is a finite-dimensional vector space over K.
(c) Let M be a finitely generated R-module and $f: M \to K$ a K-morphism. Show that if r is in R the map $(rf): M \to K$ defined by $rf(m) = f(rm)$ for all m in M is a K-morphism. Finally, show that the map $R^{\mathrm{op}} \times \mathrm{Hom}_K(M, K) \to \mathrm{Hom}_K(M, K)$ given by $(r, f) \to rf$ is an R^{op}-module structure on $\mathrm{Hom}_K(M, K)$. This is the only way we consider $\mathrm{Hom}_K(M, K)$ an R^{op}-module.
(d) Let $\mathrm{mod}(R)$ and $\mathrm{mod}(R^{\mathrm{op}})$ denote the full subcategories of $\mathrm{Mod}(R)$ and $\mathrm{Mod}(R^{\mathrm{op}})$ respectively consisting of finitely generated R and R^{op} modules. Show that there is a natural functor $F: \mathrm{mod}(R) \to \mathrm{mod}(R^{\mathrm{op}})$ whose map on objects is given by $F(M) = \mathrm{Hom}_K(M, K)$ and show that F is a duality.
(e) Show that a sequence $M_1 \to M_2 \to M_3$ in $\mathrm{mod}(R)$ is exact if and only if $F(M_3) \to F(M_2) \to F(M_1)$ is exact in $\mathrm{mod}(R^{\mathrm{op}})$.
(f) Show that a module M in $\mathrm{mod}(R)$ is projective if and only if $F(M)$ has the

property that if $0 \to N_1 \to N_2 \to N_3 \to 0$ is an exact sequence in $\mathrm{mod}(R^{op})$, then $\mathrm{Hom}_{R^{op}}(N_2, F(M)) \to \mathrm{Hom}_{R^{op}}(N_1, F(M)) \to 0$ is exact.

(g) Show that a morphism $f: M_1 \to M_2$ is an isomorphism if and only if $F(f): F(M_2) \to F(M_1)$ is an isomorphism.

(h) Show that $l(M) = l(F(M))$.

(i) Show that M is an indecomposable module if and only if $F(M)$ is indecomposable.

(j) Show that $\{f_i: M_i \to M\}_{i \in I}$ is a sum of a finite family of morphisms in $\mathrm{mod}(R)$ if and only if $\{F(f_i): F(M) \to F(M_i)\}_{i \in I}$ is a product in $\mathrm{mod}(R^{op})$.

(25) Let K be a field. Show that the contravariant functor $(\ , K): \mathrm{Mod}(K) \to \mathrm{Mod}(K)$ has the following properties.

(a) $(\ , K)$ is exact; that is, if $0 \to V_1 \to V_2 \to V_3 \to 0$ is an exact sequence of K-modules, then $0 \to \mathrm{Hom}_K(V_3, K) \to \mathrm{Hom}_K(V_2, K) \to \mathrm{Hom}_K(V_1, K) \to 0$ is also an exact sequence of K-modules.

(b) The (covariant) functor $(\ , K)(\ , K): \mathrm{Mod}(K) \to \mathrm{Mod}(K)$ is exact.

(c) For each vector space V define the map $\psi_V: V \to (\ , K)(\ , K)(V) = \mathrm{Hom}_K(\mathrm{Hom}_K(V, K), K)$ by $\psi_V(v)(f) = f(v)$. Prove the following:

(i) ψ_V is a K-morphism for each V in $\mathrm{Mod}(K)$.

(ii) ψ_V is a monomorphism for each V in $\mathrm{Mod}(K)$.

(iii) ψ_V is an isomorphism if and only if V is a finite-dimensional K-vector space.

(iv) If $f: V \to W$ is a morphism of K-vector spaces, then the diagram

$$\begin{array}{ccc} V & \xrightarrow{\ f\ } & W \\ \downarrow{\scriptstyle \psi_V} & & \downarrow \\ \mathrm{Hom}_K(\mathrm{Hom}_K(V, K), K) & \xrightarrow{(\ , K)(\ , K)(f)} & \mathrm{Hom}_K(\mathrm{Hom}(W, K), K) \end{array}$$

(d) Show that the family $\{\psi_V\}_{V \in \mathrm{Mod}(K)}$ is a morphism of functors from $I_{\mathrm{Mod}(K)}$ to $(\ , K)(\ , K)$.

(26) Let K be a field and $K \to R$ a finite K-algebra.

(a) Show that if M is an arbitrary R-module and f is in $\mathrm{Hom}_K(M, K)$, then for each r in R, the map $(rf): M \to K$ given by $(rf)(m) = f(rm)$ is in $\mathrm{Hom}_K(M, K)$.

(b) Prove that the map $R^{op} \times \mathrm{Hom}_K(M, K) \to \mathrm{Hom}_K(M, K)$ given by $(r, f) \to rf$ is an R^{op}-module structure on $\mathrm{Hom}_K(M, K)$.

(c) Show that the following data define a contravariant functor $(\ , K): \mathrm{Mod}(R) \to \mathrm{Mod}(R^{op})$.

(i) $(\ , K): \mathrm{Ob}\,\mathrm{Mod}(R) \to \mathrm{Ob}\,\mathrm{Mod}(R^{op})$ is given by $M \to \mathrm{Hom}_K(M, K)$.

(ii) If $f: M \to N$ is an R-morphism, then $(\ , K)(f): (\ , K)(N) \to (\ , K)(M)$ is the morphism $(f, K): \mathrm{Hom}_K(N, K) \to \mathrm{Hom}_K(M, K)$.

(d) Show that $(\ , K): \mathrm{Mod}(R) \to \mathrm{Mod}(R^{op})$ is an exact functor.

(e) For each R-module M define the map $\psi_M: M \to \mathrm{Hom}_K(\ , \mathrm{Hom}_K(M, K), K)$ by $\psi_M(m)f = f(m)$.

(i) Show that ψ_M is an injective R-morphism for each R-module M which is an isomorphism if and only if M is a finitely generated R-module.

(ii) Show that for each R-morphism $f: M \to N$, the diagram

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ \downarrow{\scriptstyle \psi_M} & & \downarrow{\scriptstyle \psi_N} \\ \mathrm{Hom}_k(\mathrm{Hom}(M, K), K) & \xrightarrow{(\cdot K)\circ(\cdot K)\circ f)} & \mathrm{Hom}_K(\mathrm{Hom}_K(N, K), K) \end{array}$$

commutes.

Definition
Let R be an arbitrary ring. An R-module I is said to be injective if given any monomorphism $A \xrightarrow{f} B$ of R-modules, the sequence $\mathrm{Hom}_R(B, I) \to \mathrm{Hom}_R(A, I) \to 0$ is exact.

(27) Let K be a field, $K \to R$ a finite-dimensional K-algebra.

(a) A finitely generated R-module I is an injective R-module if and only if $\mathrm{Hom}_R(I, K)$ is a projective R^{op}-module.

(b) If X is a finitely generated R-module, then there is a monomorphism $X \to I$ where I is a finitely generated injective R-module.

(c) If X is a finitely generated R-module, then there is a monomorphism $X \to I$ with I a finitely generated injective R-module with the property that if I' is a submodule of I and $I' \cap X = 0$, then $I' = 0$.

(d) Let M be an arbitrary R-module and M' a finitely generated submodule of R. Show there is a submodule N of M satisfying:
(i) $M' \cap M = 0$ and
(ii) M'/N is a finitely generated R-module.

(e) Let $R = T_n(K)$. Show that the left ideal I consisting of all (a_{ij}) with $a_{ij} = 0$ if $j > l$ is an injective as well as a projective $T_n(K)$-module. Are there any other left ideals of $T_n(K)$ which are injective $T_n(K)$-modules?

(28) **(a)** Let $V_1, \ldots, V_n$ be finite-dimensional vector spaces over the division rings $D_1, \ldots, D_n$, and $W_1, \ldots, W_m$ finite-dimensional vector spaces over the division rings $K_1, \ldots, K_m$. Suppose that $\prod_{i=1}^{n} \mathrm{End}_{D_i}(V_i) \approx \prod_{j=1}^{m} \mathrm{End}_{K_j}(W_j)$. Prove that $m = n$ and that there is a permutation σ of the set $\{1, \ldots, n\}$ such that $D_i \approx K_{\sigma(i)}$ and $W_{\sigma(i)} \approx V_i$ as D_i-vector spaces.

(b) Apply this to Theorem 3.4, Chapter 7, to prove the uniqueness of the division rings and vector spaces that occur in the decomposition of semisimple rings.

(29) Prove that every left nil ideal of a ring R is contained in the radical of R.

PART THREE

Chapter 9 LOCALIZATION AND TENSOR PRODUCTS

1. LOCALIZATION OF RINGS

In Chapter 5 we introduced rings of quotients of integral domains in connection with our study of UFD's. In this section we generalize this construction to arbitrary commutative rings.

Definition
A **multiplicative subset** S of a ring R is a subset of R containing 1 and not containing 0 which is a submonoid of the multiplicative monoid of R.

The reader should notice that, when R is an integral domain, the definition of multiplicative subset just given coincides with our previous definition.

Example 1.1 Let R be a commutative ring and x a nonnilpotent element of R. Then the set S of all $\{x^n\}_{n\in\mathbb{N}}$ is a multiplicative subset of R.

Example 1.2 Let P be a prime ideal in a ring R. Then $S = R - \mathfrak{P}$, the set of all r in R which are not in $\mathfrak{P}$, is a multiplicative subset of R.

PROOF: The fact that $\mathfrak{P}$ is a prime ideal says that if s_1 and s_2 are not in $\mathfrak{P}$, then $s_1 s_2$ is not in $\mathfrak{P}$. Hence, $S = R - \mathfrak{P}$ is a multiplicative subset of R.

Example 1.3 Let $f: R \to R'$ be a ring morphism. If S is a multiplicative subset of R, then $f(S)$ is a multiplicative subset of R' if and only if $S \cap \operatorname{Ker} f = \emptyset$.

Suppose that S is a multiplicative subset of a ring R. Consider the set $R \times S$.

As in the case of integral domains, we define an addition and multiplication as follows:

$$(r, s)+(r', s')=(s'r+sr', ss')$$

and

$$(r, s)(r', s')=(rr', ss')$$

As in the case of Chapter 5, it is easy to see that $R\times S$ is a commutative monoid under addition with identity (0, 1) and a commutative monoid under multiplication with identity (1, 1). Also, the map $R\to R\times S$ given by $r\to(r, 1)$ is an injective map which is both a multiplicative and additive monoid morphism.

To obtain the ring of quotients of R with respect to S, we consider the relation I on $R\times S$ defined by $(r_1, s_1)I(r_{22})$ if there exists an element s in S such that $s(s_2r_1-s_1r_2)=0$. As in the case of integral domains, it is easy to check that I is an equivalence relation with respect to both monoid structures on $R\times S$ and that $(R\times S)/I$ is a ring having the property that the canonical surjective map $k: R\times S\to(R\times S)/I$ is a morphism with respect to both monoid structures on $R\times S$. Moreover, the composite map $R\to R\times S\to(R\times S)/I$ is a ring morphism. The reader should check that if R is an integral domain, then $(R\times S)/I$ is the same as the ring R_S as defined in Chapter 5.

Definition

Let S be a multiplicative subset of the ring R. The ring $(R\times S)/I$ is called the **ring of quotients of R with respect to S** and is denoted by R_S. The ring morphism $R\to R_S$, which is the composition $R\to R\times S\to(R\times S)/I$ is called the **canonical morphism**.

Further, for each element (r, s) in $R\times S$, the image of (r, s) in R_S is denoted by r/s.

Basic Properties 1.4

Let S be a multiplicative subset of the ring R.

(a) $r/s=r'/s'$ if and only if there exists an s'' in S such that $s''(s'r-sr')=0$. Hence, $1\neq 0$.
(b) $r/s+r'/s'=(s'r+sr')/ss'$.
(c) $r/s\cdot r'/s'=rr'/ss'$.
(d) r/s is invertible in R_S if and only if r is in S.
(e) The kernel of the canonical ring morphism $R\to R_S$ given by $r\to r/1$ is the set of all r in R such that $sr=0$ for some s in S. Hence, if S consists solely of regular elements in R, the ring morphism $R\to R_S$ is injective.

PROOF: Left as an exercise.

As a consequence of (d) above, we have that the ring morphism $R\to R_S$ has the property that the image of r in R_S is a unit in R_S if and only if r is in S. In fact, this property can be used to characterize the ring of quotients R_S in the category of commutative rings, as we see in the following.

Proposition 1.5

Let S be a multiplicative subset of the ring R.

(a) If $f: R \to T$ is a ring morphism with the property that $f(s)$ is a unit in T for all s in S, then there is a unique ring morphism $g: R_S \to T$ such that f is the composition $R \longrightarrow R_S \xrightarrow{g} T$. The morphism g is given by $g(r/s) = g(r)/g(s)$ for all r in R and s in S.

(b) If $h: R \to R'$ is a morphism of rings satisfying:
   (i) $h(s)$ is a unit in R' for each s in S and
   (ii) for each ring morphism $f: R \to T$ such that $f(s)$ is a unit in T for all s in S, there is a unique ring morphism $g: R' \to T$ such that f is the composition gh, then the unique ring morphism $\omega: R' \to R_S$, such that $\omega h: R \to R_S$ is the canonical morphism, is an isomorphism of rings.

PROOF: Because the proof of this proposition is entirely analogous to that given for the characterization of fields of quotients for integral domains in Chapter 5, we recommend that the reader supply the details himself.

Before developing further properties of rings of quotients, we give a few examples.

Example 1.6 Let S be the multiplicative set $\mathbf{Z} - 2\mathbf{Z}$ and $k: \mathbf{Z} \to \mathbf{Z}/6\mathbf{Z}$ the canonical ring morphism. Because $S \cap \operatorname{Ker} k = \emptyset$, we know that $S' = k(S)$ is a multiplicative subset of $\mathbf{Z}/6\mathbf{Z}$. Then the canonical ring morphism $\mathbf{Z}/6\mathbf{Z} \to (\mathbf{Z}/6\mathbf{Z})_{S'}$ is surjective and $(\mathbf{Z}/6\mathbf{Z})_{S'} \approx \mathbf{Z}/2\mathbf{Z}$ as rings.

PROOF: It is easy to see that the multiplicative subset S' of $\mathbf{Z}/6\mathbf{Z}$ is the set $\{k(1), k(3), k(5)\}$. Suppose $f: \mathbf{Z}/6\mathbf{Z} \to T$ is a ring morphism such that $f(s)$ is invertible for all s in S'. In particular, $f(k(3))$ is a unit in T. Because $k(2)k(3) = 0$, it follows that $f(k(2)) = 0$. Hence, Ker f contains the ideal of $\mathbf{Z}/6\mathbf{Z}$ generated by $k(2)$. Therefore, there is a unique morphism $g: \mathbf{Z}/2\mathbf{Z} \to T$ such that f is the composition $\mathbf{Z}/6\mathbf{Z} \to \mathbf{Z}/2\mathbf{Z} \xrightarrow{g} T$ where $\mathbf{Z}/6\mathbf{Z} \to \mathbf{Z}/2\mathbf{Z}$ is the canonical surjective morphism. In addition, the morphism $\mathbf{Z}/6\mathbf{Z} \to \mathbf{Z}/2\mathbf{Z}$ carries every element of S' to a unit in $\mathbf{Z}/2\mathbf{Z}$, namely, 1. Hence, the morphism $\mathbf{Z}/6\mathbf{Z} \to \mathbf{Z}/2\mathbf{Z}$ satisfies Proposition 1.5 (b). Therefore, we have a unique isomorphism $\omega: \mathbf{Z}/2\mathbf{Z} \to (\mathbf{Z}/6\mathbf{Z})_{S'}$ such that the composition $\mathbf{Z}/6\mathbf{Z} \to \mathbf{Z}/2\mathbf{Z} \xrightarrow{\omega} (\mathbf{Z}/6\mathbf{Z})_{S'}$ is the canonical morphism $\mathbf{Z}/6\mathbf{Z} \to (\mathbf{Z}/6\mathbf{Z})_{S'}$. Hence, the canonical morphism $\mathbf{Z}/6\mathbf{Z} \to (\mathbf{Z}/6\mathbf{Z})_{S'}$ is surjective and $(\mathbf{Z}/6\mathbf{Z})_{S'} \approx \mathbf{Z}/2\mathbf{Z}$.

Example 1.7 Let $R = \mathbf{Z}[X]/I$ where I is the ideal generated by $2X$. Then R is not an integral domain. However, the element $k(2+X)$ is a regular noninvertible element in R where $k: \mathbf{Z}[X] \to R$ is the canonical surjective ring morphism. Hence, the multiplicative subset $S = \{k(2+X)^n\}_{n \in \mathbf{N}}$ has the property that the canonical morphism $R \to R_S$ is an injective, but not surjective, epimorphism of rings.

PROOF: Because $\mathbf{Z}$ is a UFD, we know that $\mathbf{Z}[X]$ is also a UFD. We have also seen the fact that 2 is a prime element of $\mathbf{Z}$ implies 2 is a prime element of $\mathbf{Z}[X]$. Because $\mathbf{Z}[X]/X \approx \mathbf{Z}$ we know that X is a prime element in $\mathbf{Z}[X]$. Hence, $k(2)$ and $k(X)$ are not zero in R, even though $k(2)k(X) = k(2X)$ is zero in R. Therefore, R is not an integral domain.

We now show that $k(2+X)$ is regular in R. Suppose $k(f(X))k(2+X) = 0$. This means that $f(X)(2+X) = 2Xg(X)$ for some $g(X)$ in $\mathbf{Z}[X]$. Because neither 2 nor X divides $2+X$ in $\mathbf{Z}[X]$, it follows from the fact that $\mathbf{Z}[X]$ is a UFD that $2X | f(X)$. Hence, $k(f(X)) = 0$, which means that $k(2+X)$ is regular in R. Thus,

$S = \{k(2+X)^n\}_{n\in\mathbf{N}}$ consists solely of regular elements in R. This implies that the canonical morphism $R \to R_S$ is injective.

Clearly, the image of $k(2+X)$ in R_S is invertible in R_S. Therefore, to show that $R \to R_S$ is not surjective, it suffices to show that $k(2+X)$ is not invertible in R.

Suppose $k(2+X)$ were invertible in R. Then we would have elements $f(X)$ and $g(X)$ in $\mathbf{Z}[X]$ such that $f(X)(2+X) = 1+2Xg(X)$. Setting $X=0$ we get $f(0)2 = 1$, which means that 2 divides 1 in $\mathbf{Z}$. This contradiction shows that $k(2+X)$ is not invertible in R.

The fact that $R \to R_S$ is an epimorphism is a consequence of our characterization of rings of quotients given in Proposition 1.5, which we state formally in the following.

Corollary 1.8
Let S be a multiplicative subset of the ring R.

(a) The canonical morphism $R \to R_S$ is an epimorphism in the category of commutative rings.
(b) The canonical morphism is an isomorphism if and only if every element of S is invertible in R.

2. LOCALIZATION OF MODULES

One of the most useful tools in commutative ring theory is the localization of modules, a construction which parallels that for rings.

Let S be a multiplicative subset of a ring R. For each R-module M consider the relation N on $M \times S$ defined by $(m_1, s_1)\mathrm{N}(m_2, s_2)$ if there exists an s in S such that $s(s_2m_1 - s_1m_2) = 0$. N is an equivalence relation. Denoting the equivalence class of an element (m, s) by m/s, we make $(M \times S)/\mathrm{N}$ into an R_S-module as follows:

(a) $m_1/s_1 + m_2/s_2 = (s_2m_1 + s_1m_2)/s_1s_2$.
(b) $(r/s)(m/s') = rm/ss'$.

This R_S-module is denoted by M_S.

Definition
Let S be a multiplicative subset of the ring R. For each R-module M, the R_S-module M_S is called the **module of quotients of M with respect to S.**

Basic Properties 2.1
Let S be a multiplicative subset of the ring R and let M be an R-module.

(a) $m/s = m'/s'$ if and only if there is an s'' in S such that $s''(s'm - sm') = 0$.
(b) $m/s + m'/s' = (s'm + sm')/ss'$.
(c) $(r/s)(m/s') = rm/ss'$ for all r in R and s in S.
(d) For each s in S, the R_S-morphism $M_S \to M_S$ given by $m/s' \mapsto (s/1)(m/s')$ is an isomorphism.

The canonical ring morphism $R \to R_S$ enables us to consider M_S an R-module. This operation is given by $r(m/s) = rm/s$ for all r in R, m in M, and s in S. The map $M \to M_S$ given by $m \to m/1$ is easily seen to be an R-morphism.

Proposition 2.2
Let S be a multiplicative subset of R.

(a) For each R-module M the kernel of the R-morphism $M \to M_S$ consists of all m in M such that $sm = 0$ for some s in S.
(b) If a subset X of M generates M, then the image of X in M_S generates M_S as an R_S-module. Thus, if M is a finitely generated R-module, then M_S is a finitely generated R_S-module.
(c) If $f: M \to M'$ is a morphism of R-modules, then there is one and only one morphism $f_S: M_S \to M'_S$ of R_S-modules such that the diagram of R-modules

$$\begin{array}{ccc} M & \xrightarrow{f} & M' \\ \downarrow & & \downarrow \\ M_S & \xrightarrow{f_S} & M'_S \end{array}$$

commutes. The morphism $f_S: M_S \to M'_S$ is given by $f_S(m/s) = f(m)/s$ for all m in M and s in S.

Because the morphisms f_S of R_S-modules associated with morphisms f of R-modules are very important, we now list some of their useful properties.

Basic Properties 2.3
Let S be a multiplicative subset of a ring R.

(a) If a morphism $f: M \to M$ of R-modules is id_M or 0, then $f_S: M_S \to M_S$ is id_{M_S} or 0.
(b) If $f: M \to M''$ and $g: M'' \to M'''$ are morphisms of R-modules, then $(gf)_S = g_S f_S$.
(c) The map $\mathrm{Hom}_R(M, M') \to \mathrm{Hom}_{R_S}(M_S, M'_S)$ given by $f \mapsto f_S$ is a morphism of R-modules. [Recall that $\mathrm{Hom}_R(M, M')$ is always a module over the center of R. Because R is commutative, $\mathrm{Hom}_R(M, M')$ is an R-module.]
(d) The map $\mathrm{End}_R(M) \to \mathrm{End}_{R_S}(M_S)$ given by $f \mapsto f_S$ is a morphism of rings.

A less obvious property of localization than those we have considered until now is the following.

Proposition 2.4

Let $M' \xrightarrow{f} M \xrightarrow{g} M''$ be an exact sequence of R-modules. If S is a multiplicative subset of R, then the sequence of R_S-modules $M'_S \xrightarrow{f_S} M_S \xrightarrow{g_S} M''_S$ is also exact. Hence, if $\cdots \to M_{i+1} \to M_i \to M_{i-1} \to M_{i-2} \to \cdots$ is an exact sequence of R-modules, then $\cdots \to (M_{i+1})_S \to (M_i)_S \to (M_{i-1})_S \to (M_{i-2})_S \to \cdots$ is an exact sequence of R_S-modules.

PROOF: Because $fg = 0$, we have by Basic Properties 2.3 that $0 = (gf)_S = g_S f_S$. Thus, $\mathrm{Im}\, f_S \subset \mathrm{Ker}\, g_S$. We now show that $\mathrm{Ker}\, g_S \subset \mathrm{Im}\, f_S$.

Suppose $g_S(m/s) = 0$. Then $g(m)/s = 0$. Hence, there is an element s' in S such that $s'g(m) = 0$. Therefore, $g(s'm) = 0$, which means $s'm = f(m')$ for some

m' in M' because $\operatorname{Im} f = \operatorname{Ker} g$. From this it follows that $f_S(m'/s's) = m/s$, which implies $\operatorname{Im} f_S \supset \operatorname{Ker} g_S$.

The rest of the proposition is a formal consequence of what we have shown.

We now turn our attention to studying the relation between the R-submodules of the R-module M and the R_S-submodules of the R_S-module M_S. This situation is very similar to that considered in Chapter 5 for ideals in integral domains.

Suppose M is an R-module and that N is an R_S-submodule of M_S. We denote by $M \cap N$ the preimage of N under the morphism $M \to M_S$. Because $M \to M_S$ is a morphism of R-modules, $M \cap N$ is a submodule of M. Therefore, by Proposition 2.4, $(M \cap N)_S$ is an R_S-submodule of M_S.

Proposition 2.5
The R_S-submodule $(M \cap N)_S$ of M_S is the R_S-submodule N of M_S.

PROOF: First we show that $(M \cap N)_S \subset N$. Each element in $(M \cap N)_S$ is of the form m/s with m in $M \cap N$ and s in S. Because m is in N and N is an R_S-submodule of M_S, we have shown that $m/s = (1/s)(m)$ is in N. Hence, $(M \cap N)_S \subset N$.

We now show that $(M \cap N)_S \supset N$. Let y be an element of N. Then $y = m/s$ for some m in M. Because sy is in N, m is also in N. Thus, m is in $M \cap N$ and so m/s is in $(M \cap N)_S$.

Corollary 2.6
If M is a noetherian R-module, then M_S is a noetherian R_S-module. In particular, if R is a noetherian ring, then R_S is a noetherian ring.

PROOF: We need only show that each R_S-submodule N of M_S is a finitely generated R_S-module. We know by Proposition 2.5 that $N = (M \cap N)_S$. Because M is a noetherian R-module, $M \cap N$ is a finitely generated R-module. Therefore, by Proposition 2.2, we have that $(M \cap N)_S$, and hence N, is a finitely generated R_S-module.

We know that if M' is a submodule of an R-module M, then $(M')_S$ is an R_S-submodule of M_S. In order to investigate the connections between the R-submodules M' and $(M')_S \cap M$, it is convenient to introduce the following.

Definitions
Let S be a multiplicative subset of a ring R and M an R-module.

(a) The kernel of the canonical morphism $M \to M_S$ is called the **S-torsion submodule of M** and is denoted by $t_S(M)$. M is an **S-torsion module** if $M = t_S(M)$. M is **S-torsionless** if $t_S(M) = 0$.

(b) If M' is a submodule of M, then the S-closure of M' in M is the preimage of the S-torsion submodule of M/M' under the canonical morphism $M \to M/M'$. The S-closure of M' in M is denoted by $Cl_S(M')$. M' is **S-closed** if $M' = Cl_S(M')$.

Basic Properties 2.7

(a) $t_S(M)$ is the set of all m in M such that $sm = 0$ for some s in S.
(b) $t_S(M)$ is an S-torsion module and $M/t_S(M)$ is S-torsionless.
(c) $Cl_S(M')$ is the set of m in M such that sm is in M' for some s in S. $Cl_S(M')$ is S-closed and $M/Cl_S(M')$ is S-torsionless.
(d) If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules, then $0 \to M' \to Cl_S(M') \to t_S(M'') \to 0$ is exact. Hence, the morphism $M'_S \to (Cl_S(M'))_S$ is an isomorphism.

PROOF: We prove the last part of (d). Because $0 \to M' \to Cl_S(M') \to t_S(M'') \to 0$ is an exact sequence of R-modules, we know by Proposition 2.4 that the sequence of R_S-modules $0 \to (M')_S \to (Cl_S(M'))_S \to (t_S(M''))_S \to 0$ is exact. But $(t_S(M''))_S = 0$ because $t_S(M'')$ is an S-torsion module. Hence, $0 \to (M')_S \to (Cl_S(M'))_S \to 0$ is exact which means that $(M')_S \to (Cl_S(M'))_S$ is an isomorphism.

We return to the problem of describing $M'_S \cap M$ for submodules M' of M.

Proposition 2.8
For each submodule M' of M we have that $M'_S \cap M = Cl_S(M')$.

PROOF: Clearly, $M'_S \cap M$ contains M'. If x is an element of $M'_S \cap M$, then $x = m'/s$ with m' in M'. Hence, $sx = m'$ is in M'. Because x is also in M, we see that $M'_S \cap M \subset Cl_S(M')$.

Suppose x is in $Cl_S(M')$. Then x is in M and $sx = m'$ in M' for some s in S. Thus, $x = m'/s$ in M'_S which implies that x is in $M'_S \cap M$. Therefore, $M'_S \cap M \supset Cl_S(M')$ which finishes the proof that $M'_S \cap M = Cl_S(M')$.

Corollary 2.9
If N is an R_S-submodule of M_S, then $N \cap M$ is S-closed.

PROOF: We know that $N = (N \cap M)_S$ by Proposition 2.5. Therefore, $N \cap M = (N \cap M)_S \cap M$ which is S-closed by Proposition 2.8.

Corollary 2.10
Suppose $\mathfrak{P}$ is a prime ideal in R and S a multiplicative subset of R.

(a) $\mathfrak{P}_S = R_S$ if and only if $S \cap \mathfrak{P} \neq \emptyset$.
(b) If $S \cap \mathfrak{P} = \emptyset$, then:
  (i) $\mathfrak{P}$ is S-closed.
  (ii) $\mathfrak{P}_S$ is a prime ideal of R_S.

We summarize all of these results in the following.

Theorem 2.11
Let S be a multiplicative subset of R.

(a) If M is an R-module, then the map from the set of submodules of M_S to the set of S-closed submodules of M given by $N \mapsto N \cap M$ is a bijective map of sets.
(b) The map from the set of prime ideals of R_S to the set of prime ideals of R which do not meet S given by $\mathfrak{P} \mapsto \mathfrak{P} \cap R$ is a bijective map of sets.

PROOF: (a) follows from previous discussion.

(b) follows from previous discussion once one makes the following observation. If $\mathfrak{P}$ is a prime ideal of R_S, then $\mathfrak{P} \cap R$ is a prime ideal of R because it is the preimage of $\mathfrak{P}$ under the ring morphism $R \to R_S$. That $\mathfrak{P} \subset R$ does not meet S follows from the fact that $(\mathfrak{P} \cap R)_S = \mathfrak{P} \neq R$.

3. APPLICATIONS OF LOCALIZATION

We now give some illustrations of how localization can be used. Some of these results will be used later on.

We have seen that ideals contained in the radical of a ring have some special properties. In fact an ideal I is in the radical of a ring R if and only if $1 + a$ is an invertible element of R for all a in I. Hence, if I is an arbitrary proper ideal in a commutative ring R and S is a multiplicative set of R containing all elements of the form $1 + a$ where a is in I, then the ideal I_S of R_S is either all of R_S or is in the radical of R_S. Because $I_S = R_S$ if and only if $I \cap S \neq \emptyset$, it is desirable to find such a multiplicative set S with the property that $S \cap I = \emptyset$.

But this is easily accomplished because the set S of all elements of the form $1 + a$ with a in I is a multiplicative subset of R which clearly does not meet I.

Lemma 3.1

If I is a proper ideal in a ring R and S is the set of all elements $1 + a$ with a in I, then:

(a) S is a multiplicative subset of R.
(b) I_S is contained in the radical of R_S.

We now apply this to the following situation.

Proposition 3.2

Let M be a finitely generated R-module and I an ideal of R such that $IM = M$. Then there is an element a in I such that $am = m$ for all m in M.

PROOF: Let $S = \{1 + a\}_{a \in I}$. Then because $M = IM$, it is easy to show that $M_S = I_S M_S$. Because M_S is a finitely generated R_S module and by Lemma 3.1 we know that $I_S \subset \operatorname{rad}(R_S)$, it follows that $M_S = 0$, that is, M is an S-torsion module. Let $m_1, \ldots, m_t$ be a finite set of generators for the R-module M and let $s_1, \ldots, s_t$ be elements of S such that $s_i m_i = 0$ for $i = 1, \ldots, t$. Then $s = \prod_{i=1}^{t} s_i$ is an element of s such that $sm_i = 0$ for $i = 1, \ldots, t$. Because $m_1, \ldots, m_t$ generate M over R, it follows that $sm = 0$ for all m in M. But $s = 1 - a$ for some a in I; hence, $am = m$ for all m in M.

Because our next illustration involves integral domains, we introduce some standard simplifications of terminology.

Definition
Let R be an integral domain and S the multiplicative subset of nonzero elements of R.

(a) An R-module is said to be a torsion module if it is S-torsion.
(b) A module is said to be torsion-free if it is S-torsionless.

Proposition 3.3
If R is an integral domain and M is a finitely generated torsion-free R-module, then M is isomorphic to a submodule of a finitely generated free module.

PROOF: Let $\{m_1, \ldots, m_s\}$ be a set of generators for M. Among the linearly independent subsets of $\{m_1, \ldots, m_s\}$, choose a maximal one and let us say it is $\{m_1, \ldots, m_t\}$. Because M is torsion-free, any set consisting of one nonzero element is linearly independent; so clearly, $t \geq 1$. For each $i \geq 1$, we have that $\{m_1, \ldots, m_t, m_{t+1}\}$ is a linearly dependent set, so we can find $r_{i1}, \ldots, r_{it}, v_i$ in R with $v_i \neq 0$ such that $v_i m_{t+i} = \sum_{j=1}^{t} r_{ij} m_j$. Let $v = v_1 \cdots v_{s-t}$. Then $v \neq 0$ and $v m_{t+i} = \sum_{j=1}^{t} a_{ij} m_j$ where $a_{ij} = (v/v_i) r_{ij} \in R$.

Hence, $vM \subset F$, where F is the free R-module generated by $\{m_1, \ldots, m_t\}$. It is easily seen that the map $f: M \to F$ given by $f(m) = vm$ for all m in M is a morphism of R-modules. Because M is torsion-free and v is a nonzero element in R, $f: M \to F$ is an injective morphism. Hence, $f_0: M \to \operatorname{Im} f$ is an isomorphism, which completes the proof.

In the following example we show that the restriction in Proposition 3.3 that the torsion-free module M be finitely generated is essential.

Example 3.4 Let R be an integral domain which is not a field, and let K be its field of quotients. Then K is a torsion-free R-module. We shall show that K cannot be a submodule of a free module. If it were, say $K \subset F$ where F is a free module, and let $X = \{x_\alpha\}$ be a basis for F. For each $y \in K$, let $r_\alpha(y) \in R$ be defined by $y = \Sigma\, r_\alpha(y) x_\alpha$. The map $r_\alpha: K \to R$ is well defined because $\{x_\alpha\}$ is a basis for F, and clearly r_α is a morphism of K to R for each α. Because $K \neq 0$, we must have $r_\alpha(y) \neq 0$ for some $y \in K$ and some α. However, every morphism from K to R is zero. If we grant this fact, it then follows that $r_\alpha(y) = 0$ for all $y \in K$ and this shows that K cannot be a submodule of a free module. To see that every morphism of K into R is zero, suppose we have $f: K \to R$ and f is not zero.

First we show that $\operatorname{Im} f \neq 0$ implies $\operatorname{Im} f = R$. Let a be a nonzero element of R. Because $aK = K$, it follows that $f(K) = f(aK) = af(K)$. Letting a be a nonzero element in $f(K)$ we know there is a b in $f(K)$ such that $ab = a$. Because $a \neq 0$, this implies $b = 1$ and hence $f(K) = R$.

Because $f(K) = R$, we have by our previous observation that $aR = R$ for all nonzero a in R. This implies that R is a field, which contradicts our original assumption.

Our final illustration involves the notion of a local ring which we now define.

Definition
A commutative ring R is called a **local ring** if it has only one maximal ideal J. The field R/J is called the **residue class field** of the local ring R.

Basic Properties 3.5
Suppose R is a local ring with maximal ideal J.

(a) J is the radical of R.
(b) J is the set of all nonunits of R.

We now give some examples of local rings.

Example 3.6 Any field is a local ring with (0) as unique maximal ideal.

Example 3.7 We have already seen that if $\mathfrak{P}$ is a prime ideal of R, then $S = R - \mathfrak{P}$ is a multiplicative subset of R. The ring R_S is a local ring with $\mathfrak{P}_S$ its unique maximal ideal, as can be seen by applying Theorem 2.10.

This construction is so prevalent in commutative algebra that it has been given a special name and notation.

Definition
Let $\mathfrak{P}$ be a prime ideal in the ring R. The local ring R_S when $S = R - \mathfrak{P}$ is called the **local ring of** $\mathfrak{P}$ and is denoted by $R_{\mathfrak{P}}$. Moreover, for each R-module M, the $R_{\mathfrak{P}}$-module M_S is denoted by $M_{\mathfrak{P}}$. Similarly, if $f: M \to N$ is an R-morphism, the $R_{\mathfrak{P}}$-morphism $f_S: M_{\mathfrak{P}} \to N_{\mathfrak{P}}$ is denoted by $f_{\mathfrak{P}}$.

It is clear that $\mathfrak{P}_{\mathfrak{P}} = \mathfrak{P}R_{\mathfrak{P}}$ is the radical of the local ring $R_{\mathfrak{P}}$. This follows from the bijective correspondence that we have exhibited between prime ideals of R not meeting $R - \mathfrak{P}$ and the prime ideals of $R_{\mathfrak{P}}$.

The following theorem, as innocuous as it may seem, gives us one of the most important tools in commutative ring theory.

Theorem 3.8
Let M be an R-module. $M = 0$ if and only if $M_{\mathfrak{P}} = 0$ for all maximal ideals $\mathfrak{P}$ of R.

PROOF: Obviously, if $M = 0$, then $M_{\mathfrak{P}} = 0$ for all maximal ideals $\mathfrak{P}$ of R.

Suppose $M_{\mathfrak{P}} = 0$ for every maximal ideal $\mathfrak{P}$ of R. This means that for each m in M and each maximal ideal $\mathfrak{P}$, there is an element s in $R - \mathfrak{P}$ such that $sm = 0$. If $M \neq 0$, choose a nonzero element m in M. Let $I = \operatorname{ann}(m)$. Then I is not all of R and is therefore contained in some maximal ideal $\mathfrak{P}$. However, because $M_{\mathfrak{P}} = 0$, there is an s in $R - \mathfrak{P}$ such that $sm = 0$. Hence, s is in I. But $\mathfrak{P} \supset I$, so we have a contradiction.

As one indication of how this theorem is used we prove the following.

Proposition 3.9
An R-morphism $f: M \to N$ is the zero morphism if $f_{\mathfrak{P}}: M_{\mathfrak{P}} \to N_{\mathfrak{P}}$ is the zero morphism for every maximal ideal $\mathfrak{P}$ of R.

PROOF: We prove this by showing that for any multiplicative subset S of R, we have $(\operatorname{Im} f)_S = \operatorname{Im}(f_S)$.

Suppose x is in $(\operatorname{Im} f)_S$. Then $x = f(m)/s$ for some m in M and s in S. But $f(m)/s = f_S(m/s)$. So x is in $\operatorname{Im}(f_S)$. Conversely, if x is in $\operatorname{Im}(f_S)$, then $x = f_S(m/s)$ for some m/s in M_S. But $f_S(m/s) = f(m)/s$ which is in $(\operatorname{Im} f)_S$. So x is in $(\operatorname{Im} f)_S$. Hence, $(\operatorname{Im} f)_S = \operatorname{Im}(f_S)$.

Returning to our proposition, the assumption that $f_{\mathfrak{P}} = 0$ for all maximal ideals $\mathfrak{P}$ of R implies that $\mathrm{Im}(f_{\mathfrak{P}}) = 0$ for all $\mathfrak{P}$. Because $\mathrm{Im}(f_{\mathfrak{P}}) = (\mathrm{Im}\, f)_{\mathfrak{P}}$, we have that $(\mathrm{Im}\, f)_{\mathfrak{P}} = 0$ for all maximal ideals $\mathfrak{P}$ of R. By Theorem 3.8, we know that $\mathrm{Im}\, f = 0$ and hence $f = 0$.

Corollary 3.10
Let $f: M \to N$ be a morphism of R-modules.

(a) f is a monomorphism if and only if $f_{\mathfrak{P}}$ is a monomorphism for all maximal ideals $\mathfrak{P}$ in R.
(b) f is an epimorphism if and only if $f_{\mathfrak{P}}$ is an epimorphism for all maximal ideals $\mathfrak{P}$ in R.
(c) f is an isomorphism if and only if $f_{\mathfrak{P}}$ is an isomorphism for all maximal ideals $\mathfrak{P}$ in R.

4. TENSOR PRODUCTS

As in the rest of this chapter, all rings are assumed to be commutative unless specified otherwise.

In Chapter 6, Section 2, we discussed in some detail the notion of a bilinear map of modules over a commutative ring. In this section we show how the tensor product of two R-modules A and B converts the study of bilinear maps on $A \times B$ to the study of morphisms of the tensor product of A and B.

Definition
Let A and B be R-modules. A **tensor product of A and B** is an R-module $T(A, B)$ together with a bilinear map $\xi: A \times B \to T(A, B)$ satisfying the following condition:

For each R-module C and each bilinear map $\beta: A \times B \to C$ there exists a unique morphism of R-modules $\bar{\beta}: T(A, B) \to C$ such that $\bar{\beta}\xi = \beta$.

Recall that if A, B, and C are R-modules, then $B(A \times B, C)$, the group of all bilinear maps from $A \times B$ to C, is an R-module (see the end of Chapter 6, Section 2). Moreover, if $\xi: A \times B \to T(A, B)$ is a tensor product of A and B, then for each R-module C we define the map $\varphi(C): \mathrm{Hom}_R(T(A, B), C) \to B(A \times B, C)$ by $\varphi(C) \times (\beta) = \beta\xi$ for all β in $\mathrm{Hom}_R(T(A, B), C)$. It is not difficult to show that each $\varphi(C)$ is an R-morphism. We further make the convention that if X and Y are R-modules, we will denote the R-module $\mathrm{Hom}_R(X, Y)$ by $R(X, Y)$ or (X, Y), depending on whether we want to emphasize the ring R or not.

Basic Properties 4.1
Let A and B be R-modules.

(a) For each R-module C, the R-morphism $\varphi(C): (T(A, B), C) \to B(A \times B, C)$ is an isomorphism.
(b) If $\xi': A \times B \to T'(A, B)$ is another tensor product of A and B there is a unique R-morphism $h: T(A, B) \to T'(A, B)$ such that $h\xi = \xi'$. This unique morphism b is an isomorphism of R-modules.

PROOF: (a) This is essentially a recapitulation of the definition of tensor product.

(b) Because $\xi' : A \times B \to T'(A, B)$ is a bilinear map, we know there is a unique R-morphism $h : T(A, B) \to T'(A, B)$ such that $h\xi = \xi'$. Because $\xi' : A \times B \to T'(A, B)$ is a tensor product of A and B, we have the isomorphism $\varphi'(C) : (T(A, B), C) \to B(A \times B, C)$ for each R-module C. We know that the morphism $h : T(A, B) \to T'(A, B)$ induces a morphism $(h, C) : (T'(A, B), C) \to (T(A, B), C)$ for each R-module C. We leave it to the reader to check that for each R-module C, the diagram

$$\begin{array}{ccc} (T'(A, B), C) & \xrightarrow{\varphi'(C)} & B(A \times B, C) \\ \Big\downarrow{\scriptstyle (h, c)} & & \Big\| \\ (T(A, B), C) & \xrightarrow{\varphi(C)} & B(A \times B, C) \end{array}$$

commutes. The fact that the horizontal morphisms are isomorphisms implies that $(h, C) : (T'(A, B), C) \to (T(A, B), C)$ is an isomorphism for every R-module C. By Chapter 6, Basic Property 3.3 this implies that h is an isomorphism.

Having explained in what sense tensor products of R-modules are unique, we now show that every pair of R-modules has a tensor product.

Associated with any map of sets $\beta : A \times B \to C$ of the underlying set of the R-modules A, B, and C is the unique R-morphism $\beta' : F(A \times B) \to C$ satisfying $\beta' | A \times B = \beta$ where $F(A \times B)$ is the free R-module generated by the set $A \times B$. It is obvious that $\beta : A \times B \to C$ is a bilinear map if and only if the following elements of $F(A \times B)$ are in $\operatorname{Ker} \beta'$:

(i) $(a, b) + (a, b') - (a, b + b')$ for all a in A and b, b' in B.
(ii) $(a, b) + (a', b) - (a + a', b)$ for all a, a' in A, and b in B.
(iii) $(ra, b) - (a, rb)$ for all r in R, a in A, and b in B.
(iv) $r(a, b) - (ra, b)$ for all r in R, a in A, and b in B.

Let $K(A \times B)$ be the submodule of $F(A \times B)$ generated by the elements described in conditions (i), (ii), (iii), and (iv) above. One easily verified property of the submodule $K(A \times B)$ of $F(A \times B)$ is the fact that the composition $A \times B \xrightarrow{\text{inc}} F(A \times B) \xrightarrow{k} F(A \times B)/K(A \times B)$ is a bilinear map. Letting $A \underset{R}{\otimes} B = F(A \times B)/K(A \times B)$ and $\xi : A \times B \to A \underset{R}{\otimes} B$ the bilinear map $k(\text{inc})$, we verify that the pair consisting of the R-module $A \underset{R}{\otimes} B$ and the bilinear map $\xi : A \times B \to A \underset{R}{\otimes} B$ is a tensor product of A and B.

First observe that because $A \times B$ generates $F(A \times B)$, the image of ξ generates $A \underset{R}{\otimes} B$. Now suppose that $\beta : A \times B \to C$ is a bilinear map. Then by our previous discussion, the R-morphism $\beta' : F(A \times B) \to C$ contains $K(A \times B)$ in its kernel. Hence, there is a unique R-morphism $\bar{\beta} : A \underset{R}{\otimes} B \to C$ such that $\bar{\beta} k = \beta'$. But $\bar{\beta}\xi = \bar{\beta} k \,\text{inc} = \beta' \,\text{inc} = \beta' | A \times B = \beta$. Thus, we have shown that given any bilinear map $\beta : A \times B \to C$ there is an R-morphism $\bar{\beta} : A \underset{R}{\otimes} B \to C$ such that $\bar{\beta}\xi =$

β. Because we already know that Im ξ generates $A \underset{R}{\otimes} B$, $\bar{\beta} : A \underset{R}{\otimes} B \to C$ is the only R-morphism satisfying $\bar{\beta}\xi = \beta$. This completes the proof that $\xi : A \times B \to A \otimes B$ is a tensor product.

Definition
Let A and B be R-modules. We call the tensor product $\xi : A \times B \to A \otimes B$ the **standard tensor product**. For each element (a, b) in $A \times B$, we denote the element $\xi((a, b))$ by $a \otimes b$.

Basic Properties 4.2
Let A and B be R-modules.

(a) The set of all elements of the form $a \otimes b$ generates $A \underset{R}{\otimes} B$.

(b) $a \otimes (b_1 + b_2) = a \otimes b_1 + a \otimes b_2$, $(a_1 + a_2) \otimes b = a_1 \otimes b + a_2 \otimes b$, $ra \otimes b = a \otimes rb = r(a \otimes b)$ for all a, a_1, a_2 in A, all b, b_1, b_2 in B, and all r in R.

(c) $0 \otimes b = 0 = a \otimes 0$ for all a in A and b in B.

(d) $-(a \otimes b) = (-a) \otimes b = a \otimes (-b)$ for all a in A and b in B.

PROOF: Left as an exercise.

Proposition 4.3
Suppose $\{a_i\}_{i \in I}$ is a set of generators for the R-module A and $\{b_j\}_{j \in J}$ is a set of generators for the R-module B. Then:

(a) The elements $\{a_i \otimes b_j\}_{(i,j) \in I \times J}$ are a set of generators for the R-module $A \otimes B$.

(b) Let C be an R-module and let $f : \{a_i \otimes b_j\}_{(i,j) \in I \times J} \to C$ be a map of sets. Then there is an R-module morphism $g : A \underset{R}{\otimes} B \to C$ with the property $g(a_i \otimes b_j) = f(a_i \otimes b_j)$ for all $(i, j) \in I \times J$ if and only if:
  (i) $f((a_{i_1} + a_{i_2}) \otimes b_j) = f(a_{i_1} \otimes b_j) + f(a_{i_2} \otimes b_j)$ for all i_1, i_2 in I and j in J.
  (ii) $f(a_i \otimes (b_{j_1} + b_{j_2})) = f(a_i \otimes b_{j_1}) + f(a_i \otimes b_{j_2})$ for all i in I and j_1, j_2 in J.
  (iii) $f(ra_i \otimes b_j) = f(a_i \otimes rb_j) = rf(a_i \otimes b_j)$ for all r in R, i in I, and j in J.
  Further, if the map $f : \{a_i \otimes b_j\}_{I \times J} \to C$ satisfies the above condition, then there is only one R-module morphism $g : A \otimes B \to C$ such that $g(a_i \otimes b_j) = f(a_i \otimes b_j)$ for all $(i, j) \in I \times J$.

PROOF: Good exercise.

We now give some examples to illustrate these ideas.

Example 4.4 For each R-module B, there is a unique R-isomorphism $f : R \underset{R}{\otimes} B \to B$ having the property $f(r \otimes b) = rb$ for all r in R and b in B. This unique isomorphism is often considered an identification of R-modules.

PROOF: It is easily checked that the map $f' : R \times B \to B$ given by $f'(r, b) = rb$ for all r in R and B is a bilinear map. Hence, there is a unique R-morphism $f : R \underset{R}{\otimes} B \to B$ such that $f(r \otimes b) = rb$.

On the other hand, it is easily shown that the map $g : B \to R \otimes B$ given by $g(b) = 1 \otimes b$ for all b in B is also an R-morphism. But $gf(r \otimes b) = g(rb) =$

$1 \otimes rb = r \otimes b$ for all r in R and b in B. This implies that $gf = \mathrm{id}_{R \otimes B}$ because the elements of the form $r \otimes b$ generate $R \otimes B$. Also, $fg(b) = f(1 \otimes b) = b$ for all b in B. Hence, $fg = \mathrm{id}_B$. So f is an isomorphism with $f^{-1} = g$.

Example 4.5 Let M be an R-module and I an ideal of R. Then there is a unique R-isomorphism $f: (R/I) \underset{R}{\otimes} M \to M/IM$ satisfying $f(k_1(r) \otimes m) = k_2(rm)$ where $k_1: R \to R/I$ and $k_2: M \to M/IM$ are the canonical surjective R-morphisms.

PROOF: First observe that if $k_1(r) = k_1(r')$, then $k_2(rm) = k_2(r'm)$. Hence, we have a map $f': R/I \times M \to M/IM$ given by $f'(k_1(r), m) = k_2(rm)$ for all r in R and m in M. Straightforward calculations show that this is a bilinear map. Therefore, there is a unique R-morphism $f: R/I \underset{R}{\otimes} M \to M/IM$ satisfying $f(k_1(r) \otimes m) = k_2(rm)$ for all r in R and m in M.

On the other hand, consider the map $g': M \to R/I \underset{R}{\otimes} M$ where $g'(m) = k_1(1) \otimes m$ for all m in M. This map is easily seen to be an R-morphism. Suppose a is in I and m is in M. Then $g'(am) = k_1(1) \otimes am = ak_1(1) \otimes m = 0 \otimes m = 0$. Hence, $\mathrm{Ker}\, g' \supset IM$. Therefore, we have an R-morphism $g: M/IM \to R/I \underset{R}{\otimes} M$ such that $g(k_2(m)) = k_1(1) \otimes m$ for all m in M. The reader can verify, as in Example 4.4, that gf and fg are identity maps. Hence, f is an isomorphism.

Example 4.6 Suppose $f: R \to R'$ is a ring morphism. Because R' is an R-module, for any R-module M we have $R' \underset{R}{\otimes} M$. Then for each element r' in R, there is a unique R-morphism $f_{r'} = R' \underset{R}{\otimes} M \to R' \underset{R}{\otimes} M$ having the property $f_{r'}(x \otimes m) = r'x \otimes m$ for every x in R' and m in M.

Moreover, the map $R' \times (R' \underset{R}{\otimes} M) \to R' \underset{R}{\otimes} M$ given by $(r', y) = f_{r'}(y)$ for all r' in R' and y in $R' \underset{R}{\otimes} M$ is an R'-module structure on $R' \underset{R}{\otimes} M$.

In connection with this example we make the following convention. If M is an R-module and $f: R \to R'$ is a ring morphism, the only R-module structure we will consider on $R' \underset{R}{\otimes} M$ is that described above.

Example 4.7 Let S be a multiplicative subset of a ring R. For each R-module M, there is a unique R_S-isomorphism $f: R_S \underset{R}{\otimes} M \to M_S$ such that $f(r/s \otimes m) = rm/s$ for all r in R, s in S, and m in M. This isomorphism is usually considered an identification of R_S-modules.

PROOF: First observe that the map $f': R_S \times M \to M_S$ given by $f'(r/s, m) = rm/s$ is a bilinear map of R-modules. Hence, there is a unique R-morphism $f: R_S \underset{R}{\otimes} M \to M_S$ such that $f(r/s \otimes m) = rm/s$. Next, observe that f is not only an R-morphism but also an R_S-morphism.

Now define a map $g': M \times S \to R_S \underset{R}{\otimes} M$ by $g'(m, s) = 1/s \otimes m$ for all m in M and s in S. Recalling that $M \times S$ is an R-module by means of the operation $r(m, s) = (rm, s)$, it is easily seen that g' is an R-morphism. Moreover, the submodule N, consisting of (m, s) such that $s'm = 0$ for some s' in S, is contained in

Ker g'. For, if $s'm = 0$, then $g'(m, s) = 1/s \otimes m = s'/ss' \otimes m = 1/ss' \otimes s'm = 1/ss' \otimes 0 = 0$. Hence, g' indicates a unique R-morphism $g : M_S \to R_S \underset{R}{\otimes} M$ (recall that $M_s = M \times S/N$). The reader should now be able to verify that gf and fg are identity maps. This shows that f is an R_S-isomorphism with $f^{-1} = g$.

Example 4.8 If A and B are R-modules, then there is a unique R-isomorphism $f : A \underset{R}{\otimes} B \to B \underset{R}{\otimes} A$ such that $f(a \otimes b) = b \otimes a$ for all a in A and b in B. In case $A = B$, notice the morphism $f : A \underset{R}{\otimes} A \to A \underset{R}{\otimes} A$ is not necessarily the identity morphism.

Example 4.9 If A, B, and C are R-modules, there is a unique R-isomorphism $f : (A \underset{R}{\otimes} B) \otimes C \to A \underset{R}{\otimes} (B \otimes C)$ satisfying $f((a \otimes b) \otimes c) = a \otimes (b \otimes c)$ for all a in A, b in B, and c in C. This isomorphism is usually regarded as an identification.

PROOF: Consider the map $f'' : A \times B \times C \to A \underset{R}{\otimes} (B \underset{R}{\otimes} C)$ given by $f''((a, b, c)) = a \otimes (b \otimes c)$. For each c in C, we obtain the map $f''_c : A \underset{R}{\otimes} B \to A \underset{R}{\otimes} (B \underset{R}{\otimes} C)$ given by $f''_c((a, b)) = f''((a, b, c)) = a \otimes (b \otimes c)$. f''_c is a bilinear map for each c in C. Hence, for each c in C, there is a unique R-morphism $f'_c : A \underset{R}{\otimes} B \to A \underset{R}{\otimes} (B \otimes C)$ satisfying $f'_c(a \otimes b) = a \otimes (b \otimes c)$.

Define the map $f' : (A \underset{R}{\otimes} B) \times C \to A \underset{R}{\otimes} (B \underset{R}{\otimes} C)$ by $f'((x, c)) = f'_c(x)$ for each x in $A \underset{R}{\otimes} B$ and c in C. Notice that if $x = a \otimes b$, then $f'((a \otimes b, c)) = a \otimes (b \otimes c)$. The reader should verify that f' is a bilinear map. Hence, there is a unique R-morphism $f : (A \underset{R}{\otimes} B) \otimes C \to A \underset{R}{\otimes} (B \underset{R}{\otimes} C)$ such that $f((a \otimes b) \otimes c) = f'((a \otimes b, c)) = a \otimes (b \otimes c)$.

A similar argument shows that there is a unique R-morphism $g : A \underset{R}{\otimes} (B \underset{R}{\otimes} C) \to (A \underset{R}{\otimes} B) \underset{R}{\otimes} C$ such that $g(a \otimes (b \otimes c)) = (a \otimes b) \otimes c$. It is then easy to show that gf and fg are identity morphisms. This finishes the proof of the example.

Example 4.10 Let A and B be R-modules and S a multiplicative subset of R. Then the R_S-morphism $f : A_S \underset{R}{\otimes} B \to A \underset{R}{\otimes} B_S$ given by $f(a/s \otimes b) = (a \otimes b/s)$ for all a in A, b in B, and s in S is an isomorphism of R_S-modules.

PROOF: This follows from the preceding example, because $A_S \underset{R}{\otimes} B = (A \underset{R}{\otimes} R_S) \underset{R}{\otimes} B = A \underset{R}{\otimes} (R_S \underset{R}{\otimes} B) = A \underset{R}{\otimes} B_S$ where all the equalities are our agreed upon identifications as R-modules. The reader can check easily that these are R_S-isomorphisms and that the resulting identification $A_S \underset{R}{\otimes} B = A \underset{R}{\otimes} B_S$ is given by the above morphism f.

Example 4.11 Let $f : R \to R'$ be a ring morphism. Then there is a unique multiplication on the abelian group $R' \underset{R}{\otimes} R'$ satisfying $(r'_1 \otimes r'_2)(r'_3 \otimes r'_4) = r'_1 r'_3 \otimes r'_2 r'_4$

which makes $R' \underset{R}{\otimes} R'$ a commutative ring with $1 = 1 \underset{R}{\otimes} 1$. Further, the map $m : R' \underset{R}{\otimes} R' \to R'$ given by $m(r_1' \otimes r_2') = r_1' r_2'$ is a surjective ring morphism.

PROOF: This result can be established using techniques similar to those employed in Example 4.9.

Example 4.12 Let S be a multiplicative subset of a ring R. By the preceding example, $R_S \underset{R}{\otimes} R_S$ is a ring. The surjective ring morphism $m : R_S \underset{R}{\otimes} R_S \to R_S$ given by $m(r/s \otimes r'/s') = rr'/ss'$ is an isomorphism of rings, which we consider an identification.

PROOF: it is easily checked that the map $h : R_S \to R_S \underset{R}{\otimes} R_S$ given by $h(r/s) = r/s \otimes 1$ is a morphism of rings. We now show that $hm = \mathrm{id}_{R_S \underset{R}{\otimes} R_S}$. Because h and m are also R-module morphisms, it is sufficient to show that $hm(r/s \otimes r'/s') = r/s \otimes r'/s'$ for all r, r' in R and s, s' in S because the elements $r/s \otimes r'/s'$ generate $R_S \underset{R}{\otimes} R_S$ as an R-module. Now $hm(r/s \otimes r'/s') = h(rr'/ss') = rr'/ss' \otimes 1$. But $r/s \otimes r'/s' = rs'/ss' \otimes r'/s' = r/ss' \otimes s'r'/s' = r/ss' \otimes r'/1 = rr'/ss' \otimes 1$. Hence, $hm(r/s \otimes r'/s') = r/s \otimes r'/s'$ for all r, r' in R and s, s' in S. So $hm = \mathrm{id}_{R_S \underset{R}{\otimes} R_S}$ and m is therefore injective. This shows that m is an isomorphism.

5. MORPHISMS OF TENSOR PRODUCTS

Suppose we are given R-morphisms $f : A \to A'$ and $g : B \to B'$. We consider in this section how these morphisms may be put together to obtain an R-morphism from $A \underset{R}{\otimes} B$ to $A' \underset{R}{\otimes} B'$.

We observe that the map $h : A \times B \to A' \underset{R}{\otimes} B'$ given by $h((a, b)) = f(a) \otimes g(b)$ for all a in A and b in B is a bilinear map. Hence, there is a unique R-morphism $\bar{h} : A \otimes B \to A' \otimes B'$ such that $\bar{h}(a \otimes b) = h((a, b)) = f(a) \otimes g(b)$ for all a in A and b in B.

Definition

Suppose that $f : A \to A'$ and $g : B \to B'$ are R-module morphisms. The unique R-morphism from $A \underset{R}{\otimes} B$ to $A' \underset{R}{\otimes} B'$ given by $a \otimes b \to f(a) \otimes g(b)$ is denoted by $f \otimes g$ and is called the **tensor product** of f and g.

If $g = \mathrm{id}_B$, we will often write $f \otimes B$ instead of $f \otimes \mathrm{id}_B$, similarly, for $f = \mathrm{id}_A$.

The notation $f \otimes g$ is suggestive, because $(f \otimes g)(a \otimes b) = f(a) \otimes g(b)$ for all a in A and b in B. We now list some easily verified properties.

Basic Properties 5.1

Let A, A' and B, B' be R-modules.

(a) $(f_1 + f_2) \otimes g = f_1 \otimes g + f_2 \otimes g$ for all R-morphisms $f_1, f_2 : A \to A'$ and all R-morphisms $g : B \to B'$.

(b) $f\otimes(g_1+g_2)=f\otimes g_1+f\otimes g_2$ for all R-morphisms $f:A\to A'$ and all R-morphisms $g_1, g_2:B\to B'$.
(c) $rf\otimes g=f\otimes rg=r(f\otimes g)$ for all r in R, $f\in(A,A')$, $g\in(B,B')$. Hence:
(d) The map $(A,A')\times(B,B')\to(A\underset{R}{\otimes}B, A'\underset{R}{\otimes}B')$ given by $(f,g)\to f\otimes g$ is bilinear.

In addition, we also point out the following.

Basic Properties 5.2
Let A, A', A'' and B, B', B'' be R-modules.

(a) Suppose $f_1:A\to A'$, $f_2:A'\to A''$ and $g_1:B\to B'$, $g_2:B'\to B''$ are R-morphisms. Then $(f_2f_1)\otimes(g_2g_1)=(f_2\otimes g_2)(f_1\otimes g_1)$.
(b) $\mathrm{id}_A\otimes\mathrm{id}_B=\mathrm{id}_{A\underset{R}{\otimes}B}$.
(c) If $f:A\to A'$ and $g:B\to B'$ are isomorphisms, then $(f\otimes g):A\underset{R}{\otimes}B\to A'\underset{R}{\otimes}B'$ is an isomorphism with inverse $f^{-1}\otimes g^{-1}$.

Proposition 5.3
If $0\to M'\xrightarrow{f}M\xrightarrow{g}M''\to 0$ is a splittable exact sequence of R-modules, then for any R-module B, the sequence $0\to B\underset{R}{\otimes}M'\xrightarrow{\mathrm{id}_B\otimes f}B\underset{R}{\otimes}M\xrightarrow{\mathrm{id}_B\otimes g}B\underset{R}{\otimes}M''\to 0$ is also a splittable exact sequence of R-modules.

PROOF: The fact that $0\to M'\xrightarrow{f}M\xrightarrow{g}M''\to 0$ is a splittable exact sequence means that there are morphisms s and t such that $sf=\mathrm{id}_M$, $gt=\mathrm{id}_{M''}$, $st=0$, and $fs+tg=\mathrm{id}_M$. Hence, $\mathrm{id}_B\otimes s$ and $\mathrm{id}_B\otimes t$ have the following properties: $(\mathrm{id}_B\otimes s)(\mathrm{id}_B\otimes f)=\mathrm{id}_{B\underset{R}{\otimes}M'}$, $(\mathrm{id}_B\otimes g)(\mathrm{id}_B\otimes t)=\mathrm{id}_{B\underset{R}{\otimes}M''}$, $(\mathrm{id}_B\otimes s)(\mathrm{id}_B\otimes t)=0$, and $(\mathrm{id}_B\otimes f)(\mathrm{id}_B\otimes s)+(\mathrm{id}_B\otimes t)(\mathrm{id}_B\otimes g)=\mathrm{id}_{B\underset{R}{\otimes}M}$. Because $(\mathrm{id}_B\otimes f)(\mathrm{id}_B\otimes g)=0$, it follows from Chapter 6, Basic Property 10.6 that $0\to B\underset{R}{\otimes}M'\xrightarrow{\mathrm{id}_B\otimes f}B\underset{R}{\otimes}M\xrightarrow{\mathrm{id}_B\otimes g}B\underset{R}{\otimes}M''\to 0$ is an exact splittable sequence of R-modules.

We now show how Example 4.6 can be generalized using these notions.

Example 5.4 Let $f:R\to R'$ be a ring morphism. Suppose M is an R-module and M' is an R'-module. Then M' is also an R-module and we have the R-module $M'\underset{R}{\otimes}M$. For each r' in R', the map $f_{r'}:M'\to M'$ given by $f_{r'}(m')=r'm'$ is an R-morphism. Therefore, for each r' in R', we have the R-morphism $f_{r'}\otimes 1_M:M'\underset{R}{\otimes}M\to M'\underset{R}{\otimes}M$. From Basic Properties 5.1 and 5.2 it follows that the map $R'\times(M'\underset{R}{\otimes}M)\to M'\underset{R}{\otimes}M$ given by $(r',x)\to(f_{r'}\otimes 1_M)(x)$ is an R'-module structure on $M'\underset{R}{\otimes}M$. Simplifying notation, the operation of R' on $M'\otimes M$ is uniquely determined by $r'(m'\otimes m)=r'm'\otimes m$ for all r' in R', m' in M', and m in M. Thus, if $x=\Sigma_{i=1}^{n}m_i'\otimes m_i$ is an arbitrary element of $M'\underset{R}{\otimes}M$, then $r'x=\Sigma_{i=1}^{n}r'm_i'\otimes m_i$. The reader should check that when $M'=R'$, the R'-module $R'\underset{R}{\otimes}M$ we just described is the same as that of Example 4.6.

As with Example 4.6, we make the convention that $M' \otimes M$ will be considered an R'-module only as described in Example 5.4.

Example 5.5 Let $f: R \to R'$ be a ring morphism. If $g: M_1' \to M_2'$ is a morphism of R'-modules and $h: M_1 \to M_2$ is a morphism of R-modules, then the R-morphism $g \otimes h: M_1' \otimes M_1 \to M_2' \underset{R}{\otimes} M_2$ is also an R'-morphism. This and the obvious analogs of Basic Properties 5.1 and 5.2 for this situation involving both R'- and R-modules are left to the reader to verify.

Our next proposition is concerned with the following situation.

Let $f: R \to R'$ be a ring morphism. Suppose A is an R-module and B', C' are R'-modules. If $g: B' \underset{R}{\otimes} A \to C'$ is an R'-morphism, then for each a in A, define the map $g_a: B' \to C'$ by $g_a(b') = g(b' \otimes a)$. It is easy to see that g_a is an R'-morphism, that is, g_a is in $\mathrm{Hom}_{R'}(B', C')$. Because $\mathrm{Hom}_{R'}(B', C')$ is an R'-module, it is also an R-module. Letting $\psi(g): A \to \mathrm{Hom}_{R'}(B', C')$ be the map given by $[\psi(g)](a) = g_a$, it can be shown that $\psi(g)$ is a morphism of R-modules.

We can therefore define a map $\psi: \mathrm{Hom}_{R'}(B' \underset{R}{\otimes} A, C') \to \mathrm{Hom}_R(A, \mathrm{Hom}_{R'}(B', C'))$ by $g \to \psi(g)$. Again the reader can check that this map ψ is a morphism of abelian groups, that is, $\psi(g_1 + g_2) = \psi(g_1) + \psi(g_2)$.

Next we observe that not only is $\mathrm{Hom}_{R'}(B' \underset{R}{\otimes} A, C')$ an R'-module, but $\mathrm{Hom}_R(A, \mathrm{Hom}_{R'}(B', C'))$ also has a natural R'-module structure. To see this we observe that if X' is any R'-module, then the map $R' \times \mathrm{Hom}_R(A, X') \to \mathrm{Hom}_R(A, X')$ given by $(r', f) \to r'f$, where $(r'f)(a) = r'(f(a))$, is an R'-module structure on $\mathrm{Hom}_R(A, X')$. Letting $X' = \mathrm{Hom}_{R'}(B', C')$, we obtain the R'-module structure on $\mathrm{Hom}_R(A, \mathrm{Hom}_{R'}(B', C'))$ we wanted. We leave it to the reader to check that the morphism of abelian groups $\psi: \mathrm{Hom}_{R'}(B' \underset{R}{\otimes} A, C') \to \mathrm{Hom}_R(A, \mathrm{Hom}_{R'}(B', C''))$ is also a morphism of R'-modules.

Proposition 5.6
Let $f: R \to R'$ be a ring morphism. If A is an R-module and B', C' are R'-modules, then the morphism of R'-modules $\psi: \mathrm{Hom}_{R'}(B' \underset{R}{\otimes} A, C') \to \mathrm{Hom}_R(A, \mathrm{Hom}_{R'}(B', C'))$ given by $[\psi(g)(a)](b') = g(b' \otimes a)$ for all g in $\mathrm{Hom}_{R'}(B' \underset{R}{\otimes} A, C')$, a in A, and b' in B' is an isomorphism of R'-modules.

PROOF: We first show that ψ is a monomorphism. Suppose $\psi(g) = 0$. Then $\psi(g)(a) = 0$ for all a in A. This means that for every b' in B' we have $[\psi(g)(a)](b') = 0$ for each a in A. Because $[\psi(g)(a)](b') = g(b' \otimes a)$, it follows that $g(b' \otimes a) = 0$ for all elements $b' \otimes a$ in $B' \otimes A$. Hence, $g = 0$, because the elements $b' \otimes a$ generate $B' \underset{R}{\otimes} A$. Thus, Ker $\psi = 0$ so ψ is a monomorphism.

To show that ψ is surjective, let $h: A \to \mathrm{Hom}_{R'}(B', C')$ be an R-morphism. Define a map $g': B' \times A \to C'$ by $g'((b', a)) = [h(a)](b')$. The reader can check that g' is a bilinear map of R-modules. Hence, there is a unique R-morphism $g: B' \otimes A \to C'$ such that $g(b' \otimes a) = g'((b', a)) = [h(a)](b')$. To see that g is an R'-morphism and not just an R-morphism, we first recall that an element of

$B' \underset{R}{\otimes} A$ is of the form $\Sigma_{i=1}^{n} b'_i \otimes a_i$, and

$$r'\left(\sum_{i=1}^{n} b'_i \otimes a_i\right) = \sum_{i=1}^{n} r' b'_i \otimes a_i$$

for all r' in R'. Hence,

$$\begin{aligned} g\left(r'\left(\sum_{i=1}^{n} b'_i \otimes a_i\right)\right) &= g\left(\sum_{i=1}^{n} r' b'_i \otimes a_i\right) \\ &= \sum_{i=1}^{n} g(r' b'_i \otimes a_i) = \sum_{i=1}^{n} h(a_i)(r' b'_i) \\ &= \sum_{i=1}^{n} r'(h(a_i))(b_i)) = r' g\left(\sum_{i=1}^{n} b'_i \otimes a_i\right)\Big). \end{aligned}$$

The reader can check easily that $\psi(g) = h$, which shows that ψ is surjective and hence an isomorphism.

As an immediate consequence, we have the following.

Corollary 5.7
If A, B, and C are R-modules, then the R-morphism

$$\psi : \mathrm{Hom}_R(B \underset{R}{\otimes} A, C) \to \mathrm{Hom}_R(A, \mathrm{Hom}_R(B, C))$$

given by

$$[\psi(g)(a)](b) = g(b \otimes a)$$

for all g in $\mathrm{Hom}_R(B \underset{R}{\otimes} A, C)$, a in A, and b in B is an isomorphism.

Another important consequence of Proposition 5.6 is the following.

Proposition 5.8
Let $\{f_i : A_i \to A\}_{i \in I}$ be a sum for the family of R-modules $\{A_i\}_{i \in I}$. If $V : R \to R'$ is a ring morphism and B is any R'-module, then the family $(f_i \otimes \mathrm{id}_B : A_i \underset{R}{\otimes} B \to A \underset{R}{\otimes} B\}_{i \in I}$ is a sum of the family $\{A_i \underset{R}{\otimes} B\}_{i \in I}$ of R'-modules.

PROOF: We have to show that for each R'-module C, the morphism of R'-modules $\varphi : \mathrm{Hom}_{R'}(A \underset{R}{\otimes} B, C) \to \prod_{i \in I} \mathrm{Hom}_{R'}(A_i \underset{R}{\otimes} B, C)$, given by $\varphi(g) = \{g(f_i \otimes \mathrm{id}_B)\}_{i \in I}$, is an isomorphism of R-modules. We do know that $\psi : \mathrm{Hom}_{R'}(A \underset{R}{\otimes} B, C) \to \mathrm{Hom}_R(A, \mathrm{Hom}_{R'}(B, C))$ is an isomorphism of R-modules. Because A is a sum of the A_i's we have the R'-module isomorphism $\varphi' : \mathrm{Hom}_{R'}(A, \mathrm{Hom}_R(B, C)) \to \prod \mathrm{Hom}_R(A_i, \mathrm{Hom}_{R'}(B, C))$ given by $\varphi'(h) = \{hf_i\}_{i \in I}$. Applying Example 5.5, we also have the R'-module isomorphisms $\psi_i^{-1} : \mathrm{Hom}_R(A_i, \mathrm{Hom}_{R'}(B, C)) \to \mathrm{Hom}_{R'}(A_i \underset{R}{\otimes} B, C)$ for each i in I. This gives the R'-module isomorphism

$$\prod_{i \in I} \psi_i^{-1} : \prod_{i \in I} \mathrm{Hom}_R(A_i, \mathrm{Hom}_{R'}(B, C)) \to \prod_{i \in I} \mathrm{Hom}_{R'}(A_i \underset{R}{\otimes} B, C)$$

defined by

$$\prod_{i \in I} \psi_i^{-1}(\{h_i\}_{i \in I}) = \{\psi_i^{-1}(h_i)\}_{i \in I}$$

Hence, the proposition is established if we show that the morphism φ: $\mathrm{Hom}_{R'}(A \underset{R}{\otimes} B, C) \to \prod_{i \in I} \mathrm{Hom}_{R'}(A_i \underset{R}{\otimes} B, C))$ is the composition of the isomorphisms

$$\mathrm{Hom}_{R'}(A \underset{R}{\otimes} B, C) \xrightarrow{\psi} \mathrm{Hom}_R(A, \mathrm{Hom}_{R'}(B, C)) \xrightarrow{\varphi'}$$

$$\prod_{i \in I} \mathrm{Hom}_R(A_i, \mathrm{Hom}_{R'}(B, C)) \xrightarrow{\prod_{i \in I} \varphi_i^{-1}} \prod_{i \in I} \mathrm{Hom}_{R'}(A_i \underset{R}{\otimes} B, C)$$

Using the following observations, the reader should be able to write a complete proof of this fact. The proof boils down to showing that $\psi_i^{-1}(\psi(g)f_i) = g(f_i \otimes \mathrm{id}_B)$ for all i in I. For any h_i in $\mathrm{Hom}_R(A_i, \mathrm{Hom}_{R'}(B, C))$ we have $\psi_i^{-1}(h_i)(a_i \otimes b) = [h_i(a_i)](b)$. Hence, $\psi_i^{-1}(\psi(g)f_i)(a_i \otimes b) = [\psi(g)f_i(a_i)](b) = [\psi(g) \times (f_i(a_i))](b) = g(f_i(a_i) \otimes b) = g(f_i \otimes \mathrm{id}_B)(a_i \otimes b)$.

Corollary 5.9
If $v: R \to R'$ is a ring morphism and F is a free R-module with basis B, then $R' \underset{R}{\otimes} F$ is a free R'-module and the set $\{1 \otimes b\}_{b \in B}$ is a basis for the R'-module $R' \underset{R}{\otimes} F$.

PROOF: Using Example 11.8 of Chapter 6, this follows directly from Proposition 5.8.

As an application of this corollary, we have the following.

Proposition 5.10
Let $v: R \to R'$ be a ring morphism. If M is a projective R-module, then $R' \underset{R}{\otimes} M$ is a projective R'-module.

PROOF: Because a module is projective if and only if it is a summand of a free module, there is a free R-module F such that $F = M \amalg N$ for some R-module N. By Proposition 5.8 we know that $R' \underset{R}{\otimes} F = R' \underset{R}{\otimes} M \amalg R' \underset{R}{\otimes} N$. Hence, $R' \underset{R}{\otimes} M$ is a summand of the free R'-module $R' \underset{R}{\otimes} F$. Thus, $R' \underset{R}{\otimes} M$ is a projective R'-module.

We conclude this section by showing that if $A' \to A \to A'' \to 0$ is an exact sequence of R-modules and B is an R-module, then $A' \underset{R}{\otimes} B \to A \underset{R}{\otimes} B \to A'' \underset{R}{\otimes} B \to 0$ is also an exact sequence of R-modules.

Suppose $v: R \to R'$ is a ring morphism. If $f: A_1 \to A_2$ is a morphism of R-modules and B and C are R'-modules, the following diagram of R'-modules is commutative:

$$\begin{array}{ccc} \mathrm{Hom}_{R'}(B \underset{R}{\otimes} A_2, C) & \xrightarrow{\psi_2} & \mathrm{Hom}_R(A_2, \mathrm{Hom}_{R'}(B, C)) \\ \downarrow{\scriptstyle \mathrm{Hom}_{R'}(\mathrm{id}_B \otimes f, C)} & & \downarrow{\scriptstyle \mathrm{Hom}_R(f, \mathrm{Hom}_{R'}(B, C))} \\ \mathrm{Hom}_{R'}(B \underset{R}{\otimes} A_1, C) & \xrightarrow{\psi_1} & \mathrm{Hom}_R(A_1, \mathrm{Hom}_{R'}(B, C)) \end{array}$$

We now use this observation to prove the following.

Theorem 5.11
Let $v: R \to R'$ be a ring morphism. If $A' \xrightarrow{f} A \xrightarrow{g} A'' \to 0$ is an exact sequence of R-modules and B is any R'-module, then the sequence of R'-modules

$$B \underset{R}{\otimes} A' \xrightarrow{\mathrm{id}_B \otimes f} B \underset{R}{\otimes} A \xrightarrow{\mathrm{id}_B \otimes g} B \otimes A'' \longrightarrow 0$$

is exact.

PROOF: In order to show that the sequence of R'-modules

$$B \underset{R}{\otimes} A' \xrightarrow{\mathrm{id}_B \otimes f} B \underset{R}{\otimes} A \xrightarrow{\mathrm{id}_B \otimes g} B \underset{R}{\otimes} A'' \longrightarrow 0$$

is exact, it suffices to show that for every R'-module C, the sequence

$$0 \longrightarrow \mathrm{Hom}_{R'}(B \otimes A'', C) \longrightarrow \mathrm{Hom}_{R'}(B \otimes A, C) \longrightarrow \mathrm{Hom}_{R'}(B \otimes A', C)$$

is exact (see Chapter 6, 4.9). By our previous remark we have the isomorphism of sequences

$$\begin{array}{ccccc} 0 \to \mathrm{Hom}_{R'}(B \underset{R}{\otimes} A'', C) & \to & \mathrm{Hom}_{R'}(B \underset{R}{\otimes} A, C) & \to & \mathrm{Hom}_{R'}(B \underset{R}{\otimes} A', C) \\ \downarrow \psi'' & & \downarrow \psi & & \downarrow \psi' \\ 0 \to \mathrm{Hom}_R(A'', \mathrm{Hom}_{R'}(B, C)) & \to & \mathrm{Hom}_R(A, \mathrm{Hom}_{R'}(B, C)) & \to & \mathrm{Hom}_R(A'', \mathrm{Hom}_{R'}(B, C)) \end{array}$$

Because $A' \to A \to A'' \to 0$ is exact, the bottom row of this diagram is exact. Hence, the top row is exact.

In the following example we show that there are exact sequences $0 \to A' \to A \to A'' \to 0$ such that for some B, the sequence $0 \to B \underset{R}{\otimes} A' \to B \underset{R}{\otimes} A \to B \underset{R}{\otimes} A'' \to 0$ is not exact.

Example 5.12 Consider the exact sequence of $\mathbf{Z}$-modules $0 \to \mathbf{Z} \xrightarrow{f} \mathbf{Z} \to \mathbf{Z}/2\mathbf{Z} \to 0$ where f is multiplication by 2. Let $B = \mathbf{Z}/2\mathbf{Z}$. Then the morphism $\mathrm{id}_B \otimes f: B \underset{\mathbf{Z}}{\otimes} \mathbf{Z} \to B \underset{\mathbf{Z}}{\otimes} \mathbf{Z}$ is the zero morphism because $B \underset{\mathbf{Z}}{\otimes} \mathbf{Z} \approx B$ and $\mathrm{id}_B \otimes f$ is multiplication by 2. Hence, $\mathrm{id}_B \otimes f$ is not a monomorphism.

Example 5.13 Let I be an ideal in R and $k: R \to R/I$ the canonical ring morphism. If $A' \to A \to A'' \to 0$ is an exact sequence of R-modules, then $A'/IA' \to A/IA \to A''/IA'' \to 0$ is an exact sequence of R/I-modules.

PROOF: Recall that $R/I \underset{R}{\otimes} X = X/IX$ for all R-modules X by Example 4.5.

Example 5.14 Let S be a multiplicative subset of a ring R. Suppose I is an ideal of R and A is an R-module. From the exact sequence of R-modules $0 \to IA \to A \xrightarrow{k} A/IA \to 0$ we obtain the exact sequence of R_S-modules $0 \to (IA)_S \to A_S \xrightarrow{k_S} (A/IA)_S \to 0$. The reader has already shown in the course of the proof of Proposition 3.2 that $(IA)_S = I_S A_S$. Hence, the surjective R_S-morphism $A_S \to (A/IA)_S$ induces an isomorphism $A_S/I_S A_S \approx (A/IA)_S$. It is not hard to show that the inverse, f, of this isomorphism is given by $f(k(a)/s) = k_S(a/s)$.

Example 5.15 Let M be a finitely generated R-module such that $M/\mathfrak{P}M = 0$ for all maximal ideals $\mathfrak{P}$ of R. Then $M = 0$.

PROOF: Clearly, $R_{\mathfrak{P}} \underset{R}{\otimes} (M/PM) = 0$ for every maximal ideal $\mathfrak{P}$ of R. Because $R_{\mathfrak{P}} \underset{R}{\otimes} M/\mathfrak{P}M = M_{\mathfrak{P}}/P_{\mathfrak{P}}M_{\mathfrak{P}}$, we have $M_{\mathfrak{P}}/P_{\mathfrak{P}}M_{\mathfrak{P}} = 0$ for each $\mathfrak{P}$. However, M is a finitely generated $R_{\mathfrak{P}}$-module and thus $M_{\mathfrak{P}}$ is a finitely generated $R_{\mathfrak{P}}$-module. Hence, $M_{\mathfrak{P}} = 0$ for every maximal ideal $\mathfrak{P}$ of R. By Theorem 3.8, we have $M = 0$.

6. LOCALLY FREE MODULES

In this section we give some further indications of how localizations can be used in commutative ring theory. Our discussion is centered on characterizing, in terms of localizations, when finitely generated modules over noetherian rings are projective.

We begin with the following.

Proposition 6.1
Let R be a local ring. A finitely generated R-module M is projective if and only if it is a free R-module.

PROOF: Obviously if M is a free R-module, it is a projective R-module.

Suppose M is a projective R-module. Because $R/\text{rad}(R)$ is a field, we know by Chapter 8, Basic Properties 3.6 that M is a free R-module.

Corollary 6.2
If M is a finitely generated projective R-module, then $M_{\mathfrak{P}}$ is a finitely generated free $R_{\mathfrak{P}}$-module for every prime ideal $\mathfrak{P}$ of R.

PROOF: Because M is a finitely generated projective R-module, it follows that $R_{\mathfrak{P}} \underset{R}{\otimes} M = M_{\mathfrak{P}}$ is a finitely generated projective $R_{\mathfrak{P}}$-module. But by Proposition 6.1 this implies that $M_{\mathfrak{P}}$ is a free $R_{\mathfrak{P}}$-module.

This suggests the following.

Definition
An R-module M is said to be **locally free** if $M_{\mathfrak{P}}$ is a free $R_{\mathfrak{P}}$-module for each prime ideal $\mathfrak{P}$ of R.

Having seen that a finitely generated projective R-module is locally free, we now consider under what conditions a locally free R-module is projective.

In order to simplify notation, we assume throughout the rest of this section that all rings are noetherian.

Given two R-modules A and B and a multiplicative subset S of R, we have already described the morphism of R-modules $\text{Hom}_R(A, B) \to \text{Hom}_{R_S}(A_S, B_S)$ given by $f \mapsto f_S$. This morphism enables us to define an R_S-morphism $\text{Hom}_R(A, B)_S \to \text{Hom}_{R_S}(A_S, B_S)$ defined by $f/s \mapsto 1/s \cdot f_S$. Our aim now is to show that if A is a finitely generated R-module, then $\text{Hom}_R(A, B)_S \to \text{Hom}_{R_S}(A_S, B_S)$ is an R_S isomorphism for all R-modules B.

Suppose $A = R$. Using the fact that the morphism $\mathrm{Hom}_R(R, B) \to B$ given by $f \mapsto f(1)$ is an isomorphism of R-modules, it is not hard to see that the following diagram is commutative:

$$\begin{array}{ccc} \mathrm{Hom}_R(R, B)_S & \longrightarrow & \mathrm{Hom}_{R_S}(R_S, B_S) \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\beta} \\ B_S & = & B_S \end{array}$$

where $\alpha(f/s) = f(1)/s$ and $\beta(g) = g(1)$. Because α and β are isomorphisms, it follows that $\mathrm{Hom}(R, B)_S \to \mathrm{Hom}_{R_S}(R_S, B_S)$ is an isomorphism.

Now suppose A is a finitely generated free R-module; that is, $A = \coprod_{i=1}^{n} R_i$ where each $R_i = R$. Because $A_S = R_S \otimes A = R_S \otimes \coprod_{i=1}^{n} R_i = \coprod_{i=1}^{n} R_S \otimes R_i = \coprod_{i=1}^{n} (R_S)_i$, it is easy to show that $\mathrm{Hom}_R(A, B)_S \to \mathrm{Hom}_{R_S}(A_S, B_S)$ is an isomorphism in this case. We now use this to prove the following.

Proposition 6.3
Let A be a finitely generated R-module and B an arbitrary R-module. Then the R_S-morphism $\mathrm{Hom}_R(A, B)_S \to \mathrm{Hom}_{R_S}(A_S, B_S)$ is an isomorphism.

PROOF: Because A is finitely generated, there is an exact sequence $0 \to K \to F \to A \to 0$ where F is a finitely generated free R-module. But R is a noetherian ring and F is a finitely generated R-module, so K is also a finitely generated R-module. Hence, there is an epimorphism $F' \to K$ with F' a finitely generated free R-module. Thus, the composition $F' \to K \to F$ yields the exact sequence $F' \to F \to A \to 0$. From this it follows that the sequence of R_S-modules $F'_S \to F_S \to A_S \to 0$ is also exact.

Using these exact sequences, we obtain for any R-module B the following commutative diagram

$$\begin{array}{ccccccc} 0 \longrightarrow \mathrm{Hom}_R(A, B)_S & \xrightarrow{\sigma} & \mathrm{Hom}_R(F, B)_S & \xrightarrow{\tau} & \mathrm{Hom}_R(F', B)_S \\ \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle\gamma} \\ 0 \longrightarrow \mathrm{Hom}_{R_S}(A_S, B_S) & \xrightarrow{\sigma'} & \mathrm{Hom}_{R_S}(F_S, B_S) & \xrightarrow{\tau'} & \mathrm{Hom}_{R_S}(F'_S, B_S) \end{array}$$

with exact rows. Because F and F' are finitely generated free R-modules, we know by our discussion preceding the statement of this proposition, that β and γ are isomorphisms. It follows immediately that α is a monomorphism.

To show that α is surjective, let x be in $\mathrm{Hom}_{R_S}(A_S, B_S)$. Then $\tau\beta^{-1}\sigma'(x) = \gamma^{-1}\tau'\sigma'(x) = 0$. Hence, $\beta^{-1}\sigma'(x) = \sigma(y)$ for some y in $\mathrm{Hom}_R(A, B)_S$. $\sigma'\alpha(y) = \beta\sigma(y) = \beta\beta^{-1}\sigma'(x) = \sigma'(x)$. This implies that $x = \alpha(y)$ since σ' is injective. Thus, α is surjective and hence an isomorphism.

Theorem 6.4
If R is a noetherian ring, then a finitely generated R-module is projective if and only if it is locally free.

PROOF: We have already seen that every finitely generated projective module is locally free.

Suppose A is a finitely generated locally free R-module. To show that A is projective, it suffices to show that if $B \to B''$ is an epimorphism of R-modules, then $\mathrm{Hom}_R(A, B) \to \mathrm{Hom}_R(A, B'')$ is also an epimorphism. By Corollary 3.10 we know that $\mathrm{Hom}_R(A, B) \to \mathrm{Hom}_R(A, B'')$ is an epimorphism if $\mathrm{Hom}_R(A, B)_{\mathfrak{P}} \to \mathrm{Hom}_R(A, B'')_{\mathfrak{P}}$ is an epimorphism for every maximal ideal $\mathfrak{P}$ of R. Because A is finitely generated we have for every maximal ideal $\mathfrak{P}$ of R a commutative diagram

$$\begin{array}{ccc} \mathrm{Hom}_R(A, B)_{\mathfrak{P}} & \longrightarrow & \mathrm{Hom}_R(A, B'')_{\mathfrak{P}} \\ \downarrow & & \downarrow \\ \mathrm{Hom}_{R_{\mathfrak{P}}}(A_{\mathfrak{P}}, B_{\mathfrak{P}}) & \longrightarrow & \mathrm{Hom}_{R_{\mathfrak{P}}}(A_{\mathfrak{P}}, B''_{\mathfrak{P}}) \end{array}$$

with the vertical arrows isomorphisms by virtue of Proposition 6.3. Because $A_{\mathfrak{P}}$ is a free $R_{\mathfrak{P}}$-module and $B_{\mathfrak{P}} \to B''_{\mathfrak{P}}$ is an epimorphism, the bottom horizontal arrow is an epimorphism. Hence, the top horizontal arrow is an epimorphism. This finishes the proof of the theorem.

We now apply this result to obtain a result which is interesting in its own right and will be used later.

Theorem 6.5

Let R be a commutative noetherian ring, A a finitely generated projective R-module, and G and H finitely generated R-modules such that $A \amalg G \approx H \amalg G$. Then H is a projective R-module.

PROOF: We shall show that $R_{\mathfrak{P}} \underset{R}{\otimes} H$ is a free R-module for every prime ideal $\mathfrak{P}$ of R. By Theorem 6.4, this will show that H is projective.

To show that $R_{\mathfrak{P}} \underset{R}{\otimes} H$ is $R_{\mathfrak{P}}$-free, we may assume that R is a local ring and that A is consequently a free R-module. Under this additional assumption, we want to show that H is a free R-module.

Next, let us suppose that G contains no free submodules as a summand, and prove that H is free under this assumption. Later, we will be able to remove this restriction from G.

With this assumption on G, we notice that if $f : G \to R$ is any morphism from G to R, then $\mathrm{Im}\, f \subset J$ where J is the maximal ideal of R (remember that we are assuming that R is a local ring). If not, then $\mathrm{Im}\, f$ would contain a unit because J is the set of all nonunits of R, and hence be all of R. In that case, f would be an epimorphism and R would be a summand of G, contrary to our assumption about G. Thus, $\mathrm{Im}\, f \subset J$. As a corollary, if $g : G \to F$ is a morphism where F is any free module of finite rank, we have $\mathrm{Im}\, g \subset JF$. To see this, the reader should show that if $\mathrm{Im}\, g \not\subset JF$, then one of the coordinate projections from F to R carries $\mathrm{Im}\, g$ onto R.

Now consider the morphism $puj_2 : G \to A$ where $p : A \amalg G \to A$ is the projection onto A, $u : H \amalg G \to A \amalg G$ is our given isomorphism, and $j_2 : G \to H \amalg G$ is the injection. Then $\mathrm{Im}(puj_2) \subset JA$ by what we have just said. On the other hand, if $j_1 : H \to H \amalg G$ is the injection of H into $H \amalg G$, then clearly $A =$

$\mathrm{Im}(puj_1) + \mathrm{Im}(puj_2)$. Hence, letting $A' = \mathrm{Im}(puj_1)$, we have $A = A' + JA$, which is equivalent to saying that $A/A' = J(A/A')$. Because $J = \mathrm{rad}(R)$, it follows that $A/A' = 0$. Hence, $A = A'$.

Knowing that $A' = A$, we see that the morphism $puj_1 : H \to A$ is an epimorphism. Because A is free, $H \approx A \amalg Q$ where $Q = \mathrm{Ker}\, puj_1$. Thus, $A \amalg G \approx H \amalg G \approx A \amalg Q \amalg G$ and, tensoring with R/J, we see that $Q/JQ = (0)$. For we have $(A \amalg G) \underset{R}{\otimes} R/J \approx A/JA \amalg G/JG$ and $(A \amalg Q \amalg G) \underset{R}{\otimes} R/J \approx A/JA \amalg Q/JQ \amalg G/JG$. Because all of these modules are finite-dimensional vector spaces over R/J, it follows, by adding dimensions of vector spaces, that the dimension of Q/JQ is zero. Thus, $Q/JQ = 0$. But Q, being a summand of H, is finitely generated so that if $Q/JQ = (0)$, we have $Q = (0)$. Thus, given our assumption on G, we see that $puj_1 : H \to A$ is an isomorphism so that H is free.

What happens now if we remove the condition on G that it contain no free submodule as a summand? If G contains a nonzero free summand, then the set of nonzero free summands is not empty, so there is a maximal free summand G' because G is a noetherian module. Then $G = G' \amalg G''$ and clearly G'' cannot contain a free summand. We now have $(A \amalg G') \amalg G'' = A \amalg G \approx H \amalg G \approx (H \amalg G') \amalg G''$, and $A \amalg G'$ is a finitely generated free module. The modules $H \amalg G'$ and G'' are also finitely generated. By what we have already shown, we know that $H \amalg G'$ and $A \amalg G'$ are isomorphic (because G'' contains no free summand) and so $H \amalg G'$ is free. Because H is a summand of a free module of finite rank, it is a finitely generated projective module and hence free because R is a local ring. This is the result that we wanted and Theorem 6.4 tells us that H is a projective R-module.

EXERCISES

(1) Prove that the characteristic of a local ring is either zero or a power of a prime, where the characteristic of a local ring R is the smallest nonnegative integer n such that $n1 = 0$ in R.

(2) Show that the assumption, in Example 5.15, that the module M be finitely generated is necessary.

(3) Let R be a commutative ring, J an ideal, and $M = R/J$.

(a) Prove that if J is generated as an ideal by a set of idempotents, then $M_{\mathfrak{P}} = 0$ or $M_{\mathfrak{P}} = R_{\mathfrak{P}}$ for every maximal ideal $\mathfrak{P}$ of R.

(b) Prove that if J is finitely generated, then M is a projective R-module and hence J is a principal ideal generated by a single idempotent.

(c) Let K be a field and let $R = \prod_{i \in \mathbf{N}} K_i$ where $K_i = K$ for each i in $\mathbf{N}$. For each finite strictly increasing set of positive integers $i_1 < i_2 < \cdots < i_n$, let $\epsilon_{i_1 \ldots i_n}$ be the element of R all of whose coordinates are zero except those in the places $i_1, \ldots, i_n$ in which case the coordinates are 1. Let J be the ideal generated by all the elements $\epsilon_{i_1 \ldots i_n}$.

(i) Prove that each element $\epsilon_{i_1 \ldots i_n}$ is an idempotent.

(ii) Prove that J is a proper ideal of R.

(iii) Prove that J is not a principal ideal.

(d) Conclude from the above that the module $M = R/J$ is a finitely generated free R-module which is not R-projective. This shows that Theorem 6.4 would not be true if we did not assume that R is noetherian.

(4) An R-module A is **finitely presented** if there are free finitely generated F_1 and F_0, and morphisms $f: F_1 \to F_0$, $g: F_0 \to A$ such that $F_1 \xrightarrow{f} F_0 \xrightarrow{g} A \to 0$ is exact.

(a) Prove Proposition 6.3 assuming that R is a commutative, not necessarily noetherian ring, and that the module A is finitely presented.

(b) Prove that if R is any commutative ring and A is a finitely presented R-module, then A is R-projective if and only if A is locally free.

(5) (a) Prove that if A is a finitely presented R-module, and $M \xrightarrow{f} A$ is an epimorphism with M a finitely generated R-module, then Ker f is a finitely generated R-module.

(b) Prove that an R-module C is finitely presented if and only if there is an exact sequence $A \to B \to C \to 0$ of R-modules with A and B finitely presented R-modules.

(c) Show that a ring R is left noetherian if and only if every finitely generated R-module is finitely presented.

(6) Let R be a commutative ring and A an R-module.

(a) Let I be an ideal of R and let $I \underset{R}{\otimes} A \to A$ be the composite morphism $I \underset{R}{\otimes} A \xrightarrow{i \otimes A} R \underset{R}{\otimes} A \to A$ where $i: I \to R$ is the inclusion and the morphism $R \underset{R}{\otimes} A \to A$ is the natural isomorphism. Prove that if the morphism $I \underset{R}{\otimes} A \to A$ is a monomorphism for every finitely generated ideal I, then it is a monomorphism for every ideal I of R.

(b) Let B' be a submodule of an R-module B and let x be an element in the kernel of the morphism $A \underset{R}{\otimes} B' \xrightarrow{A \otimes i} A \underset{R}{\otimes} B$, where $i: B' \to B$ is the inclusion. Prove that there is a submodule $\bar{B}$ of B containing B' such that

(i) $\bar{B}/B'$ is finitely generated and

(ii) x is in the kernel of the morphism $A \underset{R}{\otimes} B' \xrightarrow{A \otimes j} A \underset{R}{\otimes} \bar{B}$, where $j: B' \to \bar{B}$ is the inclusion. [*Hint*: Use the explicit construction of $A \underset{R}{\otimes} B$ in terms of generators and relations to describe what it means for the element x to be in $\text{Ker}(A \underset{R}{\otimes} B' \to A \underset{R}{\otimes} B)$.]

(c) Assume that the morphism $I \underset{R}{\otimes} A \to A$ is a monomorphism for every ideal I of R. Prove that if $0 \to B' \to B \to B'' \to 0$ is exact and B'' is cyclic, then $0 \to A \underset{R}{\otimes} B' \to A \underset{R}{\otimes} B$ is exact. [*Hint*: Use the fact that B'' is cyclic to show that we have the following diagram with exact rows and columns:

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & 0 & \to & I & = & I & & \\
 & & & & \downarrow & & \downarrow & & \\
0 & \to & B' & \to & B' \amalg R & \to & R & \to & 0 \\
 & & \| & & \downarrow & & \downarrow & & \\
0 & \to & B' & \to & B & \to & B'' & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

where $I = \text{ann}(B'')$. Tensoring each term in the diagram with A and using the fact that the middle row splits, show that $\text{Ker}(A \underset{R}{\otimes} I \to A) \to A \underset{R}{\otimes} B' \to A \underset{R}{\otimes} B$ is exact. Then use the fact that $A \underset{R}{\otimes} I \to A$ is a monomorphism.]

(d) Prove that if the morphism $I \underset{R}{\otimes} A \to A$ is a monomorphism for every finitely generated ideal I of R, then for every monomorphism $0 \to B' \to B$, the morphism $A \underset{R}{\otimes} B' \to A \underset{R}{\otimes} B$ is a monomorphism. [*Hint*: Use part (a) and induction on the number of generators of $\text{Coker}(B' \to B)$, assuming $\text{Coker}(B' \to B)$ is finitely generated. Then use part (b) to go from the assumption that $\text{Coker}(B' \to B)$ is finitely generated, to the general case.]

(7) Let R be a commutative ring, let A be an R-module, and let $\bar{B}$ be a sequence (finite or infinite) of R-modules and R-morphisms: $\bar{B} = \cdots \to B_i \xrightarrow{f_i} B_{i+1} \xrightarrow{f_{i+1}} B_{i+2} \to \cdots$. The module A is a **flat** R-module if the sequence $A \underset{R}{\otimes} \bar{B}$ is exact whenever the sequence $\bar{B}$ is exact. By $A \underset{R}{\otimes} \bar{B}$ we mean, of course, the sequence $\cdots \to A \underset{R}{\otimes} B_i \xrightarrow{A \otimes f_i} A \underset{R}{\otimes} B_{i+1} \xrightarrow{A \otimes f_{i+1}} A \underset{R}{\otimes} B_{i+2} \to \cdots$. The R-module is **faithfully flat** if it is flat and has the additional property that $A \underset{R}{\otimes} \bar{B}$ exact implies $\bar{B}$ is exact.

(a) Prove that the following conditions on A are equivalent:
- (i) A is a flat R-module.
- (ii) For every R-monomorphism $B' \to B$, the morphism $A \underset{R}{\otimes} B' \to A \underset{R}{\otimes} B$ is a monomorphism.
- (iii) For every finitely generated ideal I of R, the morphism $I \underset{R}{\otimes} A \to A$ is a monomorphism.
- (iv) If $\Sigma_{i=1}^{n} r_i a_i = 0$ with r_i in R and a_i in A for $i = 1, \ldots, n$, then there exist elements $b_1, \ldots, b_m$ in A and elements s_{ij} in R with $1 \le i \le n$ and $1 \le j \le m$ such that $\Sigma_{i=1}^{n} r_i s_{ij} = 0$ for all j, and $a_i = \Sigma_j s_{ij} b_j$ for all i.

(b) Prove that the following conditions on A are equivalent:
- (i) A is a faithfully flat R-module.
- (ii) A is a flat R-module and $A \underset{R}{\otimes} B \neq 0$ for every nonzero R-module B.
- (iii) A is a flat R-module and, for every maximal ideal $\mathfrak{B}$ of R, $A/\mathfrak{B}A \neq 0$.

(8) Let $f: R \to R'$ be a morphism of commutative rings. If R', considered as an R-module via f, is a flat R-module, then f is called a **flat morphism**. Similarly, if R' is a faithfully flat R-module, f is called a **faithfully flat morphism**.

(a) Suppose that f is a flat (faithfully flat) morphism. Prove that if B is a flat R'-module, then it is a flat (faithfully flat) R-module.

(b) Let A be a flat (faithfully flat) R-module. Prove that $R' \underset{R}{\otimes} A$ is a flat (faithfully flat) R'-module.

(c) Let S be a multiplicative subset of R and $f: R \to R_S$ the canonical morphism. Prove that f is a flat morphism.

(9) Let R be a local ring with maximal ideal J and residue class field $K = R/J$. Let A be a finitely generated R-module.

(a) Prove that there is a finitely generated free R-module F and an epimorphism $g: F \to A$ such that $\text{Ker } g$ is contained in JF. [*Hint*: Let A/JA be generated by

the cosets of $a_1, \ldots, a_n$ in A, and let F be the free module generated by elements $x_1, \ldots, x_n$. Define g by setting $g(x_i) = a_i$ for $i = 1, \ldots, n$.]

(b) Prove that if A is a flat R-module, I is any ideal of R, and $0 \to A'' \to A' \to A \to 0$ is an exact sequence of R-modules, then $0 \to A''/IA'' \to A'/IA' \to A/IA \to 0$ is exact. Here we need not assume that R is local nor that A is finitely generated. [*Hint*: Use Exercise 7(a)(iv).]

(c) Using parts (a) and (b), prove that if R is local and A is finitely generated, then A flat implies A is free.

(10) Let R be a commutative noetherian ring and A a finitely generated R-module. Prove that A is flat if and only if A is projective. Is the condition that R be noetherian essential here?

(11) Let R be any commutative ring and A and R-module. Prove that A is flat if and only if it is locally flat (that is, $A \underset{R}{\otimes} R_{\mathfrak{B}}$ is $R_{\mathfrak{B}}$-flat for every maximal ideal $\mathfrak{B}$ of R).

(12) Let $f: R \to R'$ be a faithfully flat ring morphism and A an R-module. Assume, also, that R and R' are commutative rings (although this restriction will not be necessary once we have defined tensor products over noncommutative rings).

(a) Prove that if $R' \underset{R}{\otimes} A$ is a finitely generated R'-module, then A is a finitely generated R-module.

(b) Prove that if $R' \underset{R}{\otimes} A$ is a finitely presented R'-module, then A is a finitely presented R-module.

(13) Let R be a commutative ring and M a finitely generated projective R-module.

(a) Prove that M is finitely presented.

(b) If $\mathfrak{P}$ is a prime ideal of R, prove that there are elements $m_1, \ldots, m_n$ in M such that $\{1 \otimes m_1, \ldots, 1 \otimes m_n\}$ is a basis of $R_{\mathfrak{P}} \underset{R}{\otimes} M$.

(c) Let F be the free R-module generated by $\{x_1, \ldots, x_n\}$, let $f: F \to M$ be defined by $f(x_i) = m_i$, and let $L = \operatorname{Coker} f$. Show that $R_{\mathfrak{P}} \underset{R}{\otimes} L = 0$ and that there is an element y_1 not in $\mathfrak{P}$ such that $y_1 \lambda = 0$ for all λ in L.

(d) Let $R_{(y_1)}$ denote the ring of quotients of R with respect to the multiplicative subset $\{y_1^n\}_{n \in \mathbb{N}}$. Prove that $R_{(y_1)} \underset{R}{\otimes} L = 0$, where $L = \operatorname{Coker} f$ in part (c).

(e) Letting $K = \operatorname{Ker} f$, prove that $R_{(y_1)} \underset{R}{\otimes} K$ is a finitely generated $R_{(y_1)}$-module and that $R_{\mathfrak{P}} \underset{R}{\otimes} K = 0$.

(f) Prove that, if $\mathfrak{P}' = \mathfrak{P} R_{(y_1)}$, then $R_{\mathfrak{P}} = (R_{(y_1)})_{\mathfrak{P}'}$. Hence, show that there is an element y_2 in $R_{(y_1)}$ such that y_2 is not in $\mathfrak{P}'$ and $y_2 z = 0$ for all z in $R_{(y_1)} \underset{R}{\otimes} K$.

[*Hint*: Use the fact that $0 = R_{\mathfrak{P}} \underset{R}{\otimes} K = (R_{(y_1)})_{\mathfrak{P}'} \underset{R_{(y_1)}}{\otimes} (R_{(y_1)} \underset{R}{\otimes} K)$ and that $R_{(y_1)} \underset{R}{\otimes} K$ is a finitely generated $R_{(y_1)}$-module.]

(g) (i) Prove that for some positive integer ν, $y_1^{\nu} y_2 = y_2'$ with y_2' in R.

(ii) Prove that y_2' is not in $\mathfrak{P}$ and that for each element c in K, there is an integer μ such that $y_1^{\mu} y_2' c = 0$.

(iii) Let $y = y_1 y_2'$. Prove that $R_{(y)} \underset{R}{\otimes} L = 0 = R_{(y)} \underset{R}{\otimes} K$.

(h) Conclude from all of the above that for every maximal (or prime) ideal $\mathfrak{P}$ of R, there is an element y in $R - \mathfrak{P}$ such that $R_{(y)} \underset{R}{\otimes} M$ is a free $R_{(y)}$-module of finite rank.

(14) Let R be a commutative ring and M an R-module. Suppose that for each maximal ideal $\mathfrak{P}$ of R, there is an element y in $R - \mathfrak{P}$ such that $R_{(y)} \underset{R}{\otimes} M$ is a free $R_{(y)}$-module of finite rank.

(a) Prove that for every maximal ideal $\mathfrak{P}$ of R, $R_{\mathfrak{P}} \otimes M$ is a free $R_{\mathfrak{P}}$-module of finite rank.

(b) Prove that for every prime ideal $\mathfrak{Q}$ of R, $R_{\mathfrak{Q}} \underset{R}{\otimes} M$ is a free $R_{\mathfrak{Q}}$-module of finite rank.

(c) Show that if $R_{(y)} \underset{R}{\otimes} M$ is free of rank n, then $R_{\mathfrak{Q}} \underset{R}{\otimes} M$ is free of rank n for all prime ideals $\mathfrak{Q}$ of R not containing y.

(d) For each maximal ideal $\mathfrak{P}$ of R, let $y(\mathfrak{P})$ be an element in $R - \mathfrak{P}$ such that $R_{(y(\mathfrak{P}))} \underset{R}{\otimes} M$ is free of finite rank. Let I be the ideal of R generated by all the elements $y(\mathfrak{P})$ as $\mathfrak{P}$ runs through all maximal ideals of R. Prove that I is not a proper ideal of R, that is, $I = R$, and consequently there are finitely many maximal ideals, say $\mathfrak{P}_1, \ldots, \mathfrak{P}_t$ such that the ideal generated by $y_1, \ldots, y_t$ [where $y_i = y(\mathfrak{P}_i)$] is R.

(e) Let $\phi_i : R \to R_{(y_i)}$ be the canonical ring morphism for $i = 1, \ldots, t$, let $S = R_{(y_1)} \times \cdots \times R_{(y_t)}$, and let $\phi : R \to S$ be the ring morphism determined by the ϕ_i. Prove that ϕ is a faithfully flat ring morphism. [*Hint*: Use part (b) (iii) of Exercise 7.]

(f) Using the fact that $R_{(y_i)} \underset{R}{\otimes} M$ is a free $R_{(y_i)}$-module of finite rank, prove that $S \underset{R}{\otimes} M$ is a finitely generated S-module.

(g) Conclude from the above that M is a finitely generated R-module.

(h) Let $n_i = \operatorname{rank}(R_{(y_i)} \underset{R}{\otimes} M)$ and let $n = \max(n_i)$.

- **(i)** Prove that $S \underset{R}{\otimes} M$ is a summand of S^n where S^n denotes the sum of n copies of S and $S = R_{(y_1)} \times \cdots \times R_{(y_t)}$.
- **(ii)** Prove that $S \underset{R}{\otimes} M$ is a finitely presented S-module.
- **(iii)** Conclude that M is a finitely presented R-module.
- **(iv)** Finally, conclude that M is a projective R-module.

(15) Let R be a commutative ring and M a finitely generated R-module. Assume that for each prime ideal $\mathfrak{Q}$ of R, $R_{\mathfrak{Q}} \underset{R}{\otimes} M$ is a free $R_{\mathfrak{Q}}$-module and that, for each $\mathfrak{Q}$, there is an element y in $R - \mathfrak{Q}$ such that the rank of $R_{\mathfrak{Q}'} \underset{R}{\otimes} M$ is constant for all prime ideals $\mathfrak{Q}'$ not containing y.

(a) Let $\mathfrak{P}$ be a maximal ideal of R and let $m_1, \ldots, m_n$ be elements of M such that $\{1 \otimes m_1, \ldots, 1 \otimes m_n\}$ is a basis for $R_{\mathfrak{P}} \underset{R}{\otimes} M$ over $R_{\mathfrak{P}}$. Let F be the free R-module with basis $\{x_1, \ldots, x_n\}$ and $f : F \to M$ the R-morphism defined by $f(x_i) = m_i$. Prove that there is an element y_1 in $R - \mathfrak{P}$ such that $R_{(y_1)} \underset{R}{\otimes} F \xrightarrow{1 \otimes f} R_{(y_1)} \underset{R}{\otimes} M$ is an epimorphism.

(b) Let y_2 in $R - P$ be such that for all primes $\mathfrak{Q}$ not containing y_2, the module $R_{\mathfrak{Q}} \underset{R}{\otimes} M$ has constant rank (namely, $n = \operatorname{rank} R_P \underset{R}{\otimes} M$). Let $y = y_1 y_2$. Prove that $R_{(y)} \underset{R}{\otimes} F \xrightarrow{1 \otimes f} R_{(y)} \underset{R}{\otimes} M$ is an epimorphism.

(c) Show that if $\mathfrak{Q}'$ is a prime ideal of $R_{(y)}$, then $Q' = QR_{(y)}$ where $\mathfrak{Q}$ is a prime ideal of R not containing y.

(d) Prove that if $\mathfrak{Q}$ is a prime ideal of R not containing y, then the morphism $R_{\mathfrak{Q}} \underset{R}{\otimes} F \xrightarrow{1\otimes f} R_{\mathfrak{Q}} \underset{R}{\otimes} M$ is an isomorphism.

(e) Prove that the morphism $R_{(y)} \underset{R}{\otimes} F \to R_{(y)} \underset{R}{\otimes} M$ is an isomorphism. This shows that for each maximal ideal $\mathfrak{P}$ of R, there is an element y in $R - \mathfrak{P}$ such that $R_{(y)} \underset{R}{\otimes} M$ is a free $R_{(y)}$-module of finite rank.

The following exercise is a summary of the preceding three exercises.

(16) Let M be a module over the commutative ring R. Prove that the following statements are equivalent:

(a) M is a finitely generated projective R-module.

(b) M is a finitely presented R-module having the property that $R_{\mathfrak{P}} \underset{R}{\otimes} M$ is a free $R_{\mathfrak{P}}$-module for every maximal ideal $\mathfrak{P}$ of R.

(c) For each maximal ideal $\mathfrak{P}$ of R there is an element y in $R - \mathfrak{P}$ such that $R_{(y)} \underset{R}{\otimes} M$ is a free $R_{(y)}$-module of finite rank.

(d) M is a finitely generated R-module such that $R_{\mathfrak{Q}} \underset{R}{\otimes} M$ is a free $R_{\mathfrak{Q}}$-module for every prime ideal $\mathfrak{Q}$ of R and, for every prime ideal $\mathfrak{Q}$, there is an element y in $R - \mathfrak{Q}$ such that the rank of $R_{\mathfrak{Q}} \underset{R}{\otimes} M$ equals the rank of $R_{\mathfrak{Q}'} \underset{R}{\otimes} M$ for all prime ideals $\mathfrak{Q}'$ not containing y.

(17) Prove Theorem 6.5 without the assumption that R is a noetherian ring. [*Hint*: Show that the module M satisfies condition (d) of Exercise 16.]

(18) Let R be a commutative ring. The support of an R-module M is the set $\operatorname{Supp}(M)$ of all prime ideals $\mathfrak{P}$ in R such that $M_{\mathfrak{P}} \neq 0$. In this exercise we develop some of the basic properties of the support of R-modules.

(a) Show that if M is an R-module, then $\operatorname{Supp}(M) = \emptyset$ if and only if $M = 0$.

(b) Show that if $\mathfrak{Q} \supset \mathfrak{P}$ are prime ideals of R and $\mathfrak{P}$ is in $\operatorname{Supp}(M)$, then $\mathfrak{Q}$ is in $\operatorname{Supp}(M)$.

(c) Show that if $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules, then $\operatorname{Supp}(M) = \operatorname{Supp}(M') \cup \operatorname{Supp}(M'')$.

(d) Show that if I is an ideal of R, then $\operatorname{Supp}(R/I)$ consists precisely of the prime ideals $\mathfrak{P}$ of R containing I.

(e) Show that if M is a finitely generated R-module, then $\operatorname{Supp}(M)$ consists precisely of the prime ideals of R containing the annihilator of M.

(f) If M is a finitely generated R-module, then $\mathfrak{P}$ is in $\operatorname{Supp}(M)$ if and only if $M_{\mathfrak{P}}/P_{\mathfrak{P}}M_{\mathfrak{P}} \neq 0$.

(g) Show that $\operatorname{Supp}(M \underset{R}{\otimes} N) = \operatorname{Supp} M \cap \operatorname{Supp} N$ if M and N are finitely generated R-modules.

(h) Let R be the ring of integers $\mathbf{Z}$ and M the $\mathbf{Z}$-module $\mathbf{Q}/\mathbf{Z}$ where $\mathbf{Q}$ is the field of rational numbers. Determine $\operatorname{Supp}(\mathbf{Q}/\mathbf{Z})$. What is the annihilator of $\mathbf{Q}/\mathbf{Z}$? Is $\operatorname{Supp}(\mathbf{Q}/\mathbf{Z})$ the set of all prime ideals of $\mathbf{Z}$ containing $\operatorname{ann}(\mathbf{Q}/\mathbf{Z})$?

(19) Let R be an arbitrary commutative ring and A a finitely presented R-module. Show that for each multiplicative subset S of R and each R-module B the R_S-morphism $\operatorname{Hom}_R(A, B)_S \to \operatorname{Hom}_{R_S}(A_S, B_S)$ is an isomorphism.

(20) Let M be a module over the commutative ring R. For each nonnegative integer n define $\overset{n}{\underset{R}{\otimes}} M$ to be the ***n*-fold tensor product** $M \underset{R}{\otimes} \cdots \underset{R}{\otimes} M$ **of** M **with itself**

by induction on n as follows:

$$\overset{0}{\underset{R}{\otimes}} M = R \quad \text{and} \quad \overset{n+1}{\underset{R}{\otimes}} M = (\overset{n}{\underset{R}{\otimes}} M) \underset{R}{\otimes} M \quad \text{for} \quad n \geq 0$$

(a) Show that for each pair n, m of nonnegative integers there are unique R-module morphisms $f_{n,m}: (\overset{n}{\underset{R}{\otimes}} M) \underset{R}{\otimes} (\overset{m}{\underset{R}{\otimes}} M) \to \overset{n+m}{\underset{R}{\otimes}} M$ satisfying:

(i) If $n = 0 = m$, then $f_{0,0}: R \underset{R}{\otimes} R \to R$ has the property $f_{0,0}(r_1 \otimes r_2) = r_1 r_2$ for all r_1 and r_2 in R.

(ii) If $n = 0$ and $m > 0$, then $f_{0,m}: R \underset{R}{\otimes} (\overset{m}{\underset{R}{\otimes}} M) \to \overset{m}{\underset{R}{\otimes}} M$ has the property $f_{0,m}(r \otimes (y_1 \otimes \cdots \otimes y_m)) = r(y_1 \otimes \cdots \otimes y_m)$ for all r in R and $y_1, \ldots, y_m$ in M.

(iii) If $n > 0$ and $m = 0$, then $f_{n,0}: (\overset{n}{\underset{R}{\otimes}} M) \underset{R}{\otimes} R \to \overset{n}{\underset{R}{\otimes}} M$ has the property $(x_1 \otimes \cdots \otimes x_n) \otimes r = r(x_1 \otimes \cdots \otimes x_n)$ for all $x_1, \ldots, x_n$ in M and r in R.

(iv) If $n > 0$ and $m > 0$, then $f_{n,m}: (\overset{n}{\underset{R}{\otimes}} M) \underset{R}{\otimes} (\overset{m}{\underset{R}{\otimes}} M) \to \overset{n+m}{\underset{R}{\otimes}} M$ has the property $f_{n,m}((x_1 \otimes \cdots \otimes x_n) \otimes (y_1, \ldots, y_m)) = x_1 \otimes \cdots \otimes x_n \otimes y_1 \otimes \cdots \otimes y_m$ for all $x_1, \ldots, x_n$ in M and $y_1, \ldots, y_m$ in M.

(b) Denote by $T_R(M)$ the sum of the family of R-modules $\{\overset{n}{\underset{R}{\otimes}} M\}_{n \in \mathbf{N}}$ which we also denote by $\coprod_{n=0}^{\infty} \overset{n}{\underset{R}{\otimes}} M$.

(i) Letting $g_n: \overset{n}{\underset{R}{\otimes}} M \to T_R(M)$ be the usual inclusion morphisms, show that there is a unique R-morphism $f: T_R(M) \underset{R}{\otimes} T_R(M) \to T_R(M)$ such that for all pairs of nonnegative integers n, m we have $f(g_n(x) \otimes g_m(y)) = g_{n+m} f_{n,m}(x \otimes y)$ for all x in $\overset{n}{\underset{R}{\otimes}} M$ and y in $\overset{m}{\underset{R}{\otimes}} M$.

(ii) Show that the underlying abelian group of $T_R(M)$ together with the multiplication $T_R(M) \times T_R(M) \to T_R(M)$ given by $(x, y) \to f(x \otimes y)$ for all x in $T_R(M)$ and y in $T_R(M)$ is a ring. Show that this ring structure in $T_R(M)$ has the property that the map $R \to T_R(M)$ given by the usual injective morphism $R = \overset{0}{\underset{R}{\otimes}} M \to \coprod_{n=0}^{\infty} \overset{n}{\underset{R}{\otimes}} M$, is a morphism of rings such that the image of R under this ring morphism $R \to T_R(M)$ is contained in the center of $T_R(M)$. Hence, the rings R and $T_R(M)$ together with the morphism of rings $R \to T_R(M)$ just described is an R-algebra which is called the **tensor algebra of M over R**. Usually the ring R is considered a subring of $T_R(M)$ by viewing the injective ring morphism $R \to T_R(M)$ as an identification of R with its image under the ring morphism $R \to T_R(M)$. With this convention in mind the tensor algebra $R \to T_R(M)$ is just denoted by $T_R(M)$.

(21) Let R be a commutative ring and M an R-module. Let $h: R \to \Lambda$ be an arbitrary R-algebra. We want to describe in this exercise the R-algebra morphisms from $T_R(M)$ to Λ.

(a) Recalling that $\overset{1}{\underset{R}{\otimes}} M = R \underset{R}{\otimes} M$ we see that $\overset{1}{\underset{R}{\otimes}} M = M$, using the usual identifica-

tion of $R \underset{R}{\otimes} M$ with M. Hence, viewing the injection morphism $\overset{1}{\otimes} M \to T_R(M)$ as an identification, we see that $T_R(M) \supset M$ as an R-submodule of $T_R(M)$. Show that the subring generated by R and M is all of $T_R(M)$.

(b) Suppose $f: T_R(M) \to \Lambda$ is a morphism of R-algebras, that is, f is a morphism of rings such that $f|R = h$.

(i) Show that $f|M: M \to \Lambda$ is a morphism of R-modules where Λ is considered an R-module by means of the ring morphism $h: R \to \Lambda$.

(ii) Show that if $g: T_R(M) \to \Lambda$ is also a morphism of R-algebras, then $g = f$ if and only if $g|M = f|M$.

(iii) Suppose that $g: M \to \Lambda$ is a morphism of R-modules. Show that there is one and only one morphism of R-algebras $t: T_R(M) \to \Lambda$ such that $t|M = g$. Hence, denoting the set of R-algebra morphisms of $T_R(M)$ to $h: R \to \Lambda$ by $(T_R(M), \Lambda)$, we obtain that the map $(T_R(M), \Lambda) \to \mathrm{Hom}_R(M, \Lambda)$ given by $f \to f|M$ is an isomorphism of sets.

(iv) Suppose the R-algebra $q: R \to \Gamma$ has the property that there is an injective morphism $v: M \to \Gamma$ of R-modules such that for each R-algebra $h: R \to \Lambda$, the map from the set (Γ, Λ) of R-algebra morphisms to $\mathrm{Hom}_R(M, \Lambda)$ given by $f \to fv$ is an isomorphism of sets. Prove that the uniquely determined morphism of R-algebras $u: T_R(M) \to \Gamma$ satisfying $u|R = q$ and $u|M = v$ is an isomorphism of R-algebras.

(22) Let R be a commutative ring and M an R-module. Show that if the R-module M is isomorphic to R, then:

(a) $T_R(M)$ is a commutative R-algebra and

(b) $T_R(M)$ is isomorphic to the R-algebra $R[X]$.

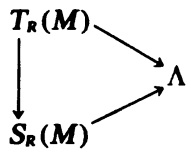
Further, show that if M is isomorphic to a free R-module with a basis consisting of two or more elements, then $T_R(M)$ is not commutative.

(23) Suppose M is a module over the commutative ring R. Let I be the ideal of $T_R(M)$ generated by the elements of $\overset{2}{\otimes} M$ of the form $x \otimes y - y \otimes x$ for all x, y in M.

(a) Show that the composition of ring morphisms $R \to T_R(M) \to T_R(M)/I$ is an injective ring morphism with the property that the image of R in $T_R(M)/I$ is contained in the center of $T_R(M)/I$. Thus, the composition $R \to T_R(M)/I$ is an R-algebra. Usually R is identified with its image under this morphism. The R-algebra $T_R(M)/I$ is called the symmetric algebra of M over R and is denoted by $S_R(M)$.

(b) Show that $S_R(M)$ is a commutative R-algebra.

(c) Show that if $R \to \Lambda$ is a commutative R-algebra and $f: T_R(M) \to \Lambda$ is a morphism of R-algebras, then there is a unique morphism of R-algebras $S_R(M) \to \Lambda$ such that the diagram



commutes, where $T_R(M) \to S_R(M)$ is the canonical surjective morphism of rings.

(d) Show that the composition of R-module morphisms $M \to T_R(M) \to S_R(M)$ is injective. This injective R-module morphism $M \to S_R(M)$ is usually considered an identification of M with its image in $S_R(M)$.

(e) Suppose $R \to \Lambda$ is a commutative R-algebra (that is, Λ is a commutative ring).

(i) Show that if $f: S_R(M) \to \Lambda$ is an R-algebra morphism, then $f|M: M \to \Lambda$ is a morphism of R-modules.

(ii) Letting $(S_R(M), \Lambda)$ denote the set of R-algebra morphisms from $S_R(M)$ to Λ, show that the map $(S_R(M), \Lambda) \to \mathrm{Hom}_R(M, \Lambda)$ given by $f \to f|M$ is an isomorphism of sets.

(iii) Suppose $R \to \Gamma$ is a commutative R-algebra and $u: M \to \Gamma$ is a morphism of R-modules such that for each commutative R-algebra $R \to \Lambda$, the map $(\Gamma, \Lambda) \to \mathrm{Hom}_R(M, \Lambda)$ given by $f \mapsto fu$ for each R-algebra morphism f from Γ to Λ is an isomorphism. Show that the unique R-algebra $v: S_R(M) \to \Gamma$ which has the property that $v|M = u$ is an isomorphism of R-algebras.

(f) Show that $T_R(M)$ is commutative if and only if the canonical R-algebra morphism $T_R(M) \to S_R(M)$ is an isomorphism.

(g) Suppose that M is a free R-module with basis $\{m_i\}_{i \in I}$. Show that the R-algebra $S_R(M)$ is isomorphic to the polynomial ring $R[X_i]_{i \in I}$.

(24) Suppose $R \to \Lambda$ and $R \to \Gamma$ are algebras over the commutative ring R. We now want to make the R-module $\Lambda \underset{R}{\otimes} \Gamma$ an R-algebra.

(a) Show that there is a unique R-module morphism $(\Lambda \underset{R}{\otimes} \Gamma) \times (\Lambda \underset{R}{\otimes} \Gamma) \to \Lambda \underset{R}{\otimes} \Gamma$ such that $(x \otimes y, x' \otimes y') \mapsto xx' \otimes yy'$ for all x, x' in Λ and y, y' in Γ.

(b) Show that this uniquely determined R-module morphism $(\Lambda \underset{R}{\otimes} \Gamma) \times (\Lambda \underset{R}{\otimes} \Gamma) \to \Lambda \underset{R}{\otimes} \Gamma$ has the property that it, together with the law of composition given by the underlying group structure on $\Lambda \underset{R}{\otimes} \Gamma$, make $\Lambda \underset{R}{\otimes} \Gamma$ a ring.

(c) Show that the ring $\Lambda \underset{R}{\otimes} \Gamma$ has the property that the map $R \to \Lambda \underset{R}{\otimes} \Gamma$ given by $r \to r(1 \otimes 1)$ is a ring morphism whose image is in the center of the ring $\Lambda \underset{R}{\otimes} \Gamma$. Hence, $R \to \Lambda \underset{R}{\otimes} \Gamma$ is an R-algebra called the **tensor product** of the R-algebras Λ and Γ.

(d) Show that the maps $f: \Lambda \to \Lambda \underset{R}{\otimes} \Gamma$ and $g: \Gamma \to \Lambda \underset{R}{\otimes} \Gamma$ given respectively by $f(x) = x \otimes 1$ for all x in Λ and $g(y) = 1 \otimes y$ for all y in Γ have the properties:

(i) f and g are R-algebra morphisms.

(ii) The images of f and g commute, that is, $f(x)g(y) = g(y)f(x)$ for all x in Λ and y in Γ.

(iii) The images of f and g together generate the ring $\Lambda \underset{R}{\otimes} \Gamma$.

The R-algebra morphisms $f: \Lambda \to \Lambda \underset{R}{\otimes} \Gamma$ and $g: \Gamma \to \Lambda \underset{R}{\otimes} \Gamma$ are called the **canonical morphisms**.

(e) Let $R \to \Sigma$ be an arbitrary R-algebra. Then associated with each R-algebra morphism $h: \Lambda \underset{R}{\otimes} \Gamma \to \Sigma$ are the R-algebra morphisms $hf: \Lambda \to \Sigma$ and $hg: \Gamma \to \Sigma$. Show:

(i) The images of hf and hg in Σ commute, that is, $hf(x)hg(y) = hg(y)hf(x)$ for all x in Λ and y in Γ.

(ii) If $h' : \Lambda \underset{R}{\otimes} \Gamma \to \Sigma$ is an R-algebra morphism, then $h = h'$ if and only if $hf = h'f$ and $hg = h'g$.

(iii) If $f' : \Lambda \to \Sigma$ and $g' : \Gamma \to \Sigma$ are R-algebra morphisms such that their images in Σ commute, that is, $g'(y)f'(x) = f'(x)g'(y)$ for all x in Λ and y in Γ, then there is a unique R-algebra morphism $h : \Lambda \underset{R}{\otimes} \Gamma \to \Sigma$ such that $f' = hf$ and $g' = hg$.

(f) We now show that the properties of the tensor product of two R-algebras given in (e) essentially describe this algebra. Suppose we are given an R-algebra $R \to \Omega$. Then the R-algebra Ω is isomorphic to $\Lambda \underset{R}{\otimes} \Gamma$ if and only if there are R-algebra morphisms $f : \Lambda \to \Omega$ and $g : \Gamma \to \Omega$ such that:

(i) The images of f and g in Ω commute.

(ii) Given any R-algebra $R \to \Sigma$ and R-algebra morphisms $f' : \Lambda \to \Sigma$ and $g' : \Gamma \to \Sigma$ such that their images in Σ commute, then there is a unique R-algebra morphism $h : \Omega \to \Sigma$ such that $f' = hf$ and $g' = hg$.

(25) Suppose $R \to \Lambda$ and $R \to \Gamma$ are commutative R-algebras.

(a) Show that $\Lambda \underset{R}{\otimes} \Gamma$ is a commutative R-algebra.

(b) Show that in the category of commutative R-algebras, $\Lambda \underset{R}{\otimes} \Gamma$ is the sum of the R-algebras Λ and Γ.

(c) Suppose that $R \to \Gamma$ is the polynomial ring $R[X_i]_{i \in I}$. Show that $\Lambda \underset{R}{\otimes} R[X_i]$ is isomorphic to the polynomial ring $\Lambda[X_i]_{i \in I}$.

(d) Suppose S is a multiplicative subset of R and $R \to \Lambda$ is the R-algebra $R \to R_S$. Further, suppose $f : R \to \Gamma$ has the property that $f(S) \subset \Gamma$ does not contain 0 and is thus a multiplicative subset of Γ. Show that $R_S \underset{R}{\otimes} \Gamma$ is isomorphic to $\Gamma_{f(S)}$.

(26) Let R be a commutative ring. Then $M_n(R)$, the ring of $n \times n$ matrices over R, can be considered an R-algebra by means of the ring morphism $R \to M_n(R)$ given by $r \to rI$ where I is the identity matrix. Show that if $R \to \Lambda$ is an arbitrary R-algebra, then $\Lambda \underset{R}{\otimes} M_n(R)$ is isomorphic to $M_n(\Lambda)$.

(27) Let S be a multiplicative subset of the commutative ring R and let $f : R \to \Lambda$ be an R-algebra.

(a) Consider the map $\Lambda_S \times \Lambda_S \to \Lambda_S$ given by $(x/s, x'/s') \mapsto xx'/ss'$. Show that the underlying abelian group of the R-module Λ_S together with this map as multiplication makes Λ_S a ring which we denote by Λ_S.

(b) Show that the map $R_S \to \Lambda_S$ given by $r/s \mapsto f(r)/s$ is a ring morphism whose image is contained in the center of Λ_S. Hence, $R_S \to \Lambda_S$ is an R_S-algebra.

(c) Show that Λ_S is isomorphic to $R_S \underset{R}{\otimes} \Lambda$.

(d) Show that Λ is a commutative R-algebra if and only if $\Lambda_{\mathfrak{P}}$ is a commutative $R_{\mathfrak{P}}$-algebra for every maximal ideal $\mathfrak{P}$ of R.

(e) Let M be an R-module and U a multiplicative subject of R.

(i) Show that the R_U-algebra $(T_R(M))_U$ is isomorphic to the R_U-algebra $T_{R_U}(M_U)$.

(ii) Show that the R_U-algebra $(S_R(M))_U$ is isomorphic to the R_U-algebra $S_{R_U}(M_U)$.

(28) Let R be a commutative algebra and $R \to \Lambda$ an arbitrary R-algebra. Suppose $R \to R'$ is a commutative R-algebra.

(a) Show that the canonical ring morphism $R' \to R' \underset{R}{\otimes} \Lambda$ given by $r' \mapsto r' \otimes 1$ for all r' in R' makes $R' \underset{R}{\otimes} \Lambda$ an R'-algebra. This is the usual way $R' \underset{R}{\otimes} \Lambda$ is considered an R'-algebra.

(b) Suppose X is a monoid. Show that the R'-algebra $R'[X]$ is isomorphic to $R' \underset{R}{\otimes} R[X]$.

(29) Suppose R is a commutative ring and $R \to \Lambda$ is an arbitrary R-algebra. Show how Examples 5.4 and 5.5 can be generalized from the case Λ is a commutative R-algebra to the case of arbitrary R-algebras.

(30) Let R be an integral domain with field of quotient K.

(a) Show that an R-module M is a torsion R-module if and only if $K \underset{R}{\otimes} M = 0$.

(b) Show that if M and N are R-modules such that $M \underset{R}{\otimes} N = 0$, then either M or N is a torsion module. [*Hint*: Use the characterization of torsion modules given in (a).]

(c) An R-module M is said to be **divisible by an element r in R** if the R-morphism $M \xrightarrow{r} M$ given by $m \to rm$ for all m in M is surjective. The R-module M is said to be divisible if it is divisible by every nonzero element of R.

(i) Show that if $f: M \to M'$ is a surjective morphism of R-modules and M is divisible, then M' is divisible.

(ii) Show that K and hence every factor module of K is a divisible R-module.

(iii) Show that if M is a torsion R-module and N is a divisible R-module, then $M \underset{R}{\otimes} N = 0$.

(31) Let $f: R \to R'$ be a morphism of commutative rings. Suppose M and N are R'-modules and hence also R-modules by means of the ring morphism f.

(a) Show that there is a unique morphism of R-modules $g: M \underset{R}{\otimes} N \to M \underset{R'}{\otimes} N$ which has the property $g(m \otimes n) = m \otimes n$ for all m in M and n in N.

(b) Show that $g: M \underset{R}{\otimes} N \to M \underset{R'}{\otimes} N$ is surjective.

(c) Show that $g: M \underset{R}{\otimes} N \to M \underset{R'}{\otimes} N$ is an isomorphism if $f: R \to R'$ is surjective. In the case $f: R \to R'$ is surjective, then the isomorphism $g: M \underset{R}{\otimes} N \to M \underset{R'}{\otimes} N$ is usually considered an identification.

(32) Let R be a PID.

(a) Show that an R-module M is a divisible R-module if and only if $(R/\mathfrak{P}) \underset{R}{\otimes} M = 0$ for each nonzero prime ideal $\mathfrak{P}$ of R.

(b) Show that an R-module M is divisible if and only if $(M \underset{R}{\otimes} R/\mathfrak{P}) \underset{R/\mathfrak{P}}{\otimes} (R/\mathfrak{P} \underset{R}{\otimes} M) = 0$ for all nonzero prime ideals $\mathfrak{P}$ of R.

(c) Show that if an R-module M has the property that $M \underset{R}{\otimes} M = 0$, then M is divisible.

(d) Show that an R-module M has the property $M \underset{R}{\otimes} M = 0$ if and only if M is a torsion R-module which is also divisible.

(33) Let R be a PID. Suppose r_1 and r_2 are two elements in R. Show that the ring $R/(r_1) \underset{R}{\otimes} R/(r_2)$ is isomorphic to $R/(r)$ where r is a greatest common divisor of r_1 and r_2.

(34) Let S be a multiplicative subset of a commutative ring R and M an R-module.

(a) Show that the R-morphism $M \to M_S$ is an isomorphism if and only if the R-morphism $M \xrightarrow{s} M$ which is given by $m \mapsto sm$ for all m in M is an isomorphism for each s in S.

(b) Let $\mathfrak{P}$ be a maximal ideal of R and M an R-module such that $\mathfrak{P}M = 0$. Show that the R-morphism $M \to M_{\mathfrak{P}}$ is an isomorphism.

(c) Show that the canonical R-morphism $f: M \to M_S$ has the following properties:

(i) For each R_S-module N and R_S-morphism $g: M_S \to N$, the composition $gf: M \to N$ is an R-morphism.

(ii) For each R_S-module N, the map $\mathrm{Hom}_{R_S}(M_S, N) \to \mathrm{Hom}_R(M, N)$ given by $g \mapsto gf$ is an isomorphism of R-modules.

(d) Suppose M is an R-module, M' is an R_S-module, and $h: M \to M'$ an R-morphism such that for each R_S-module N, the map $\mathrm{Hom}_{R_S}(M', N) \to \mathrm{Hom}_R(M, N)$ given by $g \mapsto gh$ is an isomorphism of R-modules. Show that there is a unique R_S-morphism $u: M' \to M_S$ such that $uh = f$ and that uniquely determined R_S-morphism u is an isomorphism.

(35) Let S be a multiplicative subset of a commutative ring R.

(a) Show that there is a unique functor $F: \mathrm{Mod}(R) \to \mathrm{Mod}(R_S)$ such that $F(M) = M_S$ for each R-module M and $F: \mathrm{Hom}_R(M, N) \to \mathrm{Hom}_{R_S}(F(M), F(N))$ is given by $F(f) = f_S$ for each f in $\mathrm{Hom}_R(M, N)$.

(b) The canonical ring morphism $f: R \to R_S$ enables us to view each R_S-module as an R-module. Show that there is a unique functor $G: \mathrm{Mod}(R_S) \to \mathrm{Mod}(R)$ such that $G(X)$, for each R_S-module X, is the R_S-module X viewed as an R-module and where $G: \mathrm{Hom}_{R_S}(X, X') \to \mathrm{Hom}_R(G(X), G(X'))$ is given by $G(f)$ is the R_S-morphism f viewed as an R-morphism.

(c) Show that F is a left adjoint of G.

(36) Our purpose in this exercise is to generalize the notion of the tensor product of modules over commutative rings to modules over arbitrary rings. It is suggested that the reader review the notions of left and right modules over an arbitrary ring R as discussed in the exercises for Chapter 6.

Let R be an arbitrary ring with center $C(R)$. Suppose M is a right R-module and N is a left R-module. M and N are also $C(R)$-modules because $C(R)$ is a subring of R. Hence, we can form the $C(R)$-module $M \underset{C(R)}{\otimes} N$. Let J be the $C(R)$-submodule of $M \underset{C(R)}{\otimes} N$ generated by the elements of the form $mr \otimes n - m \otimes rn$ for all m in M, n in N, and r in R. Then the tensor product of M and N over R is defined to be the $C(R)$-module $(M \underset{C(R)}{\otimes} N)/J$ which is denoted by $M \underset{R}{\otimes} N$. The image in $M \underset{R}{\otimes} N$ of the element $m \otimes n$ in $M \underset{C(R)}{\otimes} N$ is also denoted by $m \otimes n$. The reader should verify the following rules of calculation in $M \underset{R}{\otimes} N$:

(i) $(m_1 + m_2) \otimes n = m_1 \otimes n + m_2 \otimes n$

(ii) $m \otimes (n_1 + n_2) = m \otimes n_1 + m \otimes n_2$

(iii) $mr \otimes n = m \otimes rn$ for all m, m_1, m_2 in M, n, n_1, n_2 in N, and r in R as well as

(iv) $c(m \otimes n) = mc \otimes n = m \otimes cn$ for all c in $C(R)$, m in M, and n in N.

The reader should try to generalize as many of the notions and results developed for the tensor product of modules over a commutative ring to the tensor

product of modules over arbitrary rings. We indicate some of these generalizations now.

(a) Let M be a right R-module and N a left R-module and X an arbitrary $C(R)$-module. A map $f: M \times N \to X$ is said to be a **bilinear map** if f is bilinear when viewed as a map of $C(R)$-modules and also satisfies $f((mr, n)) = f((m, rn))$ for all m in M, n in N, and r in R. Show:

(i) If $f, g: M \times N \to X$ are bilinear maps, then $f + g: M \times N \to X$ defined by $(f+g)((m,n)) = f(m,n) + g(m,n)$ for all m in M, n in N is a bilinear map called **the sum of the bilinear maps** f and g.

(ii) If c is in $C(R)$ and $f: M \times N \to X$ is a bilinear map, then $(cf): M \times N \to X$ defined by $(cf)((m,n)) = c(f((m,n)))$ is a bilinear map.

(iii) The set $B(M, N; X)$ of all bilinear maps is a $C(R)$-module under the addition and multiplication defined in (i) and (ii).

(iv) If $g: X \to Y$ is a morphism of $C(R)$-modules and $f: M \times N \to X$ is a bilinear map, then the composition $gf: M \times N \to Y$ is a bilinear map.

(v) The map $\xi: M \times N \to M \underset{R}{\otimes} N$ given by $f((m,n)) = m \otimes n$ is a bilinear map whose image generates $M \underset{R}{\otimes} N$ as a $C(R)$-module.

(vi) For each $C(R)$-module X the map $\mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, X) \to B(M, N; X)$ given by $f \mapsto f\xi$ for all f in $\mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, X)$ is an isomorphism of $C(R)$-modules.

(vii) Furthermore, the $C(R)$-isomorphisms described in Part (vi) give rise to an isomorphism of the functors

$$\mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, .): \mathrm{Mod}(C(R)) \to \mathrm{Mod}(C(R))$$

and

$$B(M, N; .): \mathrm{Mod}(C(R)) \to \mathrm{Mod}(C(R))$$

and

$$\mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, .)(X) = \mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, X)$$

and

$$B(M, N; .)(X) = B(M, N; X)$$

for all X in $\mathrm{Mod}(C(R))$.

(b) Let X be a $C(R)$-module. Then we consider the $C(R)$-module $\mathrm{Hom}_{C(R)}(N, X)$ a right R-module by defining (fr) for each f in $\mathrm{Hom}_{C(R)}(N, X)$ and r in R to be the map $fr: N \to X$ given by $(fr)(n) = f(rn)$ for all n in N which is easily seen to be in $\mathrm{Hom}_{C(R)}(N, X)$. Show that for each $C(R)$-module X, the map

$$\alpha: \mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, X) \to \mathrm{Hom}_R(M, \mathrm{Hom}_{C(R)}(N, X))$$

given by $[\alpha(f)(m)](n) = f(m \otimes n)$ for all f in $\mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, X)$, m in M, and n in N is an isomorphism of $C(R)$-modules.

(c) Let X be a $C(R)$-module. We consider $\mathrm{Hom}_{C(R)}(M, X)$ a left R-module by defining for each f in $\mathrm{Hom}_{C(R)}(M, X)$ and r in R, the map $(rf): M \to X$ by $(rf)(m) = f(mr)$ for all m in M. Show that for each $C(R)$-module X, the map

$$\beta: \mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, X) \to \mathrm{Hom}_R(N, \mathrm{Hom}_{C(R)}(M, X))$$

given by $[\beta(f)(n)](m) = f(m \otimes n)$ for all f in $\mathrm{Hom}_{C(R)}(M \underset{R}{\otimes} N, X)$, m in M, and n in N is an isomorphism of $C(R)$-modules.

(d) Use these basic facts to develop the analogs for the tensor product of modules over arbitrary rings of results already obtained for tensor products of modules over commutative rings.

(37) Let R be a commutative ring. An element x in R is said to be nilpotent if $x^n = 0$ for some integer n.

(a) Show that the set N of all nilpotent elements in R is an ideal in R.

(b) Show that if x in R is not nilpotent, then there is a prime ideal of R not containing x. [*Hint*: Show that the subset $S = \{x^n\}_{n \in N}$ of R is a multiplicative subset of R and consider the ring R_S.]

(c) Show that N is the intersection of all the prime ideals of R.

Chapter 10 PRINCIPAL IDEAL DOMAINS

We have already shown that if R is a ring, then every R-module is free if and only if R is a division ring. If we relax the condition of freeness to projectivity, we get the result that every R-module is projective if and only if R is semisimple. Now for any ring R and any R-module M, we have seen that there is an epimorphism $f: F \to M$ where F is a free R-module. We therefore obtain an exact sequence $0 \to K \xrightarrow{g} F \xrightarrow{f} M \to 0$ where $K = \operatorname{Ker} f$, and we may ask whether K itself must be free or projective? To tackle this question, we shall restrict ourselves to commutative rings R, for the good reason that the noncommutative case is too complicated for us to consider here.

Our question, then, is: For what commutative rings R is it true that every submodule of a free module is free or that every submodule of a free module is projective? The condition that every submodule of a free module be projective may be replaced by the equivalent condition that every submodule of a projective be projective. For it easily follows that if the latter condition holds, the former does because free modules are projective. On the other hand, if P is any projective module, and P' is a submodule of P, then P' is projective. Because P is projective, it is a summand of some free module F and hence it is isomorphic to a submodule of F. But P', being a submodule of P, is then also isomorphic to a submodule of F and is therefore projective. We therefore want to study the cases:

(a) Every submodule of a free module is free.
(b) Every submodule of a projective module is projective.

In the first case, we are led immediately to principal ideal domains. That is, if R is a commutative ring such that every submodule of a free module is free, then

the ideals of R, being submodules of the free module R, are all free. However, in Chapter 5, we saw that in a commutative ring R, an ideal is free if and only if it is a principal ideal generated by a regular element. Thus, if all the ideals of R are free, then every ideal in R is principal and R is an integral domain. Hence, R is a principal ideal domain. Naturally, we must answer the question: If R is a principal ideal domain, is it true that every submodule of a free R-module is free? In the course of answering this question, we obtain some information concerning the second question. The main emphasis of this chapter is the study and application of PID's. Later on we will return to a more detailed study of rings satisfying condition (b).

1. SUBMODULES OF FREE MODULES

Theorem 1.1
Let R be a ring (not necessarily commutative) all of whose left ideals are projective. If F is a free R-module and M is a submodule of F, then M is a sum of left ideals of R and hence, projective. More generally, every submodule of a projective R-module is projective and every projective R-module is a sum of left ideals of R. In particular, if every left ideal of R is free, then every submodule of a free R-module is free.

PROOF: We let X be a basis for F and, for every subset Y of X we denote by (Y) the submodule of F generated by Y. The idea of our proof is to consider subsets Y of X such that $(Y) \cap M$ is a sum of left ideals of R, where M is our given submodule of F. We then take a maximal such subset Y^* and show that $Y^* = X$. Because $F = (X)$, we have $M = M \cap F = M \cap (X) = M \cap (Y^*)$ so that M is seen to be a sum of left ideals.

Let $\mathcal{G}$ be the set of ordered pairs (Y, S) where Y is a subset of X and S is a set of submodules of $(Y) \cap M$ each of which is isomorphic to a left ideal of R and such that $(Y) \cap M$ is the sum of these submodules. We order the set $\mathcal{G}$ by setting $(Y_1, S_1) \leq (Y_2, S_2)$ if $Y_1 \subset Y_2$ and $S_1 \subset S_2$. If $\{(Y_\alpha, S_\alpha)\}$ is a totally ordered subset of $\mathcal{G}$, consider the pair (Y, S) where $Y = \cup Y_\alpha$ and $S = \cup S_\alpha$. It is clear that $(Y) = \cup (Y_\alpha)$ so that $(Y) \cap M = \cup (Y_\alpha) \cap M = \cup [(Y_\alpha) \cap M]$. Because each $(Y_\alpha) \cap M$ is the sum of the submodules in the set S_α, a simple argument shows that (Y, S) is an element of $\mathcal{G}$. Thus, $\mathcal{G}$ contains a maximal element (Y^*, S^*). Suppose that $Y^* \neq X$. Then choose an element x in X but not in Y^*, and let $Y = Y^* \cup \{x\}$.

If we denote by $M + (Y^*)$ the submodule of F generated by M and (Y^*), we know that this submodule consists precisely of all elements of the form $m + v$ where $m \in M$ and $v \in (Y^*)$. Let $I = \{r \in R \mid rx \in M + (Y)^*)\}$. Then, clearly, I is a left ideal of R. If m is an element of $M \cap (Y)$, then m is in (Y) so that $m = v + rx$ where v is in (Y^*) and r is in R. Thus, $rx = m - v \in M + (Y^*)$, so that $r \in I$. If $m = v' + r'x$ is another way of writing m as an element of (Y), then $v' + r'x = v + rx$ and $v' - v = (r - r')x$. Because $v' - v \in (Y^*)$ and x is not in Y^*, the fact that $(r - r')x$ is in (Y^*) implies that $r - r' = 0$ or $r = r'$. Hence, $v - v' = 0$ and $v = v'$. Thus, if m is in $M \cap (Y)$, there is one and only one way that m may be written $m = v + rx$ with $v \in (Y^*)$ and $r \in R$. Define a map $f: M \cap (Y) \to I$ by setting

$f(m) = r$ where r is the unique element in I such that $m = v + rx$. The reader can verify easily that f is a morphism. This morphism is an epimorphism for if $r \in I$, then $rx \in M + (Y^*)$ so that $rx = m + v$ where $v \in (Y^*)$. Thus, $m = rx - v \in (Y)$ and m is therefore in $M \cap (Y)$. Because $r = f(m)$, f is surjective. The kernel of f is clearly $M \cap (Y^*)$ because, if $m = v + rx$ and $f(m) = r = 0$, then $m = v \in (Y^*)$ and $m \in M \cap (Y^*)$. Moreover, if $m \in M \cap (Y^*)$, $m = v + 0 \cdot x$ where $v \in (Y^*)$, and $f(m) = 0$.

Because $\operatorname{Ker} f = M \cap (Y^*)$, we have the exact sequence

$$0 \longrightarrow M \cap (Y^*) \xrightarrow{i} M \cap (Y) \xrightarrow{f} I \longrightarrow 0$$

where i is the inclusion. By our hypothesis on left ideals of R, I is projective so that the exact sequence is splittable. In particular, there is a monomorphism $j : I \to M \cap (Y)$ such that $fj = \mathrm{id}_I$, and $j(I)$ is a submodule of $M \cap (Y)$ isomorphic to I. Hence, $M \cap (Y)$ is the sum of $M \cap (Y^*)$ and $j(I)$. If we set $S = S^* \cup \{j(I)\}$, we see that (Y, S) is in $\mathcal{S}$ and is properly larger than (Y^*, S^*), contradicting the maximality of (Y^*, S^*). Hence, $Y^* = X$ and M is the sum of left ideals of R.

We know by Chapter 7, Basic Properties 3.1 that a sum of projective modules is projective so that we may conclude that M is a projective module. We saw in the introduction to this chapter that if every submodule of a free module is projective, then every submodule of a projective module is projective. The same argument shows that if every left ideal of R is projective, then every projective module is a sum of left ideals of R. For, since every projective is a summand of a free module, it is a submodule of a free module and hence is a sum of left ideals. If every left ideal of R is free, then every projective module is free because a sum of free modules is free. In particular, we have the following.

Corollary 1.2
Let R be a commutative ring. Then the following three conditions are equivalent.

(a) R is a principal ideal domain.
(b) Every submodule of a free R-module is free.
(c) Every submodule of a finitely generated free R-module is free. In particular, if R is a PID, then every projective R-module is free.

From the point of view of our main program, we might say that Theorem 1.1 finishes everything off because we now know that, for a commutative ring R, R is a PID if and only if every submodule of a free module is free. Also, we know that R has the property that every ideal is projective if and only if every submodule of a projective module is projective. However, it is worth our time to look at commutative rings in general and some examples of PID's in particular to see how the peculiar property of being a PID is reflected in module theory.

The reader has already seen examples of PID's. Another interesting example of a PID, actually a Euclidean ring, is the following.

Example 1.3 Let R be the set of all complex numbers $a + bi$ such that a and b are integers. Under the usual addition and multiplication of complex numbers, R is a ring, acutally a subring of the field of complex numbers. Recall that addition is

defined as

$$(a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + (b_1 + b_2)i$$

and multiplication is

$$(a_1 + b_1 i)(a_2 + b_2 i) = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)i$$

To see that R is a Euclidean ring (clearly it is an integral domain), we must define a map $g : R - \{0\} \to \mathbf{N}$ such that if $x, y \in R$ and $y \neq 0$, then there exist elements q and $r \in R$ such that $x = qy + r$ where either $r = 0$ or $g(r) < g(y)$.

Define the map $g : R - \{0\} \to \mathbf{N}$ by $g(a + bi) = a^2 + b^2$. Because a and b are integers, $a^2 + b^2$ is an integer; in fact, it is always a positive integer. Notice that if $z = a + bi$ is in R, then the element $a - bi$ is also in R, and we denote it by $\bar{z}$. The element $\bar{z}$ is called the **conjugate** of z and it should be observed that $\bar{\bar{z}} = z$. Another fact worth noting is that for all $z \in R - \{0\}$, $g(z) = z\bar{z} = g(\bar{z})$. It is also true that $g(z_1 z_2) = g(z_1)g(z_2)$. This may be verified by direct computation or by the following more indirect method. First notice that $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$. Once we have this, we have $g(z_1 z_2) = (z_1 z_2)(\overline{z_1 z_2}) = z_1 z_2 \bar{z}_1 \bar{z}_2 = z_1 \bar{z}_1 z_2 \bar{z}_2 = g(z_1)g(z_2)$. From this fact it is obvious that $g(ab) \geq g(a)$ if $g(b) \neq 0$.

Let us take elements $x = x_1 + x_2 i$ and $y = y_1 + y_2 i$. If $x = qy + r$, with $r = 0$ or $g(r) < g(y)$, we have $x\bar{y} = qy\bar{y} + r\bar{y}$ with $r\bar{y} = 0$ or $g(r\bar{y}) = g(r)g(\bar{y}) < g(\bar{y})g(\bar{y}) = g(y)g(\bar{y}) = g(y\bar{y})$. Conversely, if we can find a q' and an r' such that $x\bar{y} = q'y\bar{y} + r'$ with $g(r') < g(y\bar{y})$ (or $r' = 0$), then we have $x\bar{y}y = q'y\bar{y}y + r'y$. Thus, $r'y = (x - q'y)y\bar{y}$ and so $r'y/y\bar{y} = x - q'y \in R$ [do not forget that $y\bar{y}$ is just an ordinary integer, namely, $(y_1^2 + y_2^2)$]. Moreover, the reader can check that

$$g\left(\frac{r'y}{y\bar{y}}\right) = \frac{g(r'y)}{g(y\bar{y})}$$

and, because $g(r') < g(y\bar{y})$, we have

$$\frac{g(r'y)}{g(y\bar{y})} < \frac{g(y\bar{y})g(y)}{g(y\bar{y})} = g(y)$$

Setting $r = r'y/y\bar{y}$, we have $x - q'y = r$ or $x = q'y + r$ with $g(r) < g(y)$.

This argument shows us that in order to prove that g is a Euclidean function it is sufficient to prove that if y is a positive integer and x is an arbitrary element of r, then $x = qy + r$ with $g(r) < g(y)$. In this case, $g(y) = y^2$ and we want to find $q = q_1 + q_2 i$ so that $g(x - qy) < y^2$. Because $x - qy = (x_1 - q_1 y) + (x_2 - q_2 y)i$, we want to find integers q_1 and q_2 so that $(x_1 - q_1 y)^2 + (x^2 - q_2 y)^2 < y^2$.

If we could find integers q_1 and q_2 such that $|x_1 - q_1 y| \leq y/2$ and $|x_2 - q_2 y| \leq y/2$, we would be done. But this boils down to proving that if x is an integer and y is a positive integer, then we can find integers q and r such that $x = qy + r$ and $|r| \leq y/2$. We know from our proof of the fact that the ring $\mathbf{Z}$ is a Euclidean ring that we may find q' and r' such that $x = q'y + r'$ with $0 \leq r' < y$. If $r \leq y/2$, we are allright. Suppose $r' > y/2$ and let $r = r' - y$. Then $|r| = y - r' < y - y/2 = y/2$ and $x = q'y + r' = q'y + r + y = (q' + 1)y + r$. Hence, in this case setting $q = q' + 1$, we have $x = qy + r$ with $|r| \leq y/2$. Thus, we can always find q and r with the desired property. Hence, we have shown that there are integers q_1 and q_2 such that $|x_1 - q_1 y| \leq y/2$ and $|x_2 - q_2 y| \leq y/2$. This completes the proof that $g : R - \{0\} \to \mathbf{N}$ is a Euclidean function. Hence, R is a Euclidean ring.

Definition
The ring R defined in Example 1.3 is called the **ring of Gaussian integers** and is usually denoted by $\mathbf{Z}(\sqrt{-1})$. The map $g: R \to \mathbf{Z}$ [with $g(0) = 0$] is called the **norm map** and $g(z)$ is called the **norm of** Z.

Having shown that the Gaussian integers is a Euclidean ring, we know that it is a PID.

2. FREE SUBMODULES OF FREE MODULES

We have already seen in Chapter 6 that if F is a free module of rank n over a PID R, then every submodule F' of F is free of rank at most n. We also know that every ideal in a commutative ring is of rank at most the rank of the ring, namely, one. These results suggest the following: If F is a free module of rank n over an arbitrary commutative ring R and $F'' \subset F$ is a free submodule of F, then rank $F' \leq n$. This section is devoted to demonstrating this fact.

If R is an integral domain, we can verify this conjecture immediately. Let K be the field of quotients of R, and let F' be a free module of rank m contained as a submodule in the free R-module F of rank n. Because F' is contained in F, $K \underset{R}{\otimes} F' \subset K \underset{R}{\otimes} F$ where $K \otimes F'$ is a vector space over K of dimension m and $K \underset{R}{\otimes} F$ is a vector space over K of dimension n. We know that the finite-dimensional vector space $K \underset{R}{\otimes} F$ cannot contain a subspace of dimension greater than itself. Thus, $K \underset{R}{\otimes} F'$ must have dimension m less than or equal to n. Hence, rank $F' \leq$ rank F.

But what happens if R is not an integral domain? Can we reduce the problem to the preceding case? Suppose we have a prime ideal $\mathfrak{P}$ in R and suppose we know that $F' \underset{R}{\otimes} R/\mathfrak{P} \to F \underset{R}{\otimes} R/\mathfrak{P}$ is a monomorphism whenever F is a free R-module of rank n and F' is a free submodule. Then, because $R/\mathfrak{P}$ is an integral domain and because $F \underset{R}{\otimes} R/\mathfrak{P}$ is a free $R/\mathfrak{P}$-module of rank n having $F' \underset{R}{\otimes} R/\mathfrak{P}$ as a free $R/\mathfrak{P}$-submodule, we know that $F' \underset{R}{\otimes} R/\mathfrak{P}$ must have rank at most n. But rank $F' \underset{R}{\otimes} R/\mathfrak{P} =$ rank F' because if $F' = \Sigma_I R$, then $F' \underset{R}{\otimes} R/\mathfrak{P} = \Sigma_I R \underset{R}{\otimes} R/\mathfrak{P} = \Sigma_I R/\mathfrak{P}$. This shows that rank $F' \leq$ rank F.

However, how can we produce a prime ideal of such that $F' \underset{R}{\otimes} R/\mathfrak{P} \to F \underset{R}{\otimes} R/\mathfrak{P}$ is a monomorphism? We can do it if we show that there is a prime ideal $\mathfrak{P}$ and a monomorphism (as R-modules) $f: R/\mathfrak{P} \to R$. For then, because F and F' are free the morphisms $F' \underset{R}{\otimes} R/\mathfrak{P} \to F' \underset{R}{\otimes} R = F'$ and $F \underset{R}{\otimes} R/\mathfrak{P} \to F \underset{R}{\otimes} R = F$ would also be monomorphisms and we would have the commutative diagram

$$\begin{array}{ccc} F' \underset{R}{\otimes} R/\mathfrak{P} & \xrightarrow{i \underset{R}{\otimes} R/\mathfrak{P}} & R \underset{R}{\otimes} R/\mathfrak{P} \\ \Big\downarrow{\scriptstyle F' \underset{R}{\otimes} f} & & \Big\downarrow{\scriptstyle F \underset{R}{\otimes} f} \\ F' \underset{R}{\otimes} R & \xrightarrow{i \underset{R}{\otimes} R} & F \underset{R}{\otimes} R \end{array}$$

where $i: F' \to F$ is the inclusion. Because $i \underset{R}{\otimes} R$ and $F' \underset{R}{\otimes} f$ are monomorphisms, the composition $(i \otimes_R)(F' \underset{R}{\otimes} f) = (F \underset{R}{\otimes} f)(i \underset{R}{\otimes} R/\mathfrak{P})$ is a monomorphism, from which it follows that $i \otimes R/\mathfrak{P}$ is a monomorphism.

Granted the above argument, how do we find a prime ideal $\mathfrak{P}$ and a monomorphism $f: R/\mathfrak{P} \to R$? For that matter, how do we find a monomorphism $h: R/I \to R$ for any proper I at all? Because R is not an integral domain (if it were we would not be concerned about it), there is some nonzero element $x \in R$ which is not regular. The morphism $R \to R$ which sends $r \to rx$, then has a nontrivial kernel I, and clearly the induced morphism $R/I \to R$ is a monomorphism.

Now let us suppose that R is a noetherian ring, and let $\mathcal{T}$ be the set of all proper ideals I such that there is a monomorphism $R/I \to R$. Because R is noetherian, there is a maximal element, say I_0, in this nonempty collection. If we show that I_0 is a prime ideal, we will be done, at least in the case when R is noetherian. If I_0 is not a prime ideal, we can find elements a, b in R with $ab \in I_0$ and neither a nor b in I_0. Because I_0 is in S, there is a monomorphism $f: R/I_0 \to R$. If $x = f(\bar{1})$, where $\bar{1}$ is the identity of R/I_0, then clearly I_0 is the annihilator of x and, because $ab \in I_0$, we have $abx = 0$. However, because $b \notin I_0$, we know that $bx \neq 0$ because $0 \neq f(b) = f(b \cdot \bar{1}) = bf(\bar{1}) = bx$. Now let $J' = I_0 + Ra$; that is, J' is the ideal generated by I_0 and a. Then if $y \in J'$, we have $y = z + ra$ where $z \in I_0$ and $r \in R$. Consequently, $y(bx) = zbx + rabx = bzx + rabx = 0 + 0$ because $zx = 0$ for all $z \in I_0$. Thus, J' is contained in the annihilator, J, of bx and we have a monomorphism $R/J \to R$ where $r\bar{1} = rbx$. Because $bx \neq 0$, J cannot be all of R because $1 \cdot bx \neq 0$. But $J \supset J' \supsetneq I_0$ because $a \notin I_0$ and this contradicts the maximality of I_0. Therefore, I_0 is a prime ideal and we are done.

We summarize what we have just established in the following.

Proposition 2.1
Let R be a noetherian ring. Then there is a monomorphism of R-modules $R/\mathfrak{P} \to R$ with $\mathfrak{P}$ a prime ideal of R.

Finally, we ask what happens if R is not noetherian. Rewording our original problem, what we want to show is that if F is a free R-module of rank n, then any set of $n+1$ elements of F is not linearly independent. Let $\{x_1, \ldots, x_n\}$ be a basis for F, and let $\{y_1, \ldots, y_{n+1}\}$ be a set of elements in F. If R were noetherian, we would already know that $\{y_1, \ldots, y_{n+1}\}$ is not a linearly independent set by what we have just proved. How can we reduce this situation to the noetherian case? Here we resort to a standard type of argument.

Let $y_i = \Sigma\, a_{ij}x_j$ with $a_{ij} \in R$, and let S be the smallest subring of R containing the elements a_{ij}, that is, S is the intersection of all the subrings of R containing the a_{ij}. Now let G be the S-submodule of F generated by $x_1, \ldots, x_n$, that is, G consists of all linear combinations $\Sigma\, s_i x_i$ with $s_i \in S$. Clearly, $\{x_1, \ldots, x_n\}$ is a basis for the S-module G, and $y_1, \ldots, y_{n+1}$ are in G because $y_i = \Sigma\, a_{ij} \in S$. If we knew that S were noetherian, we could then conclude that $\{y_1, \ldots, y_{n+1}\}$ is not linearly independent over S and, hence not linearly independent over R, because R contains S. How then do we show that S is a noetherian ring? In general, how can we show that if R is a commutative ring and $\{a_1, \ldots, a_t\}$ is a finite set of elements of R, then the smallest subring of R containing $a_1, \ldots, a_t$ is noetherian?

Let $\mathbf{Z}$ be the ring of integers and let R be any commutative ring. Then we have seen that we always have a unique ring morphism $f:\mathbf{Z}\to R$ defined by $f(n)=n\cdot 1$. If a is an element of R, it is clear that the smallest subring of R containing a must be the image of the morphism $g:\mathbf{Z}[X]\to R$ given by $g(\Sigma n_i X^i)=\Sigma f(n_i)a^i$. Also, if T is a subring of R and a is an element of R, then the smallest subring of R containing both T and a is the image of the morphism $h:T[X]\to R$ defined by $h(\Sigma b_i X^i)=\Sigma b_i a^i$ where $b_i\in T$. Thus, by induction, one sees that the smallest subring of R containing $a_1,\ldots,a_t$ is the image of the morphism $k:Z[X_1,\ldots,X_t]\to R$ where $k(\Sigma n_{i_1\cdots i_t}X_1^{i_1}\cdots X_t^{i_t})=\Sigma f(n_{i_1\cdots i_t})a_1^{i_1}\cdots a_t^{i_t}$. Because the image of a noetherian ring is always noetherian we will be done if we show that $Z[X_1,\ldots,X_t]$ is noetherian.

The ring $\mathbf{Z}$, being a PID, is certainly noetherian. If we prove the following theorem, we will have solved our problem completely.

Theorem 2.2 (Hilbert's Basis Theorem)
If R is a noetherian ring, then the polynomial ring $R[X_1,\ldots,X_n]$ is also noetherian.

PROOF: Because $R[X_1,\ldots,X_n]=R'[X_n]$ where $R'=R[X_1,\ldots,X_{n-1}]$, it clearly suffices to prove the theorem for the case $n=1$. Thus, we want to show that the polynomial ring in one variable, $R[X]$, is noetherian if R is.

Let I be an ideal in $R[X]$. How can we show that it is finitely generated? Suppose we have a fixed finite set of elements $f_1(X),\ldots,f_s(X)$ in I with $f_j(X)=\Sigma_{i=0}^{n_j}b_{ji}X^i$ such that for any $g=\Sigma_{i=0}^{m}a_iX^i$ in I we have $a_m=\Sigma r_j b_{jn_j}$ with $r_j\in R$. If $m\geq\max(n_j)$, then $g_1(X)=g(X)-\Sigma_{j=1}^{s}r_jX^{m-n_j}f_j(X)$ is an element in I of degree m_1 less than m. If $m_1\geq\max(n_j)$, we may do the same thing to g_1 that we did to g and so, proceeding in this way we see that if $n=\max(n_j)$ and $g(x)$ is an element of I of degree greater than or equal to n, we may find elements $h_1(X),\ldots,h_s(X)$ in $R[X]$ such that $g(X)-\Sigma h_j(X)f_j(X)$ is in I and has degree less than n.

Let I_{n-1} denote the set of elements of I of degree less than n. It is clear that I_{n-1} is an R-module. Moreover, it is a submodule of the R-module consisting of all elements of $R[X]$ of degree less than n, and this latter module is a finitely generated R-module because it is generated as an R-module by $1, X,\ldots,X^{n-1}$. Because R is noetherian, finitely generated R-modules are noetherian, and so I_{n-1} is a finitely generated R-module. Say that I_{n-1} is generated by $k_1(X),\ldots,k_s(X)$. Then, if $g(X)$ is an element of I of degree $m\geq n$, we have as we have already seen, $g(X)-\Sigma h_j(X)f_j(X)\in I_{n-1}$ so that $g(X)-\Sigma h_j(X)f_j(X)=\Sigma r_ik_i(X)$ for some $r_1,\ldots,r_t$ in R from which it follows that $g(X)=\Sigma h_j(X)f_j(X)+\Sigma r_ik_i(X)$. If $g(X)$ is an element of I of degree less than n, then $g(X)$ is in I_{n-1} and $g(X)=\Sigma r_ik_i(X)$. Thus, the set $f_1(X),\ldots,f_s(X)$, $k_1(X),\ldots,k_t(X)$ generates I.

The question then is, how do we find these elements $f_1(X),\ldots,f_s(X)$ having the property needed for the leading coefficients of elements of I? The answer is quite simple. Let J be the set of all leading coefficients of elements of I, together with the element zero. That is, a nonzero element a of R is in J if and only if there is some element $f(X)\in I$ such that $f(X)=\Sigma_{i=0}^{d}a_iX^i$ with $a_d=a$. The set J is clearly an ideal of R. Because J is an ideal of R, J is finitely generated, say by $b_1,\ldots,b_s$. Thus, there are elements $f_1(X),\ldots,f_s(X)$ in I with $f_j(X)=\Sigma_{i=0}^{n_j}b_{ji}X^i$ and such that $b_{jn_j}=b_j$. These elements are what we have been looking for because

if $g(X)$ is an element of I, the leading coefficient of $g(X)$ is in J and is therefore a linear combination of $b_1, \ldots, b_s$ with coefficients in R. Our proof is now complete.

Corollary 2.3
If R is a commutative ring and $a_1, \ldots, a_t$ are elements of R, then the smallest subring of R containing $a_1, \ldots, a_t$ is noetherian.

It is this corollary which puts the finishing touch on our discussion of free submodules of free modules over commutative rings. In fact, we have now proven the following.

Theorem 2.4
If F is a finitely generated free module over the commutative ring R and if F' is a free submodule of F, rank $F' \leq$ rank F.

Corollary 2.5

(a) If a module M is generated by s elements, then any linearly independent subset of M has no more than s elements.
(b) If M is free and has rank at least s, then any s elements which generate M are a basis for M.
(c) If $f: M \to N$ is an epimorphism of the free module M onto the free module N, and if M and N have the same finite rank, then f is an isomorphism.

PROOF: First we prove (a). Because M is generated by s elements, there is an epimorphism $f: F \to M$ where F is a free module having a basis of s elements. If X is a linearly independent subset of M, the submodule (X) generated by X is a free submodule of M with basis X. We have the inclusion $i: (X) \to M$ and, because $F \xrightarrow{f} M$ is an epimorphism and (X) is free and hence projective, there is a morphism $g: (X) \to F$ such that $fg = i$. But $i: (X) \to M$ is a monomorphism, from which it follows that g is a monomorphism. Thus, the image of g is a free submodule of F isomorphic to (X). By Theorem 2.4 this submodule must have rank at most s and so, therefore, must (X).

(b) Of course, it follows from (a) that any basis of M must have s elements, because we are assuming that M has a set of generators consisting of s elements and that every basis of M has at least s elements.

Let $\{m_1, \ldots, m_s\}$ be a set of generators for M and let $\{x_1, \ldots, x_s\}$ be a basis for M. Then we have an epimorphism $f: M \to M$ defined by setting $f(x_i) = m_i$. Because f is an epimorphism and M is free, the exact sequence $0 \to K \to M \xrightarrow{f} M \to 0$ is splittable, where $K = \operatorname{Ker} f$. K is finitely generated and projective, being a summand of a finitely generated free module. If $\mathfrak{P}$ is any maximal ideal of R, we have the exact sequence

$$0 \to R_{\mathfrak{P}} \underset{R}{\otimes} K \to R_{\mathfrak{P}} \underset{R}{\otimes} M \to R_{\mathfrak{P}} \underset{R}{\otimes} M \to 0$$

where now all the $R_{\mathfrak{P}}$-modules are finitely generated and free. Because $\operatorname{rank}(R_{\mathfrak{P}} \underset{R}{\otimes} M) = \operatorname{rank}(R_{\mathfrak{P}} \underset{R}{\otimes} M) + \operatorname{rank}(R_{\mathfrak{P}} \underset{R}{\otimes} K)$ we have $R_{\mathfrak{P}} \underset{R}{\otimes} K = (0)$ for all $\mathfrak{P}$ and so $K = (0)$. This shows that f is an isomorphism from which it follows that $\{m_1, \ldots, m_s\}$ is a basis for M.

(c) Left as an exercise.

The proof we have given of Theorem 2.4 is not the most efficient. A different one involving exterior powers would yield a shorter proof, provided the reader were completely familiar with exterior powers. However, we are not assuming such familiarity and, moreover, we are tempted to say that the techniques we have used in proving Theorem 2.4 are perhaps more important than the result itself. Showing the existence of a monomorphism $R/\mathfrak{P} \to R$ for some prime ideal $\mathfrak{P}$ of R in the noetherian case is a special case of a fundamental step in the noetherian decomposition theory of modules, and the Hilbert Basis Theorem is one of the basic theorems of mathematics. Finally, the reduction from arbitrary commutative rings to noetherian ones is a very helpful procedure when it works. It might also be observed that, in the noetherian case at least, the proof given here can be generalized easily to show that the cardinality of a basis of F' cannot exceed that of a basis of F even when the cardinality of a basis of F is not necessarily finite.

3. FINITELY GENERATED MODULES OVER PID'S

The rest of this chapter is devoted to developing various structure theorems for finitely generated modules over PID's.

Suppose R is a PID. We have already seen in Chapter 9, Proposition 3.3, that every finitely generated torsion-free R-module is a submodule of a free R-module. Because submodules of free modules over PID's are free, we have the following.

Proposition 3.1
If R is a PID, then every finitely generated torsion-free module is free.

This result begins to show us how the module theory over PID's may be quite special. For suppose that R is a PID and that M is a finitely generated R-module. The torsion submodule of M, $t(M)$ is then a finitely generated module and $M/t(M)$ is a finitely generated torsion-free module and therefore free. Consequently, the exact sequence $0 \to t(M) \xrightarrow{i} M \xrightarrow{k} M/t(M) \to 0$ is splittable and we see that M is the sum of a finitely generated torsion module and a finitely generated free module. Therefore, to study finitely generated R-modules, it suffices to study finitely generated torsion modules, because we know what finitely generated free modules are like. It should be observed that if the finitely generated module M is also a sum $\{M' \xrightarrow{j_1} M, F \xrightarrow{j_2} M\}$ where M' is a torsion module and F is free, and $p_1: M \to M'$, $p_2: M \to F$ are the projection morphisms, then there are unique isomorphisms $h_1: M' \to t(M)$ and $h_2: F \to M/t(M)$ such that the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M' & \xrightarrow{j_1} & M & \xrightarrow{p_2} & F & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle h_1} & & \| & & \downarrow{\scriptstyle h_2} & & \\ 0 & \longrightarrow & t(M) & \xrightarrow{i} & M & \xrightarrow{k} & M/t(M) & \longrightarrow & 0 \end{array}$$

is commutative.

If the morphisms h_1 and h_2 exist, they are obviously unique. Also, if one of

them exists, they both exist and h_1 must be a monomorphism while h_2 must be an epimorphism. Finally, it is clear that h_1 will be an isomorphism if and only if h_2 is an isomorphism. Therefore, let us show that $h_1: M' \to t(M)$ exists and is an epimorphism.

We first observe that the composition $kj_1: M' \to M/t(M)$ must be zero. This comes from the general fact that if $f: M_1 \to M_2$ is a morphism from the torsion module M_1 to the torsion-free module M_2, then f must be zero. Because kj_1 is zero, the image of j_1 is contained in $t(M)$ and so we define $h_1: M' \to t(M)$ by $h_1(m') = j_1(m')$. Clearly, h_1 is a monomorphism. To see that h_1 is surjective, suppose $m \in t(M)$. Then $m = j_1p_1(m) + j_2p_2(m)$, because $j_1p_1 + j_2p_2 = \mathrm{id}_M$. Because $p_2(m)$ is in F, $p_2(m)$ must be zero for otherwise $p_2(m)$ would be a nonzero torsion element of F. Hence, $m = j_1p_1(m) = h_1(p_1(m))$ and h_1 is therefore surjective.

Because every finitely generated module over a PID R is uniquely the sum of a finitely generated torsion module and a free module of finite rank, to see what finitely generated R-modules look like, it suffices to study finitely generated torsion R-modules.

If we look at a PID R, the most conspicuous torsion modules around are the cyclic modules, that is, those of the form R/I where I is a nonzero ideal of R. Therefore, we should be reasonably content if we show that every finitely generated torsion module is a sum of such torsion modules. But what about uniqueness of such a sum? For example, consider the ring $\mathbf{Z}$ of integers, and the torsion module $M = \mathbf{Z}/(6) \amalg \mathbf{Z}/(3) \amalg \mathbf{Z}/(8)$. We know that $\mathbf{Z}/(6) = \mathbf{Z}/(2) \amalg \mathbf{Z}/(3)$ and that $\mathbf{Z}/(24) = \mathbf{Z}/(3) \amalg \mathbf{Z}/(8)$. Thus we may write:

(a) $M = \mathbf{Z}/(6) \amalg \mathbf{Z}/(3) \amalg \mathbf{Z}/(8)$ or
(b) $M = \mathbf{Z}/(2) \amalg \mathbf{Z}/(3) \amalg \mathbf{Z}/(24)$ or
(c) $M = \mathbf{Z}/(2) \amalg \mathbf{Z}/(3) \amalg \mathbf{Z}/(8) \amalg \mathbf{Z}/(3)$ or
(d) $M = \mathbf{Z}/(6) \amalg \mathbf{Z}/(24)$.

Thus, we see that a finitely generated torsion module over $\mathbf{Z}$ can be written in many ways as a sum of cyclic modules. Are there conditions one can put on a representation of a torsion module as a sum of cyclic modules which force the representation to be unique, at least up to isomorphism? For example, if we look at (c), we observe that each summand of M is a factor of $\mathbf{Z}$ by a power of a prime, and these cannot be broken down any further into a sum of nonzero modules as we show later. Can M be written as a direct sum of $\mathbf{Z}$ modulo powers of primes in some other way? It turns out that the answer is no, as we shall eventually see. If we look at (d), this seems to be an efficient decomposition of M because it involves the least number of summands. It has the property that 6 divides 24. Is it possible to write M as a sum of $\mathbf{Z}/(n_i)$ such that n_i divides n_{i+1} in any other way than (d)? We shall eventually show that the answer to this also is no. Also, from the prime factorizations of $6 = 3 \times 2$ and $24 = 3 \times 8$, we may read off (c) once we have (d). Also, from (c) we may read off (d) by writing a table:

$$\begin{matrix} 2^3, & 2 \\ 3, & 3 \end{matrix}$$

and then taking $2^3 \times 3 = 24$ and $2 \times 3 = 6$. In short, what we are saying is that if we can write the decomposition of a module as a sum of factors of $\mathbf{Z}$ by powers of primes $\amalg \mathbf{Z}/(p^i)$, then we can write it as a sum $\amalg \mathbf{Z}(\epsilon_j)$ where ϵ_j divides ϵ_{j+1} and

vice versa. The rest of this chapter is devoted to proving the existence and uniqueness of such decompositions for any finitely generated torsion module over any PID.

Let R be a PID and let M be a finitely generated torsion module over R. Then $\text{ann}(M)$ is a nonzero ideal of R. This can be seen by choosing a finite set $\{m_1, \ldots, m_s\}$ of generators of M. If we let $I_j = \{r \in R | rm_j = 0\}$, then $I_j \neq (0)$ because M is a torsion module and hence every element of M is a torsion element. Because R is a PID, we have $I_j = (a_j)$ and, because $a_1 m_1, \ldots, a_s m_i = 0$ for $i = 1, \ldots, s$, we see that $a_1, \ldots, a_s$ is in the annihilator of M and $a_1, \ldots, a_s \neq 0$. Thus, $\text{ann}(M) \neq 0$ and in fact $\text{ann}(M) = I_1 \cap \cdots \cap I_s$. Hence, $\text{ann}(M) = (a)$ where $a = \text{lcm}[a_1, \ldots, a_s]$. In general, if a module M is generated by submodules $M_1, \ldots, M_n$, the reader can readily check that $\text{ann}(M) = \text{ann}(M_1) \cap \cdots \cap \text{ann}(M_n)$. To see that torsion modules which are not finitely generated need not have nonzero annihilators, let us look at the following.

Example 3.2 Let $\mathbf{Z}$ be the ring of integers and $\mathbf{Q}$ the rational numbers. Then $\mathbf{Q}/\mathbf{Z}$ is a torsion module. For if $a/b \in Q$ represents an element x of $\mathbf{Q}/\mathbf{Z}$, we have bx represented by $b(a/b) = a \in \mathbf{Z}$ and so $bx = 0$ with $b \neq 0$. If $\text{ann}(\mathbf{Q}/\mathbf{Z}) \neq (0)$, there is an n in $\mathbf{Z}$ such that $n(\mathbf{Q}/\mathbf{Z}) = 0$. This means that for every $a/b \in \mathbf{Q}$, $n(a/b)$ is in $\mathbf{Z}$. If we choose $m \in Z$ such that n and m are relatively prime, clearly $n(1/m)$ is not in $\mathbf{Z}$. Thus, $\text{ann}(\mathbf{Q}/\mathbf{Z}) = (0)$. Let us return to the case of an arbitrary, not necessarily finitely generated torsion module M. We assume throughout the rest of this section that the annihilator (a) of M is not zero. Using the fact that R is a UFD, we may write $a = p_1^{\nu_1} \cdots p_t^{\nu_t}$ where $p_1, \ldots, p_t$ are distinct primes and $\nu_1, \ldots, \nu_t$ are positive integers. The first thing we notice is the following.

Lemma 3.3
If M is an R-module, m an element of M, and a, b two relatively prime elements of R such that $am = bm = 0$, then $m = 0$.

PROOF: Because a and b are relatively prime and R is a PID, we can find elements $r, s \in R$ such that $ra + sb = 1$. Then $m = 1 \cdot m = (ra + sb)m = ram + sbm = 0$.

As a result of this fact, we see that if q is any prime of R such that $(q) \neq (p_i)$ for $i = 1, \ldots, t$, and if $q^\nu m = 0$ for some positive integer ν and some $m \in M$, then $m = 0$. Another immediate result is that if $p_i^\mu m = 0$ for $\mu > \nu_i$, then $p_i^{\nu_i} m = 0$. For we have $p_i^\mu m = p_i^{\mu - \nu_i}(p_i^{\nu_i} m) = 0$ and $(a/p_i^{\nu_i})(p_i^{\nu_i} m) = 0$. Because $a = \Pi p_i^{\nu_i}$, the elements $p_i^{\mu - \nu_i}$ and $a/p_i^{\nu_i}$ are relatively prime so that, by Lemma 3.3, $p_i^{\nu_i} m$ must be zero.

In Chapter 6 we proved that if M is a torsion module over a PID and if $M_{(q)}$ denotes the submodule of M consisting of those elements $m \in M$ such that $q^\nu m = 0$ for some positive integer ν and prime q, then M is the sum of the submodules $M_{(q)}$ where the (q) run through all the distinct *PPD* of R. As a consequence of our previous observations we have the following.

Proposition 3.4
Let M be a torsion module with $\text{ann}(M) = (a)$, $a \neq 0$. If $a = p_1^{\nu_1} \cdots p_t^{\nu_t}$, then:

(a) $M = M_{(p_1)} \amalg \cdots \amalg M_{(p_t)}$.

(b) $\operatorname{ann}(M_{(p_i)}) = (p_i^{\nu_i})$ for all $i = 1, \ldots, t$.

PROOF: **(a)** follows from Lemma 3.3.

(b) We now show that if we let $(b_i) = \operatorname{ann}(M_{(p_i)})$, then $(b_i) = (p_i^{\nu_i})$. We first show that $p_i^{\nu_i} M_{(p_i)} = 0$. Let m be in $M_{(p_i)}$. Then $p_i^{\nu} m = 0$ for some ν. If $\nu > \nu_i$, we have seen by our previous remarks $p_i^{\nu_i} m = 0$. If $\nu \leq \nu_i$, then certainly $p_i^{\nu_i} m = 0$. Thus, $(b_i) \supset (p_i^{\nu_i})$, that is, $b_i | p_i^{\nu_i}$. Hence, $b_i = u p_i^{\mu}$ where u is a unit and $\mu \leq \nu_i$. Therefore, $(b_i) = (p_i^{\mu})$, and we have $\operatorname{ann}(M_{(p_i)}) = (p_i^{\mu}) \supset (p_i^{\nu_i})$.

To show that $(p_i^{\nu_i}) = \operatorname{ann}(M_{(p_i)})$, it suffices to produce at least one element $m \in M$ such that $p_i^{\nu_i} m = 0$ but $p_i^{\nu_i - 1} m \neq 0$. Let $b = a/p_i$. Then b is not in (a) so there is some element $m' \in M$ such that $bm' \neq 0$. Let $m = (a/p_i^{\nu_i}) m'$. Then $p_i^{\nu_i} m = p_i^{\nu_i}(a/p_i^{\nu_i}) m' = am' = 0$, but $p_i^{\nu_i - 1} m = p_i^{\nu_i - 1}(a/p_i^{\nu_i}) m' = (a/p_i) m' = bm' \neq 0$.

It is clear that our next step must be to study the modules M having the property that $\operatorname{ann}(M) = (p^{\nu})$ for some prime p and fixed positive integer ν. To get a clue as to what we might expect, suppose that $\nu = 1$. Then M is an $R/(p)$-module because $pM = 0$, and because $R/(p)$ is a field, M is a vector space over $R/(p)$ and is therefore a sum of copies of $R/(p)$. We might expect that if $\nu > 1$, the module M is still a sum of modules of the type $R/(p^{\nu_i})$ where $\nu_i \leq \nu$. In general, we cannot expect M to be a sum of modules $R/(p^{\nu})$ with ν fixed as the example $R/(p^{\nu}) \amalg R/(p)$ quickly shows. But that M contains a submodule isomorphic to $R/(p^{\nu})$ is clear.

This comes from the fact that every element of M is annihilated by p^{μ} for some $\mu \leq \nu$. Hence, if every element were killed by p^{μ} for some $\mu < \nu$, we would have $\operatorname{ann}(M) = (p^{\nu'})$ for some $\nu' < \nu$ contradicting the fact that $\operatorname{ann}(M) = (p^{\nu})$. Therefore, there is an element $m \in M$ whose annihilator is (p^{ν}) and hence the morphism $R \to M$ sending 1 to m has kernel (p^{ν}). As usual, this implies that there is a monomorphism $R/(p^{\nu}) \to M$. Notice that the element m whose annihilator is (p^{ν}) is such that $\{m\}$ is a linearly independent subset of M when M is considered an $R/(p^{\nu})$-module. Therefore, viewing M as an $R/(p^{\nu})$-module, M contains a nonempty maximal linearly independent subset $\{m_{\alpha}\}$. The submodule of M generated by this subset is then a free $R/(p^{\nu})$-module F, and we have the exact sequence

$$0 \longrightarrow F \xrightarrow{i} M \xrightarrow{k} M/F \longrightarrow 0$$

where i is the inclusion and k the canonical epimorphism.

Suppose this exact sequence were splittable, then we would have M as a sum of F and a submodule M' isomorphic to M/F. F itself is a direct sum of copies of $R/(p^{\nu})$ because F is a free $R/(p^{\nu})$-module. What about M'? Notice that because M' is a submodule of M, every element of M' is killed by p^{μ} for some $\mu < \nu$. Suppose not, and suppose $m' \in M'$ is such that the annihilator of m' is (p^{ν}). Then $\{m_{\alpha}\} \cup \{m'\}$ is a linearly independent subset of M which contradicts the fact that $\{m_{\alpha}\}$ is a maximal linearly independent subset of M. Consequently, $\operatorname{ann}(M') = (p^{\nu'})$ for some $\nu' < \nu$ and by induction on the exponent ν (the case $\nu = 1$ having been already disposed of), we may conclude that M' is a sum of modules $R/(p^{\nu_i})$ with $\nu_i \leq \nu'$. Therefore, M, being the sum of F and M', would be a sum of modules $R/(p^{\nu_i})$ where $\nu_i \leq \nu$. Hence, we have a structure theorem for M if we show that the exact sequence $0 \to F \to M \to M/F \to 0$ is splittable.

If we could show that M/F is $R/(p')$-projective, then we would know that the above sequence is splittable. However, looking at the example $0 \to \mathbf{Z}/(4) \to \mathbf{Z}/(4) \amalg \mathbf{Z}/(2) \to \mathbf{Z}/(2) \to 0$, we see that the sequence is splittable but $\mathbf{Z}/(2)$ is clearly not a projective $\mathbf{Z}/(4)$-module (why?). Is there something special about F, then, that makes the sequence splittable? The answer is yes, and the reason for this answer is found by studying so-called injective modules.

4. INJECTIVE MODULES

Definition

Let R be an arbitrary, not necessarily commutative ring and let M be an R-module. M is called an **injective module** if, for every monomorphism $f: A \to B$ of R-modules, the morphism $(B, M) \to (A, M)$ is an epimorphism.

Basic Properties 4.1

(a) If M is an injective R-module, then every exact sequence $0 \to M \xrightarrow{f} A \xrightarrow{g} B \to 0$ is splittable.

(b) If M is the product of modules $\{M_\alpha\}$, then M is injective if and only if each M_α is injective.

(c) If M is an injective R-module and $\varphi: R \to S$ is a ring morphism, then the S-module $R(S, M)$ is injective where the operation of S on $R(S, M)$ is given by $(sf)(s') = f(s's)$ for s and s' in S.

PROOF: (a) If M is injective, the morphism $(A, M) \xrightarrow{(f, M)} (M, M)$ is an epimorphism so that there is a morphism $h: A \to M$ such that $(f, M)(h) = \mathrm{id}_M$. Because $(f, M)(h) = hf = \mathrm{id}_M$, h is a splitting for the monomorphism f.

(b) The proof is similar to part (c) of Chapter 7, Basic Properties 3.1.

(c) Let $f: A \to B$ be a monomorphism of left S-modules, and let $g: A \to R(S, M)$ be an S-morphism. Because A and B are S-modules, they are R-modules induced by the morphism φ and f is also a monomorphism of R-modules. Moreover, the map $g': A \to M$ defined by $g'(a) = [g(a)](1)$ is an R-morphism, where 1 is the identity element of S. Because M is injective as an R-module, there is an R-morphism $h': B \to M$ such that $h'f = g'$. Define the map $h: B \to R(S, M)$ by $[h(b)](s) = h'(sb)$. Then h is an S-morphism and $hf = g$. This shows that $R(S, M)$ is an injective S-module.

The reader should recall that the property of projectives corresponding to Basic Properties 4.1(a) here was an if and only if statement. For projectives this followed from the fact that every module is the factor module of a projective (actually free) module. Is it also true that every R-module is a submodule of an injective module? If so, we could make Basic Properties 4.1(a) into an if and only if statement too. We approach this question obliquely.

Proposition 4.2

Let R be a ring and M an R-module. M is injective if and only if for every left ideal I of R and every R-morphism $g: I \to M$, there is a morphism $h: R \to M$ such that $h|I = g$.

PROOF: If M is injective, then this is just a special case of the definition of injectivity, so clearly the condition is necessary.

To prove sufficiency, we may consider a module B and submodule A, with a morphism $g: A \to M$. We then want to produce a morphism $h: B \to M$ such that $h|A = g$. As usual, if we want to extend a morphism all the way from A to B, we consider those submodules of B to which it can be extended, show there is a maximal one, and then show that this maximal one is B. Therefore, let $\mathcal{G}$ be the set of pairs (A', g') where A' is a submodule of B containing A and $g': A' \to M$ is a morphism extending g. We order $\mathcal{G}$ by setting $(A_1', g_1') < (A_2', g_2')$ if $A_1' \subset A_2'$ and $g_2'|A_1' = g_1'$. If $\{(A_\alpha', g_\alpha')\}$ is a totally ordered subset of $\mathcal{G}$, let $A' = \cup A_\alpha'$ and define $g': A' \to M$ by setting $g'(a') = g_\alpha'(a')$ if $a' \in A_\alpha'$. As usual, g' is a map, it is a morphism, and (A', g') is in $\mathcal{G}$. The reader should verify all of this. Now let (A^*, g^*) be a maximal element of $\mathcal{G}$. If $A^* \neq B$, choose an element $b \in B$ with $b \notin A^*$, and let $A' = A^* + (b)$. A' contains A^* properly and we shall construct a morphism $g': A' \to M$ extending g^*. This will contradict the maximality of (A^*, g^*) in $\mathcal{G}$ and so we will have proved that $A^* = B$ and g^* is our sought-for h.

We are tempted to extend g^* to $g': A' \to M$ by setting $g'(a^* + rb) = g^*(a^*) + rg'(b)$ if we could just decide how to define $g'(b)$. Let $I = \{r \in R | rb \in A^*\}$. I is clearly a left ideal of R. Define the map $j: I \to M$ by $j(r) = g^*(rb)$. It is easy to see that j is an R-morphism. Our hypothesis on M tells us that there is a morphism $j': R \to M$ such that $j'|I = j$. Let $j'(1) = m_0$ and now define a map $g': A \to M$ by setting $g'(a^* + rb) = g^*(a^*) - rm_0$ for all a^* in A^* and r in R. We claim that g' is a map. For if $a_1^* + r_1 b = a_2^* + r_2 b$, then $(r_1 - r_2)b = a_2^* - a_1^*$ which is in A^* and so $r_1 - r_2$ is in I. Therefore, $g^*(a_2^* - a_1^*) = g^*((r_1 - r_2)b) = j(r_1 - r_2) = j'((r_1 - r_2)1) = (r_1 - r_2)j'(1) = (r_1 - r_2)m_0$. As a result, $g^*(a_1) + r_1 m_0 = g^*(a_2) + r_2 m_0$ which shows that g' is a map. The reader should check that g' is a morphism, and that it clearly extends g^*. This shows that A^* must be B, and our proof of the proposition is complete.

An immediate consequence is the following.

Proposition 4.3
If R is a ring, the following statements are equivalent:

(a) R is semisimple.
(b) Every left ideal of R is a summand of R.
(c) Every R-module is injective.
(d) Every R-module is projective.

The criterion for injectivity of a module given in Proposition 4.2 permits us to prove a less obvious result for noetherian rings.

Proposition 4.4
If R is a left noetherian ring and M is a sum of the R-modules M_α, then M is injective if and only if each M_α is injective.

PROOF: If M is injective, each summand M_α is injective by Basic Properties 4.1 of injective modules. The converse is proved as follows.

Suppose each M_α is injective, that I is a left ideal of R, and $f: I \to M$ is an

R-morphism. Because R is left noetherian, I is finitely generated, say by elements $a_1, \ldots, a_n$. If $\{i_\alpha : M_\alpha \to M\}$ and $\{p_\alpha : M \to M_\alpha\}$ are the injections and projections of M as the sum of the M_α, we know that for each $m \in M$, $p_\alpha(m) = 0$ for all but a finite number of indices. For each element a_i, there are only finitely many α's such that $p_\alpha f(a_i) \neq 0$. Let $\{\alpha_j\}$ be the set of the α's such that $p_{\alpha_j} f(a_i) \neq 0$ for some i. There are only finitely many such α_j's, say $\{\alpha_1, \ldots, \alpha_t\}$, because there are only finitely many elements a_i. Let M' be the sum $M_{\alpha_1} \amalg \cdots \amalg M_{\alpha_t}$ with $M_{\alpha_i} \xrightarrow{j_{\alpha_i}} M' \xrightarrow{q_{\alpha_i}} M_{\alpha_i}$ the injections and projections. The morphism $j : M' \to M$ is the unique morphism such that $jj_{\alpha_i} = i_{\alpha_i}$ and we claim that $f = jf'$ where $f' : I \to M'$ is the unique morphism such that $q_{\alpha_i} f' = p_{\alpha_i} f$.

To see that $f = jf'$ we must show that $jf'(a_k) = f(a_k)$ for $k = 1, \ldots, n$ since $a_1, \ldots, a_n$ generate I. Using the fact that $\Sigma j_{\alpha_i} q_{\alpha_i} = \mathrm{id}_M$, we have $jf'(a_k) = j(\Sigma j_{\alpha_i} q_{\alpha_i}) f'(a_k) = \Sigma i_{\alpha_i} p_{\alpha_i} f(a_k)$. This last sum is equal to $f(a_k)$ because $p_\alpha(f(a_k)) = 0$ if $\alpha \neq \alpha_1, \ldots, \alpha_t$. Thus, $f(a_k) = jf'(a_k)$ for all $k = 1, \ldots, n$. Hence, $f = jf'$.

M' is injective because it is a sum of a finite number of injective modules and hence also the product of these modules. Therefore, there is a morphism $g' : R \to M'$ such that $g'|I = f'$. Let $g : R \to M$ be the composition fj'. Then $g|I = (jg')|I = j(g'|I) = jf' = f$ and so M is injective.

We can also obtain information from Proposition 4.2 about injective modules over PID's, because the ideals in a PID are so much less complicated than the ideals of a general noetherian ring. We first give some definitions.

Definitions

Let R be an integral domain and M an R-module.

(a) An element $m \in M$ is said to be **divisible by a nonzero element** $a \in R$ if $m = am'$ for some $m' \in M$.
(b) An element $m \in M$ is **divisible** if it is divisible by every nonzero element of R.
(c) M is called a **divisible module** if every element of M is divisible.

Basic Properties 4.5

Let R be an integral domain.

(a) A factor module of a divisible module is divisible.
(b) The set of divisible elements of a module M is a divisible submodule of M.
(c) A torsion-free divisible module is injective.
(d) An injective module is divisible.
(e) A necessary and sufficient condition that a module over a PID be injective is that it be divisible.

PROOF: (a) and (b) are left as exercises.

(c) If I is any ideal of R, M a torsion-free divisible R-module, and $f : I \to R$ a morphism, we want to produce a morphism $g : R \to M$ such that $g|I = f$. For any $a_1, a_2 \in I$ we have $f(a_1 a_2) = a_1 f(a_2) = a_2 f(a_1)$. If f is not the zero morphism (if it is, we just set $g = 0$), there is an element $a \in I$ such that $f(a) \neq 0$. Let $m_0 \in M$ be such that $am_0 = f(a)$ (this may be done because M is divisible), and define $g : R \to M$ by $g(r) = rm_0$. If a' is in I, we have $ag(a') = aa'm_0 = a'am_0 = a'f(a) = af(a')$

so that $a(g(a')-f(a'))=0$. Because M is torsion-free and $a \neq 0$, $g(a')=f(a')$, and therefore $g|I=f$.

(d) This follows from the following observation which is not difficult to verify. Let R be an integral domain. An R-module M is divisible if and only if given any nonzero element a in R and R-morphism $f:(a)\to M$, there is an R-morphism $g:R\to M$ such that $g|(a)=f$. Combining this with Proposition 4.2, we see that every injective R-module M is divisible. This same observation also proves (e).

Example 4.6 Let R be an integral domain, and K its field of quotients. Then K is torsion-free. If $a\in R$ and $b/c\in K$, we have $b/c=a(b/ac)$ provided $a\neq 0$. Thus, K is divisible and therefore injective. The module K/R is divisible because it is a factor module of K. In particular, if R is a PID, then K/R is an injective R-module.

We now show that if R is a PID, then every R-module is a submodule of an injective R-module. Let M be an R-module, let $f:F\to M$ be an epimorphism of a free R-module onto M, and let $L=\operatorname{Ker} f$. Now F is a submodule of an injective R-module. This may be seen by writing $F=\amalg R_i$ where each R_i is isomorphic to R. Then $F=\amalg R_i\subset\amalg K_i$ where each K_i is isomorphic to the field of quotients K of R. It is not difficult to check that $\amalg K_i$ is divisible and torsion-free. Therefore, $\amalg K_i$ is injective. Because L is a submodule of F, L is a submodule of $\amalg K_i$ and $\amalg(K_i/L)$ is divisible, being a factor module of a divisible module. However, R is a PID so that $\amalg(K_i/L)$ is injective. We have $F/L\subset\amalg(K_i/L)$ and M is isomorphic to F/L. Therefore, M is isomorphic to a submodule of an injective R-module. From this it is easy to prove the following.

Theorem 4.7
If R is any, not necessarily commutative, ring and M is any R-module, then M is a submodule of an injective R-module.

PROOF: We always have the ring morphism $\varphi:\mathbf{Z}\to R$ where $\mathbf{Z}$ is the ring of integers and where $\varphi(n)=n\cdot 1$. M is thus a $\mathbf{Z}$-module and because $\mathbf{Z}$ is a PID, M is contained in an injective Z-module, N. By Basic Properties 4.1 of injective modules, we know that $\mathbf{Z}(R,N)$ is an injective R-module. Also, since $M\subset N$, $\mathbf{Z}(R,M)\subset\mathbf{Z}(R,N)$. In addition, we know that $M\approx R(R,M)\subset\mathbf{Z}(R,M)$ because an R-morphism of R into M is certainly a $\mathbf{Z}$-morphism of R into M. Thus, $M\subset\mathbf{Z}(R,M)\subset\mathbf{Z}(R,N)$ where $\mathbf{Z}(R,N)$ is an injective R-module.

From this we immediately have the following.

Corollary 4.8
If R is any ring and M an R-module, then M is injective if and only if every exact sequence $0\to M\to A\to B\to 0$ is splittable.

5. THE FUNDAMENTAL THEOREM FOR PID'S

The reader will recall that injective modules were introduced in connection with our discussion of $R/(p^\nu)$ modules ($\nu>0$) where p is a prime element in the PID R.

We showed that if M is a faithful $R/(p^\nu)$-module, then there is an exact sequence of $R/(p^\nu)$-modules $0 \to F \to M \to M/F \to 0$ where F is a free $R/(p^\nu)$-module and M/F is not a faithful $R/(p^\nu)$-module. The reader will also recall that if this exact sequence of $R/(p^\nu)$-modules is splittable, then we obtain a structure theorem for M. We now show that this sequence is splittable by showing that F is an injective $R/(p^\nu)$-module.

Proposition 5.1
Let R be a PID and b a nonzero element of R. Then $R/(b)$ is an injective $R/(b)$-module.

PROOF: We have the ring morphism $\varphi : R \to R/(b)$. If we can find an injective R-module M such that $R/(b) \approx R(R/(b), M)$, then by Basic Properties 4.1 we know that $R/(b)$ is $R/(b)$-injective. In Example 4.6 we showed that if R is a PID and K its field of quotients, then K/R is an injective R-module. If we show that $R/(b) \approx R(R/(b), K/R)$, we will be done.

Consider the morphism $K/R \xrightarrow{b} K/R$ which sends x to bx. We claim that the kernel of this morphism is the image of the morphism $R \xrightarrow{f} K/R$ which sends r to $k(r)$ where $k : K \to K/R$ is the canonical morphism. If $bx = 0$ and $x = (u/v)$, then $0 = bx = k(bu/v)$ implies that $bu = vc$ for some $c \in R$. Hence, $x = k(u/v) = k(c/b) = f(c)$. Clearly, $bf(r) = bk(r/b) = k(r) = 0$, so our assertion about $\operatorname{Im} f$ is true. Now the reader can check that $\operatorname{Ker} f = (b)$ so that, because $\operatorname{Im} f \approx R/\operatorname{Ker} f \approx R/(b)$, we have the exact sequence

$$0 \to R/(b) \to K/R \xrightarrow{b} K/R \to 0$$

the latter zero being justified because K/R is divisible.

From the exact sequence above we obtain the exact sequence $0 \to R(R/(b), R/(b)) \to R(R/(b), K/R) \xrightarrow{(R/(b),\, b)} R(R/(b), K/R)$, and the morphism $(R/(b), b)$ is easily seen to be zero. Therefore, $R(R/(b), R/(b)) \approx R(R/(b), K/R)$. Finally, we know that $R/(b) = R/(b)(R/(b), R/(b)) = R(R/(b), R/(b))$. Thus, $R/(b)$ is an injective $R/(b)$-module.

With this result we are able to prove the following.

Proposition 5.2
If R is a PID and M is an R-module whose annihilator is (p^ν) for some prime element $p \in R$ and positive integer ν, then M is the sum of modules $R/(p^{\nu_i})$ with $\nu_i \leq \nu$. If M is finitely generated, M is the sum of a finite number of such modules.

PROOF: First we observe that if F is a free $R/(p^\nu)$-module, then F is an injective $R/(p^\nu)$-module. To see this recall that $R/(p^\nu)$ is a noetherian ring. Because $F = \amalg R/(p^\nu)$ and each $R/(p^\nu)$ is $R/(p^\nu)$-injective, it follows from Proposition 4.4 that F is an injective $R/(p^\nu)$-module.

Suppose now that M is an R-module whose annihilator is (p^ν), that is, M is a faithful $R/(p^\nu)$-module. We have already seen that under these circumstances there is an exact sequence of $R/(p^\nu)$-modules $0 \to F \to M \to M/F \to 0$ with F a free $R/(p^\nu)$-module such that $\operatorname{ann}(M/F) = (p^\mu)$ with $\mu < \nu$. Because F is an injective $R/(p^\nu)$-module, we have that $M = F \amalg M/F$. By induction on ν, we know that M/F

is a sum of modules $R/(p^{\nu_i})$ where $\nu_i \leq \mu$. Because $F = \coprod R/(p^{\nu})$, we have M is a sum of modules $R/(p^{\nu_i})$ with $\nu_i \leq \nu$.

We leave the proof of the rest of the proposition to the reader.

Putting Propositions 5.2 and 3.4 together, we get the following.

Theorem 5.3

Let R be a PID and M a torsion module with nontrivial annihilator (a). If $a = p_1^{\nu_1}, \ldots, p_t^{\nu_t}$ is a prime factorization of a, then M is the sum of modules $R/(p_i^{\nu_{ij}})$ where $\nu_{ij} \leq \nu_i$ for $i = 1, \ldots, t$. If M is a finitely generated torsion module, M is the sum of a finite number of modules $R/(p_i^{\nu_{ij}})$.

Can we go further? That is, can we decompose the modules $R/(p^{\nu})$ even further as sums? Suppose $R/(p^{\nu}) = M_1 \coprod M_2$. Then $(p^{\nu}) = \operatorname{ann}(M_1 \coprod M_2) = \operatorname{ann}(M_1) \cap \operatorname{ann}(M_2)$ so we have $\operatorname{ann}(M_1) = (p^{\nu_1})$ and $\operatorname{ann}(M_2) = (p^{\nu_2})$ with either ν_1 or ν_2 equal to ν. It can be shown that $R/(p) \underset{R}{\otimes} R/(p^{\nu}) \approx R/(p)$, and we know that $R/(p) \underset{R}{\otimes} (M_1 \coprod M_2) \approx (R/(p) \underset{R}{\otimes} M_1) \coprod (R/(p) \underset{R}{\otimes} M_2)$. Because $R/(p)$ is a one-dimensional vector space over $R/(p)$, we must have either $R/(p) \underset{R}{\otimes} M_1 = 0$ or $R/(p) \underset{R}{\otimes} M_2 = 0$. Say $R/(p) \underset{R}{\otimes} M_2 = 0$. Then $M_2 = (p)M_2$ because $R/(p) \underset{R}{\otimes} M_2 \approx M_2/(p)M_2$. But if $M_2 = (p)M_2$, then $M_2 = (p^{\mu})M_2$ for every μ and, because $(p^{\nu_2})M_2 = 0$, we have $M_2 = 0$. Consequently, the modules $R/(p^{\nu})$ cannot be reduced further.

Our next question is about the uniqueness of this type of decomposition of torsion modules M with nontrivial annihilator. Suppose $M = \coprod_{(p)} \coprod_{i_{(p)} \in I_{(p)}} R/(p^{\nu_{i(p)}})$ where (p) runs through $PPD(R)$, and let $N_{(p)} = \coprod_{i_{(p)} \in I_{(p)}} R/(p^{\nu_{i(p)}})$. Then $M_{(p)} = N_{(p)}$, a fact we leave to the reader to verify.

Therefore, if $M = \coprod_{(p)} \coprod_{i_{(p)} \in I_{(p)}} R/(p^{\nu_{i(p)}})$ and $M = \coprod_{(p)} \coprod_{j_{(p)} \in J_{(p)}} R/(p^{\mu_{j(p)}})$, we know that for each (p) in $PPD(R)$ we have $\coprod_{i_{(p)} \in I_{(p)}} R/(p^{\nu_{i(p)}}) = \coprod_{j_{(p)} \in J_{(p)}} R/(p^{\mu_{j(p)}})$. Thus, to show that these two decompositions of M are essentially the same, it suffices to show that for each (p) any two decompositions of $M_{(p)}$ are essentially the same. This amounts to proving that for each (p), there is a bijective map $f: I_{(p)} \to J_{(p)}$ such that $\nu_{k_{(p)}} = \mu_{f(k_{(p)})}$. We prove this only in the case $M_{(p)}$ is finitely generated.

Proposition 5.4

Let R be a PID and M a finitely generated R-module whose annihilator is (p^{ν}) for some prime element p in R. Then any two decompositions of M as a sum of modules $R/(p^{\mu})$ are essentially the same in the sense described above.

PROOF: Suppose $M = \coprod_{i=1}^{n} R/(p^{\nu_i}) = \coprod_{j=1}^{m} R/(p^{\mu_j})$. Assume the exponents $\nu_1, \ldots, \nu_n$ and $\mu_1, \ldots, \mu_m$ have been so labeled that $\nu_1 \geq \nu_2 \cdots \geq \nu_n$ and $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m$. We want to show that $m = n$ and that $\nu_i = \mu_i$ for $i = 1, \ldots, m$. This is a special case of the following much more general theorem.

Theorem 5.5
Let R be a noetherian commutative ring and $I_1 \subset I_2 \subset \cdots \subset I_n$, $J_1 \subset \cdots \subset J_m$ two sequences of proper ideals of R such that

$$\frac{R}{I_1} \amalg \cdots \amalg \frac{R}{I_n} \approx \frac{R}{J_1} \amalg \cdots \amalg \frac{R}{J_m}$$

Then $n = m$ and $I_k = J_k$ for $k = 1, \ldots, n$.

PROOF: To show that $m = n$, we take a maximal ideal M containing I_n and, tensoring, one obtains

$$\frac{R}{M} \underset{R}{\otimes} \frac{R}{I_1} \amalg \cdots \amalg \frac{R}{M} \underset{R}{\otimes} \frac{R}{I_n} \approx \frac{R}{M} \underset{R}{\otimes} \frac{R}{J_1} \amalg \cdots \amalg \frac{R}{M} \underset{R}{\otimes} \frac{R}{J_m}$$

These are vector spaces over R/M which, on the one hand, are n-dimensional because $R/M \underset{R}{\otimes} R/I_k = R/M$, and at most m-dimensional because $R/M \underset{R}{\otimes} R/J_k = (0)$ or R/M. Hence, $n \leq m$. Similarly, $m \leq n$ so $m = n$.

To show that $I_n = J_n$, we tensor everything with R/I_n. Then $R/I_n \underset{R}{\otimes} R/I_k = R/I_n$ because $I_k \subset I_n$. Therefore, we see that $R/I_n \amalg \cdots \amalg R/I_n = R/(I_n, J_1) \amalg \cdots \amalg R/(I_n, J_n)$ where (I_n, J_k) denotes the ideal generated by I_n and J_k. Because the sum of n copies of R/I_n is a free R/I_n-module, and because $R/(I_n, J_k)$ is a direct summand of this R/I_n-module for each k, we know that $R/(I_n, J_k)$ is a projective R/I_n-module. Thus, the canonical epimorphism $R/I_n \to R/(I_n, J_k)$ is a splittable R/I_n-epimorphism, and we have $R/I_n \approx R/(I_n, J_k) \amalg Z_k$ where $Z_k \approx (I_n, J_k)/I_n$ and Z_k is a finitely generated R/I_n-module. Because $R/I_n \amalg \cdots \amalg R/I_n \approx R/(I_n, J_1) \amalg \cdots \amalg R/(I_n, J_n)$ and because $R/I_n \approx R/(I_n, J_k) \amalg Z_k$, we get $(\amalg R/(I_n, J_k)) \amalg (\amalg Z_k) \approx \amalg R/(I_n, J_k)$, with $\amalg Z_k$ finitely generated. Tensoring both sides by R/M where M is any maximal ideal of R, we get $(R/M \underset{R}{\otimes} \amalg R/(I_n, J_k)) \amalg (R/M \underset{R}{\otimes} \amalg Z_k) \approx R/M \underset{R}{\otimes} (\amalg R/(I_n, J_k))$. Knowing that both sides are finite-dimensional vector spaces over R/M, we conclude that $R/M \underset{R}{\otimes} (\amalg Z_k) = 0$ for every maximal ideal M of R. Therefore, by Chapter 9, Example 5.15, we know that $\amalg Z_k = (0)$. Hence, $Z_k = (0)$ for $k = 1, \ldots, n$, and, in particular, $Z_n = (0)$. Because Z_n was the kernel of $R/I_n \to R/(I_n, J_n)$, we must have $J_n \subset I_n$, if $Z_n = (0)$. A similar argument shows that $I_n \subset J_n$, so we have $I_n = J_n$.

Now suppose that $I_k = J_k$ for $k = n, n-1, \ldots, t+1$, and let us prove that $I_t = J_t$. Tensoring $R/I_1 \amalg \cdots \amalg R/I_n$ by R/I_t, we get $R/I_t \amalg \cdots \amalg R/I_t \amalg R/I_{t+1} \amalg \cdots \amalg R/I_n$ isomorphic to $R/(I_t, J_1) \amalg \cdots \amalg R/(I_t, J_t) \amalg R/I_{t+1} \amalg \cdots \amalg R/I_n$, because $I_k = J_k$ for $k = t+1, \ldots, n$. Letting $F = R/I_t \amalg \cdots \amalg R/I_t$ (t times), $G = R/I_{t+1} \amalg \cdots \amalg R/I_n$, and $H = R/(I_t, J_1) \amalg \cdots \amalg R/(I_t, J_t)$, we have

$$F \amalg G \approx H \amalg G$$

where F, G, and H are finitely generated R/I_t-modules and F is a free R/I_t-module. We know by Chapter 9, Theorem 6.5 that this implies H is a projective R/I_t-module.

Because H is a projective R/I_t-module, $R/(I_t, J_k)$ is also a projective R/I_t-module for $k = 1, \ldots, t$ because it is a summand of H. Thus, the canonical epimorphism of R/I_t-modules $R/I_t \to R/(I_t, J_k)$ is splittable for each k, and we have $R/I_t \approx R/(I_t, J_k) \amalg Z_k'$ where each Z_k' is a finitely generated R/I_t-module. Thus, we have $H \amalg (\coprod_{k=1}^{t} Z_k') \amalg G \approx H \amalg G$. Using the same argument as before, we see that $Z_k' = 0$ for $k = 1, \ldots, t$. In particular, $Z_t' = 0$ and thus we have $J_t \subset I_t$. Similarly, $I_t \subset J_t$ and so $I_t = J_t$. This completes the proof of the theorem.

The foregoing arguments give us the following.

Theorem 5.6
If R is a PID and M is a finitely generated torsion module, then M is the sum of a finite number of modules $R/(p_i^{r_{ij}})$ where the p_i are prime elements of R and the r_{ij} are positive integers. Any two such decompositions are identical up to order.

Now suppose that M is a finitely generated torsion module and that we have its decomposition $M = \coprod_{i=1}^{t} (\coprod_{j=1}^{n_i} R/p_i^{r_{ij}})$ where $(p_1), \ldots, (p_t)$ are distinct elements of $PPD(R)$. Let us assume that for each $i = 1, \ldots, t$, we have $r_{i1} \geq r_{i2} \geq \cdots \geq r_{in_i}$. By choosing n to be the largest of $n_1, \ldots, n_t$, and by setting $r_{ij} = 0$ for $n_i < j \leq n$, we obtain $r_{i1} \geq r_{i2} \geq \cdots \geq r_{in}$ for $i = 1, \ldots, t$.

Define the elements $\epsilon_j \in R$ for $j = 1, \ldots, n$ by

$$\epsilon_j = p_1^{r_{1j}} p_2^{r_{2j}} \cdots p_t^{r_{tj}}$$

Then for no j is ϵ_j a unit and we have, for each $j = 2, \ldots, n$, that ϵ_j divides ϵ_{j-1} or, equivalently, $(\epsilon_{j-1}) \subset (\epsilon_j)$. Moreover, by the Chinese Remainder Theorem (Chapter 5, Proposition 6.6) because $R/(\epsilon_j) \approx R/(p_1^{r_{1j}}) \amalg \cdots \amalg R/(p_t^{r_{tj}})$, we see that $M \approx R/(\epsilon_1) \amalg \cdots \amalg R/(\epsilon_n)$. Thus, every finitely generated torsion module M over a PID is a sum $R/(\epsilon_1) \amalg \cdots \amalg R/(\epsilon_n)$ where $(\epsilon_1) \subset (\epsilon_2) \subset \cdots \subset (\epsilon_n)$ and $(\epsilon_n) \neq R$. By Theorem 5.5 we know that type of decomposition of M is unique.

We have seen that a finitely generated module is the sum of its torsion submodule and a free module, and a finitely generated free module is the finite sum of copies of $R = R/(0)$. Thus, because (0) is contained in every ideal, we see that an arbitrary finitely generated R-module M is the sum of modules $R/(\epsilon_1) \amalg \cdots \amalg R/(\epsilon_n)$ where $(\epsilon_1) \subset \cdots \subset (\epsilon_n) \neq R$ and we now allow (ϵ_i) to be (0). This sum is unique for the number of ϵ_i''s equal to zero determines the free part of M while the rest give the decomposition of $t(M)$.

Notice, too, that (0) is a prime ideal of R. Thus, $R = R/(0)$ is R modulo a power of a prime ideal. We combine all of these observations in the following.

Theorem 5.7
Let R be a PID and M a finitely generated R-module. Then:

(a) M is the finite sum $R/(\epsilon_1) \amalg \cdots \amalg R/(\epsilon_n)$ where $(\epsilon_1) \subset \cdots \subset (\epsilon_n) \neq R$ and M may be written as such a direct sum in only one way.

(b) M is the finite sum $\coprod_i \coprod_j R/(p_i^{a_{ij}})$ where each (p_i) is a prime ideal of R (possibly zero). M may be written as such a sum in only one way.

Theorem 5.7 is called the fundamental theorem for finitely generated modules over PID's.

Definition

(a) The ideals $(\epsilon_1), \ldots, (\epsilon_n)$ of R uniquely associated with the finitely generated module M as in Theorem 5.7 are called the **invariant factors of** M.
(b) The nonzero prime ideals (p_i) of R uniquely associated with the finitely generated module M as in Theorem 5.7 are called the **elementary divisions of** M.

The next chapter is devoted to various applications of the fundamental theorem.

EXERCISES

(1) Throughout this exercise R is a noetherian commutative ring. Let M be an R-module. We say that a prime ideal $\mathfrak{P}$ of R is an **associated prime ideal of** M if there is an element x in M such that $\mathfrak{P}$ is the annihilator of x; that is, P consists of all r in R such that $rx = 0$. We denote by $\mathrm{Ass}_R(M)$, or more simply $\mathrm{Ass}(M)$, the set of prime ideals associated with M.
(a) Show that a prime ideal $\mathfrak{P}$ of R is in $\mathrm{Ass}(M)$ for an R-module M if and only if there is a monomorphism $R/\mathfrak{P} \to M$.
(b) For each prime ideal $\mathfrak{P}$ of R show that $\mathrm{Ass}_R(R/\mathfrak{P}) = \{\mathfrak{P}\}$.
(c) Let M be an R-module. For each element x in M, let $\mathrm{ann}(x)$ denote the annihilator of x. If $\mathcal{S}$ is the set of all ideals of the form $\mathrm{ann}(x)$ for all nonzero x in M, then the maximal elements of the set $\mathcal{S}$ are in $\mathrm{Ass}(M)$. [*Hint*: Generalize the procedure used in showing that there is always an injective morphism $R/\mathfrak{P} \to R$ of R-modules for some prime ideal $\mathfrak{P}$ if R is not the zero ring.]
(d) An R-module $M \neq (0)$ if and only if $\mathrm{Ass}(M) \neq \phi$.
(e) Let $0 \to M' \to M \to M'' \to 0$ be an exact sequence of R-modules. Show that $\mathrm{Ass}(M') \subset \mathrm{Ass}(M) \subset \mathrm{Ass}(M') \cup \mathrm{Ass}(M'')$.
(f) Let M be an R-module. We say that an element r in R is a zero divisor for M if there is a nonzero element x in M such that $rx = 0$. Show that an element r in R is a zero divisor for M if and only if r is in $\mathfrak{P}$ for some $\mathfrak{P}$ in $\mathrm{Ass}(M)$. Hence, $\bigcup_{\mathfrak{P} \in \mathrm{Ass}(M)} \mathfrak{P}$ is the set of zero divisors for M. [*Hint*: To show that if r in R is a zero divisor in M, then r is in $\mathfrak{P}$ for some $\mathfrak{P}$ in $\mathrm{Ass}(M)$, consider the submodule Rx of M where x is a nonzero element of M such that $rx = 0$.]
(g) Let M be a finitely generated R-module. Show that there is a finite chain $0 = M_0 \subset M_1 \subset \cdots \subset M_n = M$ of submodules of M such that for $0 \leq i < n$ we have $M_{i+1}/M_i \approx R/\mathfrak{P}_i$ for some prime ideal $\mathfrak{P}_i$ of R. [*Hint*: Let $\mathcal{S}$ be the set of all submodules of M for which the statement above is valid. Show that if $M \neq (0)$, then $\mathcal{S} \neq \phi$ and that M is a maximal element of $\mathcal{S}$.
(h) Show that if M is a finitely generated R-module, then $\mathrm{Ass}(M)$ is a finite set. [*Hint*: This follows from (g) using (b) and (e).]
(2) Suppose R is a commutative noetherian ring and $\{M_i\}_{i \in I}$ is a family of R-modules. Prove that $\mathrm{Ass}(\coprod_{i \in I} M_i) = \bigcup_{i \in I} \mathrm{Ass}(M_i)$. Use this result to show that

there is an R-module M which is not finitely generated such that $\mathrm{Ass}_R(M)$ is finite.

(3) Let $\mathbf{Z}$ be the ring of integers and $\mathbf{Q}$ the field of rational numbers.

(a) Show that the $\mathrm{Ass}_{\mathbf{Z}}(\mathbf{Q}/\mathbf{Z})$ consists precisely of all the nonzero prime ideals of $\mathbf{Z}$.

(b) Let M be a finite abelian group of order n. How are the set $\mathrm{Ass}(M)$ and the prime divisors of n related?

(4) Let R be a commutative noetherian domain. Show that an R-module M is a torsion R-module if and only if the ideal (0) is not in $\mathrm{Ass}(M)$.

(5) (a) Let R be an arbitrary commutative ring. Show that if an ideal I of R is contained in a finite union $\bigcup_{i=1}^{n} \mathfrak{P}_i$ of prime ideals $\mathfrak{P}_i$ of R, then $I \subset \mathfrak{P}_i$ for some $i = 1, \ldots, n$. [*Hint*: Proceed by induction on n to show that if $I \not\subset \mathfrak{P}_i$ for each $i = 1, \ldots, n$, then $I \not\subset \bigcup \mathfrak{P}_i$.]

(b) Use (a) to establish the following result. Let M be a finitely generated module over a commutative noetherian ring R and I an ideal of R. Then the following statements are equivalent:

(i) There is a nonzero element x in M such that $Ix = 0$.

(ii) For each r in I, there is a nonzero element x in M such that $rx = 0$.

(iii) There is a prime ideal $\mathfrak{P}$ in $\mathrm{Ass}(M)$ such that $I \subset \mathfrak{P}$.

(c) Let $f: R \to R'$ be a ring morphism of commutative, noetherian rings. Suppose M is an R'-module which we consider an R-module by means of the ring morphism $f: R \to R'$. Show:

(i) If $\mathfrak{P}'$ is a prime ideal of R' in $\mathrm{Ass}_{R'}(M)$, then $f^{-1}(P')$ is in $\mathrm{Ass}_R(M)$.

(ii) If M is a finitely generated R'-module and $\mathfrak{P}$ is a prime ideal of R in $\mathrm{Ass}_R(M)$, then there is a prime ideal $\mathfrak{P}'$ of R' in $\mathrm{Ass}_{R'}(M)$ such that $\mathfrak{P}' \supset f(P)$.

(iii) If R is a PID and M is a finitely generated R'-module, then $\mathrm{Ass}_R(M)$ is a finite set.

(d) Let M be a finitely generated R-module with R a noetherian ring. Then the underlying abelian group M of M has the property that $\mathrm{Ass}_{\mathbf{Z}}(M)$ is finite.

(e) Let R be a noetherian ring and M an R-module whose underlying abelian group is isomorphic to $(\mathbf{Q}/\mathbf{Z})$ where $\mathbf{Q}$ is the field of rational numbers. Show that M is not a finitely generated R-module.

(6) Let R be an arbitrary ring. Show that for a monomorphism $f: A \to B$ of R-modules, the following statements are equivalent:

(a) If b is a nonzero element of B, then there is an element r in R such that rb is a nonzero element of $f(A)$.

(b) If B' is a nonzero submodule of B, then $B' \cap f(A) \neq 0$.

(c) An R' morphism $g: B \to C$ is a monomorphism if the composition $gf: A \to C$ is a monomorphism. A monomorphism $f: A \to B$ is said to be an **essential monomorphism** if it satisfies any of the above equivalent conditions. If A is a submodule of a module B, then B is said to be an **essential extension** of A, if the $\mathrm{inc}: A \to B$ is an essential monomorphism.

(7) Suppose R is an arbitrary ring.

(a) Show that for each R-module A, the morphism id_A is an essential monomorphism.

(b) Show that the composition of two essential monomorphisms is an essential monomorphism.

(c) Suppose A is a submodule of a module B. Show that if $\{E_j\}_{j\in J}$ is a totally ordered family of submodules of B which are essential extensions of A, then the submodule $E = \bigcup_{j\in J} E_j$ is an essential extension of A.

(d) Suppose A is a submodule of B. Show that there is a maximal essential extension E of A contained in B, that is, $E \supset A$ is an essential extension of A and if $E' \supset E$ is an essential extension of A, then $E' = E$. Show also that E is the only essential extension of E in B.

(e) Suppose A is a submodule of the injective R-module B. Show that A is the only essential extension of A in B if and only if A has no essential extensions except itself, that is, if A is a submodule of any R-module which is an essential extension of A, then $A = A'$.

(8) Let R be an arbitrary ring. Suppose A is a submodule of the R-module B and $f: A \to C$ a morphism of R-modules. Then the push-out $B \overset{A}{\times} C$ of the morphisms

$$\begin{array}{ccc} A & \xrightarrow{\text{inc}} & B \\ \downarrow{\scriptstyle f} & & \\ C & & \end{array}$$

gives rise to the following commutative exact diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{\text{inc}} & B & \longrightarrow & B/A & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \downarrow & & \| & & \\ 0 & \longrightarrow & C & \longrightarrow & B \overset{A}{\times} C & \longrightarrow & B/A & \longrightarrow & 0 \end{array}$$

(a) Show that the monomorphism $C \to B \overset{A}{\times} C$ is essential if and only if a submodule A' of B is the submodule A whenever A' contains A and there is an R-morphism $f': A' \to C$ with the property $f'|A = f$.

(b) Show that an R-module A is injective if and only if A has no essential extensions except itself.

(c) An R-module E containing the R-module A is said to be an **injective envelope** of A if E is an injective R-module which is also an essential extension of A. Show that every R-module has an injective envelope.

(d) Show that if $E_1 \supset A$ and $E_2 \supset A$ are injective envelopes of A, then there is an isomorphism $f: E_1 \to E_2$ such that $f(a) = a$ for all a in A.

(9) Show that a ring R is left noetherian if and only if every sum of injective R-modules is injective. The proof that if R is left noetherian, then a sum of injective R-modules is injective is given in the text. We now outline a proof that if every sum of injective R-modules is injective, then R is left noetherian. Let $I_0 \subset I_1 \subset \cdots \subset I_n \subset \cdots$ be an ascending chain of left ideals in R and let I be the left ideal $\bigcup_{n=1}^{\infty} I_n$. For each n let E_n be an injective R-module containing R/I_n and $f_n: I \to E_n$ the composition $I \xrightarrow{\text{inc}} R \to R/I_n \xrightarrow{\text{inc}} E_n$.

(a) Define $g: I \to \prod_{n=0}^{\infty} E_n$ by $g(r) = (f_n(r))_{n \in N}$ for all r in I and show that $\operatorname{Im} g$ is contained in the submodule $\coprod_{n=0}^{\infty} E_n$ of $\prod_{n=0}^{\infty} E_n$. Let $f: I \to \coprod_{n=0}^{\infty} E_n$ be defined by $f(r) = g(r)$ for all r in R.

(b) By hypothesis $\coprod_{n=0}^{\infty} E_n$ is injective; hence, there is a morphism $h: R \to \coprod_{n=0}^{\infty} E_n$ such that $h|I = f$. Show that there is an n such that $h(R) \subset E_0 \amalg E_1 \amalg \cdots \amalg E_n$.

(c) Show that because $f(I) \supset E_0 \amalg \cdots \amalg E_n$, it follows that $I = I_{n+1}$.

(10) Let S be a multiplicative subset of the commutative noetherian ring R.

(a) Show that if E is an injective R-module, then E_S is an injective R_S-module. [*Hint*: Use the fact that if M is a finitely generated R-module, then for all R-modules X, the morphism of R-modules $\operatorname{Hom}_R(M, X)_S \to \operatorname{Hom}_{R_S}(M_S, X_S)$ is an isomorphism.]

(b) Viewing each R_S-module as an R-module by means of the canonical morphism $R \to R_S$, show that every injective R_S-module is also an injective R-module.

(c) Let $\mathfrak{P}$ be a prime ideal of R and let E be an $R_{\mathfrak{P}}$-module which is an injective envelope of the $R_{\mathfrak{P}}$-module $R_{\mathfrak{P}}/\mathfrak{P}R_{\mathfrak{P}}$. Show that viewing E as an R-module, then E is an injective envelope for the R-module $R/\mathfrak{P}$.

(d) Let R be a PID with field of quotients K and (p) a prime ideal in $PPD(R)$. Show that $(K/R)_{(p)}$ is an injective envelope for each of the R-modules R/p^nR for all positive n.

(11) Let R be an integral domain. Show that its field of quotients is an injective envelope for R.

(12) Let $0 \to M_1 \to M_2 \to M_3 \to 0$ be an exact sequence of R-modules over an arbitrary ring R. Suppose we are given monomorphisms $f_i: M_i \to E_i$ with E_i injective for $i = 1, 3$. Show that there is a commutative exact diagram

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & M_1 & \longrightarrow & M_2 & \longrightarrow & M_3 & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle f_3} & & \\
0 & \longrightarrow & E_1 & \longrightarrow & E_1 \amalg E_3 & \longrightarrow & E_3 & \longrightarrow & 0
\end{array}$$

where $E_1 \to E_1 \amalg E_3$ is the usual injection morphism and $E_1 \amalg E_3 \to E_3$ is the canonical projection morphisms.

(13) Let R be a commutative noetherian ring. Show:

(a) $\operatorname{Ass}(R/\mathfrak{P}) = \{\mathfrak{P}\}$ for each prime ideal $\mathfrak{P}$ of R.

(b) If $A \subset B$ is an essential extension of R-modules, then $\operatorname{Ass}(A) = \operatorname{Ass}(B)$.

(c) If E is an injective envelope for a finitely generated R-module A, then there is a finite family $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ of prime ideals in R satisfying:

(i) $E = E_1 \amalg \cdots \amalg E_n$ where E_i is an injective envelope of $R/\mathfrak{P}_i$ for each $i = 1, \ldots, n$.

(ii) The distinct prime ideals amongst the $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are precisely the prime ideals in $\operatorname{Ass}(A)$.

(14) Let R be a commutative noetherian ring and M an R-module. Show:

(a) $\mathrm{Ass}(M) \subset \mathrm{Supp}(M)$.

(b) If S is a multiplicative subset of R, then $\mathrm{Ass}_{R_S}(M_S)$ consists of precisely those prime ideals $\mathfrak{P}R_S$ in R_S where $\mathfrak{P}$ is a prime ideal of R in $\mathrm{Ass}(M)$ such that $S \cap \mathfrak{P} = \emptyset$.

(15) Suppose A is a finitely generated module over the commutative noetherian ring R. Show that for each R-module B, we have $\mathrm{Ass}(\mathrm{Hom}_R(A, B)) = \mathrm{Ass}(B) \cap \mathrm{Supp}(A)$. We outline a proof as follows:

(a) Using the fact that $\mathrm{Hom}_{R_{\mathfrak{P}}}(A_{\mathfrak{P}}, B_{\mathfrak{P}}) = \mathrm{Hom}_R(A, B)_{(\mathfrak{P})}$ for each prime ideal $\mathfrak{P}$ of R, show that $\mathrm{Supp}(\mathrm{Hom}_R(A, B)) \subset \mathrm{Supp}(A)$ and hence $\mathrm{Ass}(\mathrm{Hom}_R(A, B)) \subset \mathrm{Supp}(A)$.

(b) Because A is a finitely generated R-module, we know there is an exact sequence $F \to A \to 0$ with F a finitely generated free R-module. Show this implies that $\mathrm{Hom}_R(A, B)$ is isomorphic to a submodule of a finite sum of copies of B which implies $\mathrm{Ass}(\mathrm{Hom}_R(A, B)) \subset \mathrm{Ass}(B)$; and so combining with (a), we have $\mathrm{Ass}(\mathrm{Hom}_R(A, B)) \subset \mathrm{Supp}(A) \cap \mathrm{Ass}(B)$.

(c) Suppose $\mathfrak{P}$ is in $\mathrm{Supp}(A) \cap \mathrm{Ass}(B)$. Then $A_{\mathfrak{P}} \neq 0$, and there is an exact sequence $0 \to R/\mathfrak{P} \to B$. Show:

(i) $\mathrm{Ass}_{R_{\mathfrak{P}}}(\mathrm{Hom}_{R_{\mathfrak{P}}}(A_{\mathfrak{P}}, R_{\mathfrak{P}}/\mathfrak{P}R_{\mathfrak{P}})) \neq 0$, and so

(ii) $\mathrm{Ass}_R(\mathrm{Hom}_R(A, R/\mathfrak{P})) = \{\mathfrak{P}\}$, which implies

(iii) $\mathfrak{P} \in \mathrm{Ass}(\mathrm{Hom}_R(A, B))$. This implies

(d) $\mathrm{Ass}(\mathrm{Hom}_R(A, B)) = \mathrm{Ass}(B) \cap \mathrm{Supp}(A)$.

(16) Show that the following statements are equivalent for a ring R (not necessarily commutative):

(a) Every submodule of a projective R-module is projective.

(b) Every factor of an injective R-module is injective. [*Hint*: To show that (b) implies (a) observe that given an exact sequence $0 \to M' \xrightarrow{\text{inc}} M \to M'' \to 0$ of R-modules and an injective module E containing M, we obtain the commutative exact diagram

$$\begin{array}{ccccccccc} & & & & 0 & & 0 & & \\ & & & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & M' & \xrightarrow{\text{inc}} & M & \longrightarrow & M'' & \longrightarrow & 0 \\ & & \| & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & M' & \xrightarrow{\text{inc}} & E & \longrightarrow & E'' & \longrightarrow & 0 \end{array}$$

with E'' an injective R-module because it is a factor of the injective R-module E.]

(17) Let R be an arbitrary ring.

(a) Show that an R-module E is an injective R-module if and only if $\mathrm{Ext}^1_R(X, E) = 0$ for each R-module X.

(b) Suppose $0 \to M \to E \to E'' \to 0$ is an exact sequence of R-modules with E an injective module. Show that for each R-module X, there is an exact sequence of abelian groups

$$\mathrm{Hom}_R(X, E) \longrightarrow \mathrm{Hom}_R(X, E'') \longrightarrow \mathrm{Ext}^1_R(X, M) \longrightarrow 0$$

Chapter 11 APPLICATIONS OF THE FUNDAMENTAL THEOREM

1. DIAGONALIZATION

Because the ring, $\mathbf{Z}$, of integers is our most familiar PID, we take a quick look at what some of the results of Chapter 10 tell us about $\mathbf{Z}$-modules, or just plain abelian groups. We know that every finitely generated abelian group is the sum of a finitely generated free abelian group, and a finitely generated torsion group. Because we know that all cyclic groups $\mathbf{Z}/(a)$ are finite if $a \neq 0$, we know that every finitely generated torsion group is finite. In fact, if G is a finitely generated torsion group, we have $G \approx \mathbf{Z}/(\epsilon_1) \amalg \cdots \amalg \mathbf{Z}/(\epsilon_n)$ where $0 \neq (\epsilon_1) \subset \cdots \subset (\epsilon_n) \neq \mathbf{Z}$. The order of $\mathbf{Z}/(\epsilon_k)$ is $|\epsilon_k|$, and so the order of G is $\Pi|\epsilon_k|$.

Alternatively, we can write $G = \coprod_i \coprod_j \mathbf{Z}/(p_i^{r_{ij}})$, and then the order of G is $\prod_i p_i^{\sum_j r_{ij}}$. Theoretically, we are now in a position to compute the number of nonisomorphic abelian groups there are of a given order. For example, suppose we are given an integer m and we are asked to find all abelian groups of order m. We may write $m = p_1^{r_1} \cdots p_t^{r_t}$. If G is an abelian group of order m, we know that $G = \coprod_{i=1}^{t} \coprod_j \mathbf{Z}/(p_i^{r_{ij}})$ with $\Sigma_j \nu_{ij} = \nu_i$. Moreover, if we take any set of positive integers $\{\nu_{ij}\}$ such that $\Sigma_j \nu_{ij} = \nu_i$, we obtain an abelian group $\coprod_i \coprod_j \mathbf{Z}/(p_i^{r_{ij}})$ of order m and distinct sets $\{\nu_{ij}\}$, $\{\nu'_{ij}\}$ yield nonisomorphic abelian groups.

If we let $\pi(\nu)$ be the number of distinct ways that the positive integer ν may be written as a sum of positive integers, then it is clear that the number of

nonisomorphic abelian groups of order m is $\pi(\nu_1)\cdots\pi(\nu_t)$ where $m = p_1^{\nu_1}\cdots p_t^{\nu_t}$. Thus, if $m = 25$, we have $25 = 5^2$ so that, because $\pi(2) = 2$, there are precisely two abelian groups of order 25, namely, $\mathbf{Z}/(5)\amalg\mathbf{Z}/(5)$ and $\mathbf{Z}/(25)$. In general, $\pi(\nu)$ is extremely difficult to calculate, and that is why we said that we are "theoretically" in a position to determine all abelian groups of a given order.

Having discussed finite abelian groups we now turn our attention to applying the fundamental theorem to finitely generated modules over arbitrary PID's.

Theorem 1.1
Let F be a free R-module of rank n over a PID R. If F' is a submodule of F such that F/F' is a torsion module, then there exists a basis $x_1, \ldots, x_n$ for F and nonzero elements $a_1, \ldots, a_n$ in R satisfying:

(a) $(a_1)\subset(a_2)\subset\cdots\subset(a_n)$.
(b) $a_1x_1, \ldots, a_nx_n$ is a basis for F'.

Furthermore, if $x_1', \ldots, x_n'$ is any other basis for F and $b_1, \ldots, b_n$ are elements of R which satisfy (a) and (b) relative to $x_1', \ldots, x_n'$, then $(b_i) = (a_i)$ for all $i = 1, \ldots, n$.

PROOF: Because F is a finitely generated free R-module and R is a PID, we know that F' is a free R-module with rank $F' \leq$ rank F. We now show that because F/F' is a torsion module, rank F' = rank F.

Lemma 1.2
Let R be an arbitrary integral domain. Suppose $F' \subset F$ is a free submodule of the free R-module F with rank $F = n$. Then rank F' = rank F if and only if F/F' is a torsion module.

PROOF: Consider the exact sequence $0\to F'\to F\to F/F'\to 0$. Then $0\to K\underset{R}{\otimes}F'\to K\underset{R}{\otimes}F\to K\underset{R}{\otimes}(F/F')\to 0$ is an exact sequence of K-modules where K is the field of quotients of R. Because F/F' is a torsion module if and only if $K\underset{R}{\otimes}(F/F') = 0$, we have that the K-morphism $k\underset{R}{\otimes}F'\to K\underset{R}{\otimes}F$ is an isomorphism if and only if F/F' is a torsion module. But $K\underset{R}{\otimes}F'\to K\underset{R}{\otimes}F$ is a K-isomorphism if and only if the K-vector spaces $K\underset{R}{\otimes}F'$ and $K\underset{R}{\otimes}F$ have the same K-dimension. But rank F' and dimension $K\underset{R}{\otimes}F'$ are the same as are rank F and dimension $K\underset{R}{\otimes}F$. Hence, F/F' is a torsion module if and only if rank F = rank F'.

With this preliminary result out of the way we can proceed to show that there is a basis $x_1, \ldots, x_n$ for F and nonzero elements $a_1, \ldots, a_n$ in R satisfying Theorem 1.1 **(a)** and **(b)**. Our proof is based on the following.

Lemma 1.3
Let R be a PID. Suppose $k: R\to R/(\epsilon)$ is the canonical epimorphism and that $h: F\to R/(\epsilon)$ is an epimorphism from a free R-module F of finite rank. Then there is an epimorphism $f: F\to R$ and an isomorphism $t: R/(\epsilon)\to R/(\epsilon)$ such that $kf = th$.

PROOF: Because $k: R \to R/(\epsilon)$ is an epimorphism and F is a projective R-module, we know there is a morphism $g: F \to R$ such that $kg = h$. Hence, $k(\text{Im } g) = R/(\epsilon)$. We leave it to the reader to take care of the case $\text{Im } g = (0)$.

Suppose $\text{Im } g = (a)$ with $a \neq 0$. Then $R/(\epsilon) = k(\text{Im } g) = (a, \epsilon)/(\epsilon)$, which implies that $(a, \epsilon) = R$. Hence, a and ϵ are relatively prime, so there exist elements t and s such that $ta + s\epsilon = 1$. From this it follows that the morphism $t: R/(\epsilon) \to R/(\epsilon)$ given by $k(r) \to tk(r)$ for all r in R is an isomorphism of R-modules.

Because $\text{Im } g$ is a free R-module with basis a, the epimorphism $g_0: F \to \text{Im } g$ has a splitting $g': \text{Im } g \to F$ and $g'(a) = x_1$ is part of a basis $x_1, \ldots, x_n$ of F. Hence, there is a morphism $v: F \to R$ such that $v(x_1) = \epsilon$ and $v(x_i) = 0$ for $i \neq 1$. Define the morphism $f: F \to R$ by $f = tg + sv$. We leave it to the reader to check that this morphism has the desired properties.

We now turn to proving the first part of Theorem 1.1. If rank $F = 1$, the reader can verify immediately that if $F' \subset F$ is a nonzero free submodule of F, then given any basis x_1 of F there is an element a_1 in R such that $a_1 x_1$ is a basis for F'.

Suppose now that rank $F = n > 1$. Then F/F' is a finitely generated torsion module so $F/F' = R/(\epsilon_1) \amalg \cdots \amalg R/(\epsilon_m)$ with $(\epsilon_1) \subset (\epsilon_2) \cdots \subset (\epsilon_m)$. We therefore have the projection morphism $p: F/F' \to R/(\epsilon_1)$ whose kernel is $R/(\epsilon_2) \amalg \cdots \amalg R/(\epsilon_n)$. Define $h: F \to R/(\epsilon_1)$ to be the composition $F \to F/F' \to R/(\epsilon_1)$ where $F \to F/F'$ is the canonical epimorphism. By Lemma 1.3, we have the commutative diagram

$$\begin{array}{ccccc} F & \longrightarrow & F/F' & \longrightarrow & 0 \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle tp} & & \\ R & \xrightarrow{k} & R/(\epsilon_1) & \longrightarrow & 0 \\ \downarrow & & \downarrow & & \\ 0 & & 0 & & \end{array}$$

with exact rows and columns. Clearly, $f(F') \subset (\epsilon_1)$ since $kf(x') = 0$ for all x' in F'. Because (ϵ_1) is the annihilator F/F', we have that $\epsilon_1 F \subset F'$. Because there is an x_1 in F such that $f(x_1) = 1$, it follows that $f(\epsilon_1 x_1) = \epsilon_1$, so that $f(F') = (\epsilon_1)$. Letting $f': F' \to (\epsilon_1)$ be the morphism given by $f'(x') = f(x')$ for all x' in F', we obtain the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & F' & \longrightarrow & F & \longrightarrow & F/F' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f'} & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle tp} & & \\ 0 & \longrightarrow & (\epsilon_1) & \longrightarrow & R & \xrightarrow{k} & R/(\epsilon_1) & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 0 & & 0 & & 0 & & \end{array}$$

with exact rows and columns.

Let $F_1 = \text{Ker } f$ and $F_1' = \text{Ker } f'$. Then $F_1' \subset F_1$ and $F_1' = F' \cap F_1$. This shows that $\epsilon_1 F_1 \subset F_1'$ because $\epsilon_1 F \subset F'$. Let x_1 be an element of F such that $f(x_1) = 1$. Then $F = F_1 \amalg (x_1)$ and $F' = F_1' \amalg (\epsilon_1 x_1)$ where (x_1) and $(\epsilon_1 x_1)$ denote the submodules generated by x_1 and $\epsilon_1 x_1$. Thus, rank $F_1 = \text{rank } F_1' = n - 1$ where $n = \text{rank } F$. Then, by induction, there is a basis $x_2, \ldots, x_n$ of F_1 and nonzero elements

$a_2, \ldots, a_n$ in R such that $a_2x_2, \ldots, a_nx_n$ is a basis for F_1' and $(a_2) \subset (a_3) \subset \cdots \subset (a_n)$. From this it follows that $(a_2) = \text{ann}(F_1/F_1')$. Because $\epsilon_1 F_1 \subset F_1'$, we know that ϵ_1 is in $\text{ann}(F_1/F_1') = (a_2)$. Thus, $(\epsilon_1) \subset (a_2) \cdots \subset (a_n)$. Setting $a_1 = \epsilon_1$, we see that the basis $x_1, \ldots, x_n$ of F and the elements $a_1, \ldots, a_n$ of R satisfy the conditions of Theorem 1.1. This completes the proof of the first part of the theorem.

The uniqueness part of the proof is left as an exercise.

Corollary 1.4

Let F_1 and F be free modules of finite rank over the PID R, and let $f: F_1 \to F$ be a morphism. Then F_1 has a basis $\{z_1, \ldots, z_t\}$, F has a basis $\{x_1, \ldots, x_s\}$, and there are nonzero elements $a_1, \ldots, a_m \in R$ with $(a_1) \subset \cdots \subset (a_m)$ such that $f(z_i) = a_i x_i$ for $i = 1, \ldots, m$ and $f(z_j) = 0$ for $j > m$. If $\{z_1', \ldots, z_t'\}$, $\{x_1', \ldots, x_s'\}$ and $\{a_1', \ldots, a_m'\}$ satisfy the same conditions, then $m' = m$, $(a_i') = (a_i)$ for $i = 1, \ldots, m$, and the submodules generated by $\{x_1', \ldots, x_m'\}$ and $\{x_1, \ldots, x_m\}$ are equal.

PROOF: Let $\text{Im } f = F'$. We know that $F/F' = T \amalg G$ where T is the torsion submodule of F/F' and G is a free R-module. Let $k': F \to F/F'$ be the canonical epimorphism and $k^{-1}(T) = L$. Then L is a free submodule of F containing F' and $\text{rank } L = \text{rank } F'$ because $L/F' = T$, which is a torsion module. Also, $F/L \approx G$. Hence, $F = L \amalg L'$ where L' is a free R-module isomorphic to G. Applying Theorem 1.1 we can find a basis $x_1, \ldots, x_m$ for L and nonzero elements $a_1, \ldots, a_m$ in R such that $a_1x_1, \ldots, a_mx_m$ is a basis for F' and $(a_1) \subset (a_2) \cdots \subset (a_m)$.

Letting $x_{m+1}, \ldots, x_s$ be a basis for L', we have that $x_1, \ldots, x_m, x_{m+1}, \ldots, x_s$ is a basis for F. Because the morphism $f_0: F_1 \to F'$ is a splittable epimorphism, there is a splitting $g: F' \to F_1$ for f_0 and $F_1 = g(F') \amalg \text{Ker } f_0$. Let $z_i = g(x_i)$ for $i = 1, \ldots, m$ and $z_{m+1}, \ldots, z_t$ be a basis for $\text{Ker } f_0$. Then $z_1, \ldots, z_m, z_{m+1}, \ldots, z_t$ is a basis for F_1. It is easily verified that the bases $x_1, \ldots, x_s$ and $z_1, \ldots, z_t$ for F and F_1, respectively, together with the elements $a_1, \ldots, a_m$ in R satisfy the conditions of the corollary.

The uniqueness proof is again left as an exercise.

If F_1 and F are free modules of finite ranks t and s, respectively, over a commutative ring R, and if $\{z_1, \ldots, z_t\}$, $\{x_1, \ldots, x_s\}$ are bases for F_1 and F, respectively, any morphism $f: F_1 \to F$ gives rise to a matrix (a_{ij}) with t rows and s columns where $f(z_i) = \Sigma\, a_{ij}x_j$. Conversely, any matrix (a_{ij}) with t rows and s columns gives rise to a morphism $f: F_1 \to F$ by defining $f(z_i) = \Sigma\, a_{ij}x_j$. In other words, the setup is exactly the same as for finite-dimensional vector spaces over a field. What Corollary 1.4 tells us is that if R is a PID and $f: F_1 \to F$ is a morphism, then we may find bases $\{z_1, \ldots, z_t\}$ and $\{x_1, \ldots, x_s\}$ of of F_1 and F, respectively, such that with respect to these bases the morphism f has the matrix

$$\begin{pmatrix} a_1, 0, \ldots, 0 & 0, \ldots, 0 \\ 0, a_2, \ldots, 0 & 0, \ldots, 0 \\ \vdots & \\ 0 \qquad\quad a_n, & 0, \ldots, 0 \\ 0 \qquad\quad 0 & 0, \ldots, 0 \\ \vdots \qquad\quad \vdots & \vdots \\ 0, \ldots, \quad 0, & 0, \ldots, 0 \end{pmatrix}$$

where $(0) \neq (a_1) \subset \cdots \subset (a_t)$. For this reason, Corollary 1.4 is referred to as the **diagonalization theorem for matrices over a PID**. Notice that if $s = t$ and f is a monomorphism, then $s = t = n$ and Coker f is a torsion module. Conversely, if $s = t$ and Coker f is a torsion module, then f is a monomorphism. (Why?)

2. DETERMINANTS

Having reached the point of considering matrices, we digress for a bit and study determinants. We assume that the reader is acquainted with matrices and determinants over fields, and we shall use this familiarity as a starting point for introducing determinants over an arbitrary commutative ring.

Suppose that we have a 2×2 matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ with a_{ij} in a field R. Then we ordinarily say that the determinant of this matrix is $a_{11}a_{22} - a_{12}a_{21}$. Of course, from the definition, we see that there is no reason to suppose that R is a field; it might just as well be any commutative ring. Therefore, we see that to every 2×2 matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ with a_{ij} in a commutative ring R, we may associate an element of R, namely, $a_{11}a_{22} - a_{21}a_{12}$.

Now let us suppose that F is a free R-module of rank 2, and that $\{x_1, x_2\}$ is a basis for F. Then if y_1 and y_2 are any two elements of F, we have $y_1 = a_{11}x_1 + a_{12}x_2$ and $y_2 = a_{21}x_1 + a_{22}x_2$ with a_{ij} in R. Therefore, we can unambiguously associate to the ordered pair of elements (y_1, y_2) the matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and, to this matrix we can associate its determinants $a_{11}a_{22} - a_{21}a_{12}$. It should be observed that this matrix associated to the ordered pair (y_1, y_2) depends not only on the chosen basis $\{x_1, x_2\}$ but also on the order in which they are written.

Thus, having chosen a basis for F in a given order, we can define a map σ from the set $F \times F$ to R which to every element (y_1, y_2) of $F \times F$ assigns the element $a_{11}a_{22} - a_{21}a_{12}$ of R with a_{ij} defined as above. The reader might also note that if $f: F \to F$ is defined to be the endomorphism of F which sends x_1 to y_1 and x_2 to y_2, then the matrix of f with respect to the basis $\{x_1, x_2\}$ is precisely $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

The map $\sigma: F \times F \to R$ has a few properties that are worth noting. First, σ is bilinear. Second, if $y_1 = y_2$, then $\sigma(y_1, y_2) = 0$. For this reason, σ is said to be a skew-symmetric bilinear map of $F \times F$ to R. The term "skew-symmetric" is explained by the following observation. If it were true that $\sigma(y_1, y_2) = \sigma(y_2, y_1)$ for all y_1, y_2 in F, we would be tempted to say that σ is "symmetric," that is, its value does not depend on the order of the ordered pair (y_1, y_2). However, we can show that this is, in general, not the case; in fact, $\sigma(y_1, y_2) = -\sigma(y_2, y_1)$ for all (y_1, y_2) in $F \times F$. To see this, consider $\sigma(y_1 + y_2, y_1 + y_2)$. From the fact that $\sigma(y, y) = 0$ for all $y \in F$, we know that $\sigma(y_1 + y_2, y_1 + y_2) = 0$. However, by the property of bilinearity of σ, we have $0 = \sigma(y_1 + y_2, y_1 + y_2) = \sigma(y_1, y_1 + y_2) + \sigma(y_2, y_1 + y_2) = \sigma(y_1, y_1) + \sigma(y_1, y_2) + \sigma(y_2, y_1) + \sigma(y_2, y_2) = 0 + \sigma(y_1, y_2) + \sigma(y_2, y_1) + 0$. Thus, $0 = \sigma(y_1, y_2) + \sigma(y_2, y_1)$ or $\sigma(y_1, y_2) = -\sigma(y_2, y_1)$. For this reason, σ is said to be skew-symmetric, the prefix "skew" referring to the minus sign.

In all of the foregoing discussion, we have considered 2×2 matrices and free modules of rank 2. There is nothing sacred about the number two and so we are led to make the following:

Definition
Let M be a module over the commutative ring R. A map $\beta: M\times\cdots\times M\to R$ from the n-fold Cartesian product of M to R is said to be **n-linear** if for all $m_1,\ldots,m_n$, $m\in M$, and all r, s in R, we have $\beta(m_1,\ldots,rm_i+sm,m_{i+1},\ldots,m_n)=r\beta(m_1,\ldots,m_i,m_{i+1},\ldots,m_n)+s\beta(m_1,\ldots,m,m_{i+1},\ldots,m_n)$ for each $1=1,\ldots,n$. The n-linear map $\beta: M\times\cdots\times M\to R$ is said to be **skew-symmetric** if $\beta(m_1,\ldots,m_n)=0$ whenever $m_i=m_{i+1}$ for some $i=1,\ldots,n-1$.

Basic Properties 2.1

(a) If $\beta: M\times\cdots\times M\to R$ is an n-linear skew-symmetric map, then $\beta(m_1,\ldots,m_j,\ldots,m_k,\ldots,m_n)=-\beta(m_1,\ldots,m_k,\ldots,m_j,\ldots,m_n)$ for all $1\le j<k\le n$. Consequently, $\beta(m_1,\ldots,m_n)=0$ whenever $m_j=m_k$ for some $j\neq k$.

(b) Let $\mathrm{Sk}_n(M)$ be the set of all skew-symmetric n-linear maps from $M\times\cdots\times M$ to R. For $\beta_1,\beta_2\in\mathrm{Sk}_n(M)$ and $r\in R$, define $\beta_1+\beta_2: M\times\cdots\times M\to R$ by $(\beta_1+\beta_2)(m_1,\ldots,m_n)=\beta_1(m_1,\ldots,m_n)+\beta_2(m_1,\ldots,m_n)$ and define $r\beta_1: M\times\cdots\times M\to R$ by $(r\beta_1)(m_1,\ldots,m_n)=r\cdot\beta_1(m_1,\ldots,m_n)$. Then $\beta_1+\beta_2$ is in $\mathrm{Sk}_n(M)$, $r\beta_1$ is in $\mathrm{Sk}_n(M)$, and with this definition of addition and of the operation of R on $\mathrm{Sk}_n(M)$, $\mathrm{Sk}_n(M)$ is an R-module.

(c) If $f: M\to N$ is a morphism of R-modules, then we have a map $f\times\cdots\times f: M\times\cdots\times M\to N\times\cdots\times N$. If $\beta\in\mathrm{Sk}_n(N)$, then the map $\beta\cdot(f\times\cdots\times f): M\times\cdots\times M\to R$ is in $\mathrm{Sk}_n(M)$. Thus, we have a map $\mathrm{Sk}_n(f): \mathrm{Sk}_n(N)\to\mathrm{Sk}_n(M)$ defined by $\mathrm{Sk}_n(f)(\beta)=\beta\cdot(f\times\cdots\times f)$. This map is a morphism of R-modules. Finally:

(d) If g is a morphism $g: N\to W$, we have the equality $\mathrm{Sk}_n(gf)=\mathrm{Sk}_n(f)\mathrm{Sk}_n(g)$.

PROOF: (a) We prove this assertion by induction on $k-j$. When $k-j=1$, the proof proceeds as in our discussion of bilinear maps. For we have

$$\begin{aligned}0&=\beta(m_1,\ldots,m_j+m_{j+1},m_j+m_{j+1},\ldots,m_n)\\&=\beta(m_1,\ldots,m_j,m_j+m_{j+1},\ldots,m_n)+\beta(m_1,\ldots,m_{j+1},m_j+m_{j+1},\ldots,m_n)\\&=0+\beta(m_1,\ldots,m_j,m_{j+1},\ldots,m_n)+\beta(m_1,\ldots m_{j+1},,m_j,\ldots,m_n)+0\end{aligned}$$

Thus,

$$0=\beta(m_1,\ldots,m_j,m_{j+1},\ldots,m_n)+\beta(m_1,\ldots,m_{j+1},m_j,\ldots,m_n)$$

or

$$\beta(m_1,\ldots,m_j,m_{j+1},\ldots,m_n)=-\beta(m_1,\ldots,m_{j+1},m_j,\ldots,m_n)$$

Having disposed of the case $k-j=1$, we take care of the case $k-j=p$ with $p>1$ as follows. We have

$$\begin{aligned}\beta(m_1,\ldots,m_j,\ldots,m_{k-1},m_k,\ldots,m_n)&=-\beta(m_1,\ldots,m_j,\ldots,m_k,m_{k-1},\ldots,m_n)\\&=\beta(m_1,\ldots,m_k,\ldots,m_j,m_{k-1},\ldots,m_n)\\&=-\beta(m_1,\ldots,m_k,\ldots,m_{k-1},m_j,\ldots,m_n)\end{aligned}$$

and we are done.

Now if $m_j = m_k$ for some $1 \le j < k \le n$, we have

$$\beta(m_1, \ldots, m_j, \ldots, m_k, \ldots, m_n) = -\beta(m_1, \ldots, m_j, m_k, m_{j+1}, \ldots, m_n) = 0$$

and this completes the proof of (**a**).

(**b**), (**c**), and (**d**) The reader can verify easily that $\beta_1 + \beta_2$ and $r\beta_1$ are again in $\mathrm{Sk}_n(M)$. The only thing that need be observed in order to show that $\mathrm{Sk}_n(M)$ is an R-module is that the zero map, that is, the map which sends $(m_1, \ldots, m_n)$ to 0 for all $m_1, \ldots, m_n \in M$, is in $\mathrm{Sk}_n(M)$ and is the zero element of $\mathrm{Sk}_n(M)$. The fact that $\mathrm{Sk}_n(f)$ is a morphism follows from straightforward computation. So does the equality $\mathrm{Sk}_n(gf) = \mathrm{Sk}_n(f)\mathrm{Sk}_n(g)$.

Proposition 2.2
If F is a free R-module of rank n, then $\mathrm{Sk}_n(F)$ is a free R-module of rank 1.

PROOF: Let F be a free R-module with basis $\{x_1, \ldots, x_n\}$. We shall prove that $\mathrm{Sk}_n(F)$ is free of rank 1 by induction on n. It is clear that when $n = 1$, $\mathrm{Sk}_1(M) = (M, R)$ for any module M, because to say that a map $\beta : M \to R$ is 1-linear is equivalent to saying that it is a morphism, and the condition of skew-symmetry is vacuous. Thus, if F is a free module with basis $\{x_1\}$, we have $\mathrm{Sk}_1(F) = (F, R)$ and (F, R) is free of rank 1 because the morphism $f : F \to R$ defined by $f(x_1) = 1$ is a basis for (F, R).

Assuming that our proposition is true for free modules of rank $n-1$ we shall prove it for free modules of rank n, by establishing an isomorphism between $\mathrm{Sk}_n(F)$ and $\mathrm{Sk}_{n-1}(F')$ where F' is the submodule of F generated by the subset $\{x_1, \ldots, x_{n-1}\}$ of the basis $\{x_1, \ldots, x_n\}$ of F.

Define $h : \mathrm{Sk}_n(F) \to \mathrm{Sk}_{n-1}(F')$ by $h(\beta)(y_1, \ldots, y_{n-1}) = \beta(x_n, y_1, \ldots, y_{n-1})$ for all $\beta \in \mathrm{Sk}_n(F)$. From the fact that β is n-linear and skew-symmetric, it follows easily that $h(\beta)$ is $(n-1)$-linear and skew-symmetric, so that $h(\beta)$ is in $\mathrm{Sk}_{n-1}(F')$. It is also trivial to show that h is a morphism. To see that h is an isomorphism, it will suffice to show that there is a map $h' : \mathrm{Sk}_{n-1}(F') \to \mathrm{Sk}_n(F)$ such that the compositions hh' and $h'h$ are the appropriate identity maps. It then follows that h' is itself a morphism and is the inverse of h.

Let $\beta' : F' \times \cdots \times F' \to R$ be an element of $\mathrm{Sk}_{n-1}(F')$; and let $m_1, \ldots, m_n$ be elements of F. Because F is the sum of F' and the submodule generated by x_n, each m_i may be written uniquely as $m_i = y_i + r_i x_n$ with $y_i \in F'$ and $r_i \in R$. For each $i = 1, \ldots, n$, let $(y_1, \ldots, \hat{y}_i, \ldots, y_n)$ denote the $(n-1)$-tuple $(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ in $F' \times \cdots \times F'$. Now define $h'(\beta') : F \times \cdots \times F \to R$ by $h'(\beta')(m_1, \ldots, m_n) = \Sigma_{i=1}^{n} (-1)^{i+1} r_i \beta'(y_1, \ldots, \hat{y}_i, \ldots, y_n)$. Our remarks about the uniqueness of the elements y_i and r_i indicate that $h'(\beta')$ is a map from the n-fold Cartesian product $F \times \cdots \times F$ to R. What we shall show is that $h'(\beta')$ is actually in $\mathrm{Sk}_n(F)$; that is, $h'(\beta')$ is n-linear and skew-symmetric.

To show n-linearity, let $m_1, \ldots, m_n, m$ be elements of F, and let r, s be elements of R. We have, as before, $m_i = y_i + r_i \cdot x_n$, and we set $m = z + tx_n$ with $z \in F'$ and $t \in R$. Then $rm_i = ry_i + rr_i x_n$ and $sm = sz + st x_n$ so that

$$rm_i + sm = (ry_i + sz) + (rr_i + st)x_n$$

Computing $h'(\beta')(m_1, \ldots, rm_j + rm, m_{j+1}, \ldots, m_n)$, we get

$$\begin{aligned}
&h'(\beta')(m_1, \ldots, rm_j + sm, m_{j+1}, \ldots, m_n) \\
&\qquad = \sum_{i<j} (-1)^{i+1} r_i \beta'(y_1, \ldots, y_i, \ldots, ry_j + sz, \ldots, y_n) \\
&\qquad\quad + (-1)^{j+1}(rr_j + st)\beta'(y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_n) \\
&\qquad\quad + \sum_{j<i} (-1)^{i+1} r_i \beta'(y_1, \ldots, ry_j + sz, \ldots, \hat{y}_i, \ldots, y_n) \\
&\qquad = \sum_{i<j} (-1)^{i+1} r_i r \beta'(y_1, \ldots, \hat{y}_i, \ldots, y_j, \ldots, y_n) \\
&\qquad\quad + \sum_{i<j} (-1)^{i+1} r_i s \beta'(y_1, \ldots, \hat{y}_i, \ldots, z, \ldots, y_n) \\
&\qquad\quad + (-1)^{j+1} r r_j \beta'(y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_n) \\
&\qquad\quad + (-1)^{j+1} st \beta'(y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_n) \\
&\qquad\quad + \sum_{j<i} (-1)^{i+1} r r_i \beta'(y_1, \ldots, y_j, \ldots, \hat{y}_i, \ldots, y_n) \\
&\qquad\quad + \sum_{j<i} s r_i \beta'(y_1, \ldots, z, \ldots, \hat{y}_i, \ldots, y_n) \\
&\qquad = r h'(\beta')(m_1, \ldots, m_j, \ldots, m_n) + s h'(\beta')(m_1, \ldots, m, \ldots, m_n)
\end{aligned}$$

Thus, $h'(\beta')$ is n-linear. To see that $h'(\beta')$ is skew-symmetric, we take $m_1, \ldots, m_n \in F$, and assume that $m_j = m_{j+1}$ for some j. We want to show that $h'(\beta')(m_1, \ldots, m_n) = 0$. But

$$\begin{aligned}
&h'(\beta')(m_1, \ldots, m_n) \\
&\qquad = \sum (-1)^{i+1} r_i \beta'(y_1, \ldots, \hat{y}_i, \ldots, y_n) \\
&\qquad = \sum_{i<j} (-1)^{i+1} r_i \beta'(y_1, \ldots, \hat{y}_i, \ldots, y_j, y_{j+1}, \ldots, y_n) \\
&\qquad\quad + \sum_{j+1<i} (-1)^{i+1} r_i \beta'(y_1, \ldots, y_j, \ldots, y_{j+1}, \ldots, \hat{y}_i, \ldots, y_n) \\
&\qquad\quad + (-1)^{j+1} r_j \beta'(y_1, \ldots, \hat{y}_j, y_{j+1}, \ldots, y_n) \\
&\qquad\quad + (-1)^{j+2} r_{j+1} \beta'(y_1, \ldots, y_j, \hat{y}_{j+1}, \ldots, y_n)
\end{aligned}$$

Since $m_j = m_{j+1}$, we have $y_j = y_{j+1}$ and $r_j = r_{j+1}$ so that

$$(-1)^{j+1} r_j \beta'(y_1, \ldots, \hat{y}_j, y_{j+1}, \ldots, y_n) + (-1)^{j+2} r_{j+1} \beta'(y_1, \ldots, t_j, \hat{y}_{j+1}, \ldots, y_n) = 0$$

Moreover, because β' itself is skew-symmetric, each of the terms $r_i \beta'(y_1, \ldots, \hat{y}_i, \ldots, y_j, y_{j+1}, \ldots, y_n)$ and $r_i \beta'(y_1, \ldots, y_j, y_{j+1}, \ldots, \hat{y}_i, \ldots, y_n)$ is zero, and so $h'(\beta')(m_1, \ldots, m_n) = 0$.

We have now shown that for each $\beta' \in \mathrm{Sk}_{n-1}(F')$, the map $h'(\beta')$ is in $\mathrm{Sk}_n(F)$, and so we define the map $h' : \mathrm{Sk}_{n-1}(F') \to \mathrm{Sk}_n(F)$ by sending β' to $h'(\beta')$. All that remains to be shown is that if $\beta \in \mathrm{Sk}_n(F)$ and $\beta' \in \mathrm{Sk}_{n-1}(F')$, then $h'h(\beta) = \beta$ and $hh'(\beta') = \beta'$. First we consider $h'h(\beta)$. We have $h'h(\beta) \times (m_1, \ldots, m_n) = \Sigma(-1)^{i+1} r_i h(\beta)(y_1, \ldots, \hat{y}_i, \ldots, y_n)$ where the notation is that which we have consistently been using. But $\Sigma(-1)^{i+1} r_i h(\beta)(y_1, \ldots, \hat{y}_i, \ldots, y_n) = \Sigma(-1)^{i+1} r_i \beta(x_n, y_1, \ldots, \hat{y}_i, \ldots, y_n)$ by the definition of $h(\beta)$. Hence, to

show that $h'h(\beta) = \beta$, we must show that

$$\beta(m_1, \ldots, m_n) = \Sigma (-1)^{i+1} r_i \beta(x_n, y_1, \ldots, \hat{y}_i, \ldots, y_n)$$

But

$$\beta(m_1, \ldots, m_n) = \beta(y_1 + r_1 x_n, \ldots, y_n + r_n x_n) = \beta(y_1, \ldots, y_n) + \sum_{i=1}^{n} \beta(y_1, \ldots, y_{i-1}, r_i x_n, y_{i+1}, \ldots, y_n)$$

This last equality is obtained by using the n-linearity of β and observing that the terms of the form $\beta(y_1, \ldots, r_i x_n, \ldots, r_j x_n, \ldots, y_n)$ are all zero because $\beta(y_1, \ldots, r_i x_n, \ldots, r_j x_n, \ldots, y_n) = r_i r_j \beta(y_1, \ldots, x_n, \ldots, x_n, \ldots, y_n)$ and β is skew-symmetric. Similarly, all terms having more than two entries which are multiples of x_n are zero and so $\beta(m_1, \ldots, m_n)$ reduces to what we have written.

Now $\beta(y_1, \ldots, r_i x_n, \ldots, y_n) = r_i \beta(y_1, \ldots, x_n, \ldots, y_n)$ and applying Basic Properties 2.1, we have $\beta(y_1, \ldots, x_n, \ldots, y_n) = (-1)^{i+1} \beta(x_n, y_1, \ldots, \hat{y}_i, \ldots, y_n)$ if x_n is in the ith spot. Hence, $\Sigma \beta(y_1, \ldots, r_i x_n, \ldots, y_n) = \Sigma (-1)^{i+1} r_i \beta(x_n, y_1, \ldots, \hat{y}_i, \ldots, y_n)$, and we would be done provided we could show that $\beta(y_1, \ldots, y_n) = 0$. The fact that $\beta(y_1, \ldots, y_n) = 0$ comes from the following.

Lemma 2.3

Let M be an R-module, M' a submodule generated by p elements, and let $\{y_1, \ldots, y_n\}$ be a subset of M' with $n > p$. If $\beta : M \times \cdots \times M \to R$ is an n-linear skew-symmetric map, then $\beta(y_1, \ldots, y_n) = 0$.

PROOF: Let $x_1, \ldots, x_p$ generate M' and let $y_i = \sum_{j=1}^{p} a_{ij} x_j$ for $i = 1, \ldots, n$. Using the n-linearity of β, we have $\beta(y_1, \ldots, y_n) = \beta(\Sigma a_{1j_1} x_{j_1}, \ldots, \Sigma a_{nj_n} x_{j_n}) = \Sigma a_{1j_1} \cdots a_{nj_n} \beta(x_{j_1}, \ldots, x_{j_n})$. However, because the indices j_k just range from 1 to p, and because $n > p$, we must always have at least two of the x's equal. Hence, we have $\beta(x_{j_1}, \ldots, x_{j_n}) = 0$ for all $j_1, \ldots, j_n$ and thus $\beta(y_1, \ldots, y_n) = 0$.

In the case at hand, we have $y_1, \ldots, y_n \in F'$ with F' of rank $n-1$. Hence, $\beta(y_1, \ldots, y_n) = 0$ and we have shown that $h'h(\beta) = \beta$.

Next we confirm the fact that $hh'(\beta') = \beta'$. For this, we compute $hh'(\beta')(y_1, \ldots, y_{n-1})$ and we get $hh'(\beta')(y_1, \ldots, y_{n-1}) = h'(\beta')(x_n, y_1, \ldots, y_{n-1}) = \beta'(y_1, \ldots, y_n)$ because $y_i = y_i + 0\, x_n$ for $i = 1, \ldots, n-1$. Thus, $hh'(\beta') = \beta'$ and our proof of Proposition 2.2 is complete.

Having established these results, how do we proceed to determinants? Consider a free module F of rank n, and let $f : F \to F$ be an endomorphism of F. Then, by Basic Properties 2.1, we have an endomorphism $\mathrm{Sk}_n(f) : \mathrm{Sk}_n(F) \to \mathrm{Sk}_n(F)$. Using Proposition 2.2, we know that $\mathrm{Sk}_n(F)$ is a free module of rank one. Putting that together with the following fact, we shall be in a position to define determinants.

Lemma 2.4

Let R be a commutative ring and G a free R-module of rank one. Let $\theta : R \to (G, G)$ be the morphism defined by $\theta(r)(g) = rg$. [$\theta(r)$ may also be described as $\theta(r) = r\, \mathrm{id}_G$, because (G, G) is an R-module.] Then θ is an isomorphism.

PROOF: That θ is a morphism is easily established and, because Ker θ is the annihilator of G, θ is in fact a monomorphism. To see that θ is an epimorphism, we choose a basis $\{x\}$ of G and let $f: G \to G$ be any element of (G, G). Then $f(x) = rx$ for some $r \in R$ so that $f(y) = ry$ for all $y \in G$. Hence, $f = \theta(r)$ and θ is an epimorphism; hence, an isomorphism.

It is important to notice that although it required choosing a basis of G to prove that θ is an isomorphism, the morphism θ itself is defined without recourse to a basis of G. In practical, computational terms, what this means is the following. If $\{x\}$ and $\{y\}$ are two bases of G, and $f: G \to G$ is an endomorphism of G, we know that $f(x) = rx$ and $f(y) = ry$, where $f = \theta(r)$. In short, the element r is determined by f and is independent of the basis chosen for G. The element r is $\theta^{-1}(f)$ and is determined by evaluating f on *any* basis of G.

Definition

Let F be a free R-module of rank n, and let $f: F \to F$ be an endomorphism of F. The **determinant of** f is defined to be $\theta^{-1}(\mathrm{Sk}_n(f))$ where $\theta: R \to (\mathrm{Sk}_n(F), \mathrm{Sk}_n(F))$ is the isomorphism described in Lemma 2.4. The determinant of f will be denoted either by $\det f$ or $|f|$.

Basic Properties 2.5

(a) Let F be a free module of finite rank n. Then $|r\,\mathrm{id}_F| = r^n$. In particular $|\mathrm{id}_F| = 1$.

(b) If F is a free R-module of finite rank and f and g are endomorphisms of F, then $|fg| = |f||g|$. Hence, $|fg| = |gf|$.

(c) Let F be a free R-module of finite rank n and f an endomorphism of F. If $\{x_1, \ldots, x_n\}$ is a basis for F, let $f(x_i) = \sum_{j=1}^{n} a_{ij}x_j$ for $i = 1, \ldots, n$. Then $|f| = \sum_{\pi} \mathrm{sig}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)}$ where π runs through all permutations of the set $\{1, \ldots, n\}$, and $\mathrm{sig}(\pi) = \pm 1$ depending on whether π is an even or odd permutation.

(d) Let F be a free R-module of rank n and f an endomorphism of F. If $\{x_1, \ldots, x_n\}$ is a basis of F and $f(x_i) = f(x_j)$ for some $i \neq j$, then $|f| = 0$.

PROOF: (a) is left to the reader.

(b) To prove (b), recall that $\mathrm{Sk}_n(f)(\beta) = |f|\beta$ and $\mathrm{Sk}_n(g)(\beta) = |g|\beta$ for all $\beta \in \mathrm{Sk}_n(F)$. Moreover, it is clear that for any $h: F \to F$, if $\mathrm{Sk}_n(h)(\beta) = r\beta$ for all $\beta \in \mathrm{Sk}_n(F)$, then $|h| = r$. Now, we compute $\mathrm{Sk}_n(fg)(\beta)$ for any $\beta \in \mathrm{Sk}_n(F)$. We have $\mathrm{Sk}_n(fg)(\beta) = (\mathrm{Sk}_n(g)\mathrm{Sk}_n(f))(\beta) = |f|(|g|\beta) = |f||g|\beta$. Hence, we have $|fg| = |f||g|$. Because $|f||g| = |g||f| = |gf|$, we also have $|fg| = |gf|$.

(c) To compute $|f|$, we choose a basis of $\mathrm{Sk}_n(F)$ and see what $\mathrm{Sk}_n(f)$ does to that basis. The first problem then is how to go about finding a basis of $\mathrm{Sk}_n(F)$. In the proof of Proposition 2.2, we chose a basis $\{x_1, \ldots, x_n\}$ of F, defined F' to be the submodule generated by $\{x_1, \ldots, x_{n-1}\}$, and established an isomorphism $h': \mathrm{Sk}_{n-1}(F') \to \mathrm{Sk}_n(F)$. Thus, if we know a basis for $\mathrm{Sk}_{n-1}(F')$, we know one for $\mathrm{Sk}_n(F)$. In the case $n = 1$, we have already seen that a basis for $\mathrm{Sk}_1(F)$ may be obtained by considering the morphism ξ which sends x_1 to 1 in R where $\{x_1\}$ is a basis for F. Furthermore, an element β of $\mathrm{Sk}_1(F)$ is completely determined by its value on x_1 and $\beta(x_1) = r$ if and only if $\beta = r\xi$.

Let us now assume that if $\{x_1, \ldots, x_{n-1}\}$ is a basis for F', then: (1) any element β of $\mathrm{Sk}_{n-1}(F')$ is completely determined by its value on $(x_1, \ldots, x_{n-1})$ in $F' \times \cdots \times F'$; (2) there is an element $\xi' \in \mathrm{Sk}_{n-1}(F')$ such that $\xi'(x_1, \ldots, x_{n-1}) = 1$; (3) this element ξ' is a basis for $\mathrm{Sk}_{n-1}(F')$; and (4) $\beta(x_1, \ldots, x_{n-1}) = r$ if and only if $\beta = r\xi'$ for all $\beta \in \mathrm{Sk}_{n-1}(F')$. We shall show that if F is free with basis $\{x_1, \ldots, x_n\}$, then the corresponding four statements are true for $\mathrm{Sk}_n(F)$.

To prove (1), suppose that $\beta_1, \beta_2 \in \mathrm{Sk}_n(F)$, and that $\beta_1(x_1, \ldots, x_n) = \beta_2(x_1, \ldots, x_n)$. Using the morphism $h: \mathrm{Sk}_n(F) \to \mathrm{Sk}_{n-1}(F')$ established earlier (where F' is generated by $(x_1, \ldots, x_n)$, we have $(h(\beta_1))(x_1, \ldots, x_{n-1}) = \beta_1(x_n, x_1, \ldots, x_{n-1}) = (-1)^{n-1}\beta_1(x_1, \ldots, x_n) = (-1)^{n-1}\beta_2(x_1, \ldots, x_n) = \beta_2(x_n, x_1, \ldots, x_{n-1}) = (h(\beta_2))(x_1, \ldots, x_{n-1})$. Hence, by our assumption about F', we have $h(\beta_1) = h(\beta_2)$, so (because h is an isomorphism) $\beta_1 = \beta_2$.

To prove (2), we take the element $\xi' \in \mathrm{Sk}_{n-1}(F')$ such that $\xi'(x_1, \ldots, x_{n-1}) = 1$. Using the morphism $h': \mathrm{Sk}_{n-1}(F') \to \mathrm{Sk}_n(F)$ which is the inverse of h, we have $h'(\xi')(x_1, \ldots, x_n) = (-1)^{n+1}\xi'(x_1, \ldots, x_{n-1}) = (-1)^{n+1}$. Thus, $h'((-1)^{n+1}\xi') \times (x_1, \ldots, x_n) = 1$, and we may choose $\xi = h'((-1)^{n+1}\xi')$.

Because ξ' is a basis of $\mathrm{Sk}_{n-1}(F')$ and h' is an isomorphism, it is clear that ξ is a basis of $\mathrm{Sk}_n(F)$, and (3) is established.

Finally, to prove (4), it is clear that if $\beta = r\xi$, then $\beta(x_1, \ldots, x_n) = r$. But conversely, if $\beta(x_1, \ldots, x_n) = r$, then $h(\beta)(x_1, \ldots, x_{n-1}) = \beta(x_n, x_1, \ldots, x_{n-1}) = (-1)^{n-1}\beta(x_1, \ldots, x_n) = (-1)^{n-1}r$, so we have $h(\beta) = (-1)^{n-1}r\xi'$, because (4) holds for F'. However, $\beta = h'h(\beta) = h'((-1)^{n-1}r\xi') = rh'((-1)^{n-1}\xi') = r\xi$ and we are done.

Having a basis for $\mathrm{Sk}_n(F)$, we are now in a position to compute $|f|$. We have $f(x_i) = \Sigma\, a_{ij}x_j$ where $\{x_1, \ldots, x_n\}$ is a basis of F. Let ξ be the element of $\mathrm{Sk}_n(F)$ such that $\xi(x_1, \ldots, x_n) = 1$. From the foregoing discussion, we need only compute $\mathrm{Sk}_n(f)(\xi)(x_1, \ldots, x_n)$ because this element of R is precisely $|f|$. By definition, $\mathrm{Sk}_n(f)(\xi)(x_1, \ldots, x_n) = \xi(f(x_1), \ldots, f(x_n)) = \xi(\Sigma\, a_{1j_1}x_{j_1}, \ldots, \Sigma\, a_{nj_n}x_{j_n}) = \Sigma\, a_{1j_1}a_{2j_2} \cdots a_{nj_n}\xi(x_{j_1}, \ldots, x_{j_n})$. Observe that this sum ranges over all n-tuples of indices $(j_1, \ldots, j_n)$ with each index ranging over $\{1, \ldots, n\}$. However, if any two of the indices j_k and j_l are equal, we have $\xi(x_{j_1}, \ldots, x_{j_n}) = 0$. Thus, we may restrict ourselves to just those n-tuples $(j_1, \ldots, j_n)$ in which the j_k are all distinct. But each such n-tuple determines a unique permutation π of the set $\{1, \ldots, n\}$, and we may write $(\pi(1), \ldots, \pi(n))$ for the n-tuple $(j_1, \ldots, j_n)$. With this notation, we write $\Sigma\, a_{1j_1}a_{2j_2} \cdots a_{nj_n}\xi(x_{j_1}, \ldots, x_{j_n}) = \Sigma\, a_{1\pi(1)}a_{2\pi(2)} \cdots a_{n\pi(n)}\xi(x_{\pi(1)}, \ldots, x_{\pi(n)})$ where the sum ranges over all permutations π of the set $\{1, \ldots, n\}$. It is a triviality to show (using skew-symmetry of ξ) that $\xi(x_{\pi(1)}, \ldots, x_{\pi(n)}) = \mathrm{sig}(\pi)\xi(x_1, \ldots, x_n) = \mathrm{sig}(\pi)$ [because $\xi(x_1, \ldots, x_n) = 1$], and therefore we have $\mathrm{Sk}_n(f)(\xi)(x_1, \ldots, x_n) = \Sigma_\pi\, \mathrm{sig}(\pi)a_{1\pi(1)}a_{2\pi(2)} \cdots a_{n\pi(n)}$.

(d) Left as an exercise.

Example 2.6 Let F be free of rank two, with basis $\{x_1, x_2\}$ and let $f: F \to F$ be a morphism given by $f(x_1) = a_{11}x_1 + a_{12}x_2$ and $f(x_2) = a_{21}x_1 + a_{22}x_2$. By the above, we have $|f| = a_{11}a_{22} - a_{12}a_{21}$ and we see that we have our usual definition of determinant.

Example 2.7 Let F be a free module of rank n and $f: F \to F$ an endomorphism. Then $F^* = (F, R)$ is also free of rank n, and we have the endomorphism $f^*: F^* \to F^*$, where $f^* = (f, R)$. We can show that $|f| = |f^*|$. To do this, choose a

basis $\{x_1, \ldots, x_n\}$ for F. Then $\{\xi_1, \ldots, \xi_n\}$ is a basis for F^* where $\xi_i(x_j) = \delta_{ij}$ (the Kronecker delta). Simple linear algebra tells us that if $f(x_i) = \Sigma\, a_{ij}x_j$ and $f^*(\xi_j) = \Sigma\, b_{ij}\xi_i$, then $b_{ij} = a_{ji}$. Property (c) above tells us that $|f| = \Sigma_\pi \pm a_{1\pi(1)} \cdots a_{n\pi(n)}$ and that $|f^*| = \Sigma_\pi \pm a_{\pi(1)1} \cdots a_{\pi(n)n}$. However, because $\operatorname{sig}(\pi) = \operatorname{sig}(\pi^{-1})$ and because $a_{\pi(i)} = a_{\pi^{-1}\pi(i)\pi(i)}$, we have

$$\begin{aligned}\operatorname{sig}(\pi)a_{1\pi(1)} \cdots a_{n\pi(n)} &= \operatorname{sig}(\pi)a_{\pi^{-1}\pi(1)\pi(1)} \cdots a_{\pi^{-1}\pi(n)\pi(n)} \\ &= \operatorname{sig}(\pi^{-1})a_{\pi^{-1}(1)1} \cdots a_{\pi^{-1}(n)n}\end{aligned}$$

Thus $|f| = |f^*|$.

Example 2.8 If F is a free R-module of finite rank and $f: F \to F$ is an automorphism, then $|f|$ is a unit. We will see in a little while that $|f|$ is a unit if and only if f is an automorphism.

3. MATRICES

Having defined the determinant of an endomorphism of a free module, we are now ready to define the determinant of a square matrix over a commutative ring R. To make sure, though, that we are all talking about the same thing, let us denote by $[n]$ the set of integers $\{1, \ldots, n\}$ and make the following:

Definition
An **m-by-n matrix** (written $m \times n$ matrix) over a commutative ring R is a map from the set $[m] \times [n]$ to R. If $m = n$, the matrix is called a **square matrix**.

If $A: [m] \times [n] \to R$ is an $m \times n$ matrix, we usually write the element $A(i, j)$ as a_{ij} and denote the matrix by $A = (a_{ij})$. This leads to the usual way of illustrating the matrix as a rectangular array:

$$\begin{pmatrix} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \vdots & & & \\ a_{m1}, & a_{m2}, & \ldots, & a_{mn} \end{pmatrix}$$

Definition
If $A: [m] \times [n] \to R$ is an $m \times n$ matrix, its **tranpose** is a matrix ${}^tA: [n] \times [m] \to R$ defined by ${}^tA(i, j) = A(j, i)$. If $A_1, A_2: [m] \times [n] \to R$ are matrices, we denote by $A_1 + A_2: [m] \times [n] \to R$ the matrix defined by $(A_1 + A_2)(i, j) = A_1(i, j) + A_2(i, j)$, and call it the **sum of A_1 and A_2**.

If r is an element of R, we define $rA_1: [m] \times [n] \to R$ by $(rA_1)(i, j) = r(A_1(i, j))$.

Finally, if $A_1: [m] \times [n] \to R$ and $A_2: [n] \times [p] \to R$, we define $A_1A_2: [m] \times [p] \to R$ by $A_1A_2(i, j) = \Sigma_{k=1}^{n} A_1(i, k)A_2(k, j)$. This matrix is called the **product of A_1 and A_2**.

The reader has seen all of these definitions before when R is a field. There is essentially no difference between the formal properties of matrices defined over a commutative ring and those defined over a field. In particular, if m and n are fixed,

the set of all $m \times n$ matrices over R, denoted by $\mathcal{M}(m, n)$ is an R-module under the addition and scalar multiplication defined above. Multiplication of matrices is associative, and ${}^t(A_1A_2) = {}^tA_2{}^tA_1$.

If F and G are free modules of ranks m and n, respectively, with respective bases $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$, then the map $h:(F, G) \to \mathcal{M}(m, n)$ defined by $h(f)(i, j) = a_{ij}$ where $f(x_i) = \Sigma_j a_{ij} y_j$, is an isomorphism of R-modules. Letting $\{\xi_1, \ldots, \xi_m\}$ and $\{\eta_1, \ldots, \eta_n\}$ denote the bases of F^* and G^*, respectively, with $\xi_i(x_j) = \delta_{ij}$ and $\eta_k(y_l) = \delta_{kl}$, we have the isomorphism $h':(G^*, F^*) \to \mathcal{M}(n, m)$, and for all $f \in (F, G)$ we have ${}^t[h(f)] = h'(f^*)$. Finally, if H is a free module of rank p with basis $\{z_1, \ldots, z_p\}$, the isomorphisms $h'':(G, H) \to \mathcal{M}(n, p)$ and $h''':(F, H) \to \mathcal{M}(m, p)$ have the property that $h'''(gf) = h(f)h''(g)$ for $f: F \to G$ and $g: G \to H$.

Now suppose that A is a square matrix of order n, that is, A is an $n \times n$ matrix. Let F be a free R-module with basis $\{x_1, \ldots, x_n\}$ and define $f: F \to F$ by $f(x_i) = \Sigma_{j=1}^n A(i, j)x_j$, that is, $f = h^{-1}(A)$ where $h:(F, F) \to \mathcal{M}(n, n)$ is the isomorphism described above. Because f has a determinant, we are tempted to define the determinant of A as the determinant of f. However, if G were another free R-module of rank n, with basis $\{y_1, \ldots, y_n\}$, and if we defined $g: G \to G$ by $g(y_i) = \Sigma_{j=1}^n A(i, j)y_j$, could we be sure that $|f| = |g|$?

To see that is is the case, we use Basic Properties 2.5, for that says that $|f| = \Sigma_\pi \operatorname{sig}(\pi)A(1, \pi(1))A(2, \pi(2)) \cdots A(n, \pi(n)) = |g|$.

Definition

Let A be a square matrix over R of order n, let F be a free R-module of rank n with basis $\{x_1, \ldots, x_n\}$, and let $f: F \to F$ be the endomorphism defined by $f(x_i) = \Sigma_j A(i, j)x_j$. The **determinant of A**, denoted by $|A|$ or $\det A$, is defined to be the determinant of f. If A is written (a_{ij}), then $|A|$ is usually written $|a_{ij}|$.

Basic Properties 3.1

(a) If A_1 and A_2 are square matrices of order n, then $|A_1A_2| = |A_1||A_2|$.
(b) If A is a square matrix, then $|{}^tA| = |A|$.
(c) If A is a square matrix of order n, then $|A| = \Sigma_\pi \operatorname{sig}(\pi)A(1, \pi(1))A(2, \pi(2)), \ldots, A(n, \pi(n))$.
(d) If A is a square matrix with two rows equal or two columns equal, then $|A| = 0$.

The proofs of these properties are left to the reader.

Although (c) above gives us a way of concretely computing the determinant of a matrix, it possibly is different from what the reader has generally taken to be the definition of the determinant. For example, if

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

one is usually told that

$$|A| = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

and one assumes that determinants of square matrices of order two are known. This is, of course, a very useful inductive way of defining determinants that can be established immediately from (c) after we set things up properly.

First, let us say that a map $f:[n_1]\to[n_2]$ is **order-preserving** if whenever $i \le j$ in $[n_1]$, we have $f(i)\le f(j)$. If $n_1 \le n_2$, we can have order-preserving maps which are injective; in fact, the number of order-preserving injective maps of $[n_1]$ into $[n_2]$ is $\binom{n_2}{n_1}$ where $\binom{n_2}{n_1}$ is the binomial coefficient. Although this is clear, we may actually get more detailed information from the following.

Example 3.2 If n is a positive integer, there are n distinct order-preserving injective maps of $[n-1]$ to $[n]$. They are the maps $\epsilon_i:[n-1]\to[n]$ defined by $\epsilon_i(k)=k$ if $k<i$, and $\epsilon_i(k)=k+1$ if $k\ge i$. The index i runs from 1 to n. If $n=1$, then $[n-1]=\emptyset$, and of course there is only one map from the empty set to [1]. Let us agree to use the same symbol ϵ_i to denote one of the above types of maps of $[n-1]$ into $[n]$ whatever the integer n. Thus, $\epsilon_2:[2]\to[3]$ is not the same map as $\epsilon_2:[3]\to[4]$, but the meaning is still clear.

Now let $f:[p]\to[n]$ be any order-preserving injective map of $[p]$ into $[n]$. Then we have $1\le f(1)<f(2)<\cdots<f(p)\le n$. Letting I be the complement of the image of f in $[n]$, we may arrange the elements of I so that $I=\{i_1, i_2, \ldots, i_{n-p}\}$ with $1\le i_1<i_2<\cdots<i_{n-p}<n$. Because $i_k \le p+k$ for $k=1,\ldots,n-p$, we have the maps $\epsilon_{i_k}:[p+k-1]\to[p+k]$ for $k=1,\ldots,n-p$. The reader may now easily convince himself that $f=\epsilon_{i_{n-p}}\cdots\epsilon_{i_1}$.

Getting back to our matrices, we have the following.

Definition
Let $A:[m]\times[n]\to R$ be an $m\times n$ matrix over R. A matrix $A':[m']\times[n']\to R$ is called a **submatrix** of A if A' is the composition $[m']\times[n'] \xrightarrow{f\times g} [m]\times[n] \xrightarrow{A} R$ where $f:[m']\to[m]$ and $g:[n']\to[n]$ are order-preserving injective maps. If A is a square matrix of order n, the submatrix $A_{ij}=A\cdot(\epsilon_i\times\epsilon_j):[n-1]\times[n-1]\to R$ is called the **complement** of $A(i,j)$, where $e_i, e_j:[n-1]\to[n]$ are the maps introduced in Example 3.2 above. The element $(-1)^{i+j}|A_{ij}|$ is called the **cofactor** of the element $A(i,j)$.

Notice that $|A_{ij}|$ is the determinant of a square matrix of order $n-1$ and that the cofactor is not just this determinant but carries a sign along with it.

Example 3.4 Let

$$A=\begin{pmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{pmatrix}$$

The following are examples of submatrices of A:

(i) $\begin{pmatrix} a_{11} & a_{12} & a_{13}\\ a_{31} & a_{32} & a_{33}\end{pmatrix}$ (ii) $\begin{pmatrix} a_{12}\\ a_{32}\end{pmatrix}$ (iii) $\begin{pmatrix} a_{22} & a_{23}\\ a_{32} & a_{33}\end{pmatrix}$

It might be instructive to note that $\begin{pmatrix} a_{31} & a_{32} & a_{33}\\ a_{11} & a_{12} & a_{13}\end{pmatrix}$ is *not* a submatrix of A.

Proposition 3.5
Let A be a square matrix of order n, and let i be some fixed integer between 1 and n. Then

$$\sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{kj}| = \delta_{ik} |A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ji} |A_{jk}|$$

for $k = 1, \ldots, n$, where δ_{ik} is the Kronecker delta.

PROOF: We know that $|A| = \Sigma_\pi \operatorname{sig}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)}$. This sum is easily seen to be $\Sigma_{j=1}^{n} a_{ij} \Sigma_{\pi'} \operatorname{sig}(\pi') a_{1\pi'(1)} \cdots \hat{a}_{ij} \cdots a_{n\pi'(n)}$ where π' ranges over all permutations such that $\pi'(i) = j$. Thus, to show that $|A| = \Sigma_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$, it remains only to show that $\Sigma_{\pi'} \operatorname{sig}(\pi') a_{1\pi'(1)} \cdots \hat{a}_{ij} \cdots a_{n\pi'(n)} = (-1)^{i+j} |A_{ij}|$ where π' is as above.

Because A_{ij} is an $(n-1) \times (n-1)$ matrix, we have $|A_{ij}| = \Sigma_\sigma \operatorname{sig}(\sigma) A_{ij}(1, \sigma(1)) \cdots A_{ij}(n-1, \sigma(n-1))$ where σ ranges over all permutations of $\{1, \ldots, n-1\}$. Now σ may also be regarded as a permutation of $\{1, \ldots, n\}$ with $\sigma(n)$ defined to be equal to n, and we may therefore consider, for each such σ, the composition of permutations $(j, j+1, \ldots, n)\sigma(n, n-1, \ldots, i)$ where the notation $(i_1, \ldots, i_k)$ means the permutation taking i_j into i_{j+1} for $j = 1, \ldots, k-1$, taking i_k into i_1, and leaving everything else fixed. The reader may now easily check that if we denote the composition $(j, j+1, \ldots, n)\sigma(n, n-1, \ldots, i)$ by π', then $\pi'(i) = j$ and the term $A_{ij}(1, \sigma(1)) \cdots A_{ij}(n-1, \sigma(n-1))$ equals the term $a_{1\pi'(1)} \cdots \hat{a}_{ij} \cdots a_{n\pi'(n)}$. It is also easy to see that $\operatorname{sig}(\pi') = (-1)^{2n-(i+j)} \operatorname{sig}(\sigma) = (-1)^{i+1} \operatorname{sig}(\sigma)$, so that the assertion $\Sigma_\pi \operatorname{sig}(\pi') a_{1\pi'(1)} \cdots \hat{a}_{ij} \cdots a_{n\pi'(n)} = (-1)^{i+j} |A_{ij}|$ follows immediately. We therefore have established the fact that

$$|A| = \sum_{j=1}^{n} (-1)^{i+1} a_{ij} |A_{ij}| = \delta_{il} |A|$$

Using the above equality, we can now show that

$$\sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{kj}| = \delta_{ik} |A| = 0 \qquad \text{when} \qquad i \neq k$$

For consider the matrix A' where $A'(i', j') = A(i', j')$ if $i' \neq k$, but $A'(k, j') = A(i, j')$. Then A' is a matrix whose ith row is equal to the kth row, so that (by Basic Properties 3.1) $|A'| = 0$. On the other hand, $|A'_{ij}|$ is clearly equal to $-|A_{kj}|$ (why?), so we have

$$0 = -|A'| = \sum_{j=1}^{n} (-1)^{i+j+1} a_{ij} |A'_{ij}| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{kj}|$$

This finishes the proof that

$$\sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{kj}| = \delta_{ik} |A|$$

The reader may prove for himself that

$$\delta_{ik} |A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ji} |A_{jk}|$$

Corollary 3.6
Let A be a square matrix of order n, and let $\tilde{A}$ be the square matrix of order n defined by $\tilde{A}(i, j) = (-1)^{i+j} |A_{ij}|$. Then $A\tilde{A} = \tilde{A}A = |A| I$, and $|A\tilde{A}| = |A|^n$. Further-

more, if $|A|$ is a unit, then $((1/|A|)\tilde{A})A = I = A((1/|A|)\tilde{A})$, so that A has an inverse matrix $A^{-1} = (1/|A|)\tilde{A}$.

Proposition 3.7
Let F be a free module of rank n, and $f: F \to F$ an endomorphism of F. Then $|f|$ is contained in ann(Coker f) and in ann(Ker f).

PROOF: Let $\{x_1, \ldots, x_n\}$ be a basis for F and let $A = (a_{ij})$ be the matrix corresponding to f with respect to this basis. Letting $\tilde{A}$ be the matrix introduced in Proposition 3.6, we denote by $\tilde{f}$ the endomorphism of F corresponding to $\tilde{A}$. Because $A\tilde{A} = |A|I = \tilde{A}A$, we see that the compositions $f\tilde{f}: F \to F$ and $\tilde{f}f: F \to F$ are both simply the endomorphism of F obtained by multiplying each element of F by $|A|$ (or by $|f|$, since $|A| = |f|$).

Suppose, now, that $x \in \text{Ker } f$. Then $(\tilde{f}f)(x) = \tilde{f}(f(x)) = \tilde{f}(0) = 0$. However, $(\tilde{f}f)(x) = |f|x$ so that we have shown $|f|x = 0$ if $x \in \text{Ker } f$. Hence, $|f| \in \text{ann}(\text{Ker } f)$.

Now let x be any element of F. Then $|f|x = (f\tilde{f})(x) = f(\tilde{f}(x))$ so that for each $x \in F$ we see that there is a $y \in F$ [namely, $\tilde{f}(x)$] with the property that $f(y) = |f|x$. Using this fact, we can show that $|f|$ is in ann(Coker f). For, if $z \in \text{Coker } f$, let $z = k(x)$ where $x \in F$ and $k: F \to \text{Coker } f$ is the canonical morphism. Then $|f|z = |f|k(x) = k(|f|x) = k(f(y)) = (kf)(y) = 0$ because $kf = 0$, and thus $|f|$ annihilates every element of Coker f.

Corollary 3.8
If F is a free module rank and $f: F \to F$ is an endomorphism, then f is an automorphism if and only if $|f|$ is a unit in the ring R.

PROOF: In Example 2.8 we saw that if f is an automorphism, then $|f|$ is a unit. Conversely, if $|f|$ is a unit, we must have $\text{Ker } f = 0 = \text{Coker } f$ because $|f| \in \text{ann}(\text{Coker } f) \cap \text{ann}(\text{Ker } f)$.

4. FURTHER APPLICATIONS OF THE FUNDAMENTAL THEOREM

In this section, K will denote a field and V will be a finite-dimensional vector space over K. We denote by $K[X]$ the polynomial ring in one indeterminate over K, and we recall that $K[X]$ is a PID; in fact, it is a Euclidean ring.

If $T: V \to V$ is a linear transformation (or endomorphism) of the vector space V, we may make V a module over $K[X]$ by defining $Xv = T(v)$ for every $v \in V$. Thus, if $P(X) = \Sigma\, a_i X^i$ is an element of $K[X]$ and $v \in V$, we have $P(X) \cdot v = \Sigma\, a_i T^i(v)$. Because V is a finite-dimensional vector space, V is obviously finitely generated as a $K[X]$-module. In fact, the morphism $\epsilon: K[X] \underset{K}{\otimes} V \to V$ defined by $\epsilon(P(X) \otimes v) = P(X) \cdot v$ is clearly an epimorphism of $K[X]$-modules, and $K[X] \underset{K}{\otimes} V$ is a free $K[X]$-module whose rank is equal to the dimension of the K-vector space V.

Not only is V a finitely generated $K[X]$-module, it is actually a torsion module over $K[X]$. To see this, we must show that if $v \in V$, then there is some element $P(X)$ of $K[X]$ such that $P(X)v = 0$. Because V is finite-dimensional over K, say

dim $V = n$, we know that the $n+1$ elements $v, T(v), T^2(v), \ldots, T^n(v)$ are linearly dependent over K. Thus, we may find $n+1$ elements $a_0, a_1, \ldots, a_n$ of K, not all of which are zero, such that $\Sigma_{i=0}^{n} a_i T^i(v) = 0$. Letting $P(X) = \Sigma_{i=0}^{n} a_i X^i$, we have $P(X)v = 0$, with $P(X) \neq 0$. Because V is a torsion module and $\epsilon : K[X] \underset{K}{\otimes} V \to V$ is an epimorphism of the free $K[X]$-module $K[X] \underset{K}{\otimes} V$ onto V, it follows that Ker ϵ is a free $K[X]$-module of rank n, where $n = \dim_K V$. (Do not forget that $K[X]$ is a PID.) We shall now begin to study Ker ϵ.

First, it is clear that if $P(X) \in K[X]$ and $v \in V$, then $XP(X) \otimes v - P(X) \otimes Xv \in$ Ker ϵ. Thus, any finite sum of elements of this form is again in Ker ϵ. If $Q(X) \in K[X]$, then $Q(X)[XP(X) \otimes v - P(X) \otimes Xv] = XQ(X)P(X) \otimes v - Q(X)P(X) \otimes Xv$ so that the set of all finite sums of elements of the form $XP(X) \otimes v - P(X) \otimes Xv$ is a submodule of $K[X] \underset{K}{\otimes} V$ contained in Ker ϵ. Let us call this submodule L. What we propose to show is that L is precisely Ker ϵ. Once we have succeeded in proving this, we will have shown that the morphism $\gamma : K[X] \underset{K}{\otimes} V \to K[X] \underset{K}{\otimes} V$ defined by $\gamma(P(X) \otimes v) = XP(X) \otimes v - P(X) \otimes Xv$ has, as its image, the kernel of ϵ. Thus, the sequence of $K[X]$-modules

$$K[X] \underset{K}{\otimes} V \xrightarrow{\gamma} K[X] \underset{K}{\otimes} V \xrightarrow{\epsilon} V \longrightarrow 0$$

will have been shown to be exact. In addition, using the fact that V is a torsion module together with our observation at the very end of Section 2, we will know that γ is a monomorphism, and hence the sequence

$$0 \longrightarrow K[X] \underset{K}{\otimes} V \xrightarrow{\gamma} K[X] \underset{K}{\otimes} V \xrightarrow{\epsilon} V \longrightarrow 0 \tag{1}$$

will have been shown to be exact.

To see that $L =$ Ker ϵ, consider first elements in $K[X] \underset{K}{\otimes} V$ of the form $aX^m \otimes v$ where $a \in K$ and $v \in V$. If $m = 0$, then $X^m = 1$ and $a \otimes v = 1 \otimes av$ so that $a \otimes v = 1 \otimes e(a \otimes v)$. If $m = 1$, we have $aX \otimes v = aX \otimes v - a \otimes Xv + a \otimes Xv = aX \otimes v - a \otimes Xv + 1 \otimes aXv$. Letting $l = aX \otimes v - a \otimes Xv$, we have $l \in L$ and $aX \otimes v = l + 1 \otimes e(aX \otimes v)$. Suppose, now, that we have shown that $aX^m \otimes v = l + 1 \otimes e(aX^m \otimes v)$ for $m = 0, \ldots, k$ and all $v \in V$, where l is some element in L depending, of course, on a, m, and v. Then $aX^{k+1} \otimes v = X \cdot aX^k \otimes v - aX^k \otimes Xv + aX^k \otimes Xv = l_1 + aX^k \otimes Xv$ where $l_1 = XaX^k \otimes v - aX^k \otimes Xv \in L$. By our induction assumption, we have $aX^k \otimes Xv = l_2 + 1 \otimes e(aX^k \otimes Xv)$ with $l_2 \in L$. Note that $e(aX^k \otimes Xv) = aX^{k+1}v = e(aX^{k+1} \otimes v)$ so that, setting $l = l_1 + l_2$, we have $aX^{k+1} \otimes v = l + 1 \otimes e(aX^{k+1} \otimes v)$. We have therefore shown that every element of the form $aX^m \otimes v$ may be written as $l + 1 \otimes e(aX^m \otimes v)$ with $l \in L$. It follows immediately, then, that every element y in $K[X] \underset{K}{\otimes} V$ may be written $y = l + 1 \otimes e(y)$ with $l \in L$.

Now suppose that y is in Ker ϵ. Writing $y = l + 1 \otimes e(y)$, with $l \in L$, we see that $y = l$ because $1 \otimes e(y) = 1 \otimes 0 = 0$. Hence, Ker $\epsilon \subset L \subset$ Ker ϵ, so that $L =$ Ker ϵ. This proves that the sequence (1) is exact.

To tie up some of the above material with the reader's past experience, let us choose a convenient basis for $K[X] \underset{K}{\otimes} V$ and see what the corresponding matrix

of γ looks like. If $\{v_1, \ldots, v_n\}$ is a basis for V over K, we know that $\{1 \otimes v_1, \ldots, 1 \otimes v_n\}$ is a basis for $K[X] \underset{K}{\otimes} V$ over $K[X]$. Because $\gamma(1 \otimes v_i) = X \otimes v_i - 1 \otimes Xv_i = X(1 \otimes v_i) - 1 \otimes T(v_i)$, we have

$$\gamma(1 \otimes v_i) = X(1 \otimes v_i) - \sum_{j=1}^{n} 1 \otimes a_{ij}v_j = \sum_{j=1}^{n} (\delta_{ij}X - a_{ij})(1 \otimes v_i)$$

where $T(v_i) = \Sigma_{j=1}^{n} a_{ij}v_j$, with $a_{ij} \in K$. Thus, the matrix of γ with respect to this basis is

$$\begin{pmatrix} X - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & X - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & X - a_{nn} \end{pmatrix} \qquad (*)$$

where

$$A = \begin{pmatrix} a_{11} \cdots a_{1n} \\ \vdots \quad \vdots \\ a_{n1} \cdots a_{nn} \end{pmatrix}$$

is the matrix of T with respect to the basis $\{v_1, \ldots, v_n\}$.

Definition
The matrix (*) is called the **characteristic matrix** of the matrix A. The element $|\gamma|$ of $K[X]$ is called the **characteristic polynomial** of the transformation T and the determinant of the matrix (*) is called the **characteristic polynomial of the matrix** A.

Because V is the cokernel of γ, we know that $|\gamma|$ is in the annihilator of V (see Proposition 3.7). If we write

$$|\gamma| = a_0 + a_1X + \cdots + a_nX^n$$

we see, first of all, that $a_n = 1$ and that the transformation

$$a_0I + a_1T + a_2T^2 + \cdots + T^n : V \to V$$

is the zero morphism of V into itself because $|\gamma|v = (\Sigma\, a_iX^i)v = \Sigma\, a_i(T^i(v)) = 0$ where $I = \mathrm{id}_V$. Therefore, we obtain the Cayley-Hamilton theorem.

Theorem 4.1
If K is a field and $A = (a_{ij})$ is a square matrix of order n, then:

(a) The characteristic polynomial of A is a monic polynomial of degree n.
(b) A is a root of its characteristic polynomial.

Example 4.2 Consider the matrix

$$A = \begin{pmatrix} 2 & 1 & 0 \\ -1 & 3 & 2 \\ 4 & 0 & -6 \end{pmatrix}$$

Then its characteristic polynomial is $X^3+X^2-23X+34$;

$$A^2=\begin{pmatrix} 3 & 5 & 2 \\ 3 & 8 & -6 \\ -16 & 4 & 36 \end{pmatrix}$$

$$A^3=\begin{pmatrix} 9 & 18 & -2 \\ -26 & 27 & 52 \\ 108 & -4 & -208 \end{pmatrix}$$

$$A^3+A^2-23A+34I=\begin{pmatrix} 9+3-46+34 & 18+5-23+0 & -2+2+0+0 \\ -26+3+23+0 & 27+8-69+34 & 52-6-46+0 \\ 108-16-92+0 & -4+4+0+0 & -208+36+138+34 \end{pmatrix}$$

and each of the sums is zero as we hoped and, hopefully, expected.

Returning to our general discussion, we make use of the fact that $K[X]$ is a PID and exploit the results of Section 1. By Corollary 1.4 we know that there are bases $\{z_1,\dots,z_n\}$ and $\{y_1,\dots,y_n\}$ of the free $K[X]$-module $K[X]\otimes_K V$ such that $\gamma(y_i)=P_i(X)z_i$ where $P_i(X)$ are elements of $K[X]$ and $(0)\neq(P_1(X))\subset\cdots\subset(P_n(X))$. Because we may choose any unit multiple of $P_i(X)$ in place of $P_i(X)$, we may assume that each $P_i(X)$ is a monic polynomial. The reader should not overlook the fact that some $(P_i(X))$ may equal (R) and thus some $P_i(X)=1$. The polynomials $P_i(X)$ are the invariant factors of the $K[X]$-module V, and $V\approx K[X]/(P_1(X))\amalg K[X]/(P_2(X))\amalg\cdots\amalg K[X]/(P_n(X))$ as a $K[X]$-module. The ideal $(P_1(X))$ is the annihilator of V as a $K[X]$-module.

Definition

The monic polynomial $P_1(X)$ in the above decomposition of V as a $K[X]$-module is called the **minimal polynomial of the transformation** T. The nonunit polynomials $P_i(X)$, $i=1,\dots,r$ are called the **invariant factors of** T.

In a little while we shall see what the modules $K[X]/(P_i(X))$ look like as vector spaces over K. First, however, let us look at $|\gamma|$, the characteristic polynomial of T, and see how it may be expressed in terms of the invariant factors of T.

Using our above basis $\{z_1,\dots,z_n\}$ and $\{y_1,\dots,y_n\}$ for $K[X]\otimes_K V$, we may write $y_i=\Sigma\, b_{ij}z_j$ and $z_j=\Sigma\, c_{jk}y_k$ with b_{ij} and c_{ij} in $K[X]$. Therefore, $z_i=\Sigma_k\, c_{ik}y_k=\Sigma_{k,j}\, c_{ik}b_{kj}z_j$ so that we have $CB=I_n$ where C is the matrix (c_{ij}), B is the matrix (b_{ij}), and I_n is the $n\times n$ matrix defined by $I_n(i,j)=\delta_{ij}$. I_n is called the identity matrix. Thus, we have $|CB|=1$ or $|C||B|=1$ so that both $|C|$ and $|B|$ are units of $K[X]$. In particular, because the only units of $K[X]$ are the nonzero polynomials of degree zero (that is, constants), we have that $|B|$ and $|C|$ are constants.

Letting $P_i(X)=P_i$, we have $\gamma(y_i)=P_iz_i=\Sigma\, P_ic_{ik}y_k$ so that, with respect to the basis $\{y_1,\dots,y_n\}$, γ has the matrix (d_{ik}) where $d_{ik}=P_ic_{ik}$. The matrix (d_{ik}) is the

product of the matrices P and C where

$$P = \begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & P_n \end{pmatrix}$$

and C is as above. Thus, $|\gamma| = |d_{ik}| = |PC| = |P||C|$. The element $|\gamma|$ is the characteristic polynomial of T and is monic. The element $|P|$ is clearly equal to $P_1P_2 \cdots P_n$ and, because each P_i is monic, this product is monic. Finally, $|C|$ is a unit, hence a constant. Thus, $|C| = 1$ because $|P||C| = |\gamma| =$ a monic polynomial. This gives us the following.

Proposition 4.3
The characteristic polynomial of a transformation T is the product of the invariant factors of T.

If L is any field containing K and λ is any element of L, then for any polynomial $Q(X) \in K[X]$ we obtain an element $Q(\lambda) \in L$. For if $Q(X) = a_0 + a_1X + \cdots + a_nX^n$ with $a_i \in K$, then $a_i\lambda^i \in L$ for each i and $Q(\lambda) = a_0 + a_i\lambda + \cdots + a_n\lambda^n$. If $Q(X) = Q_1(X) \cdots Q_s(X)$ where the $Q_i(X)$ are elements of $K[X]$, and if $\lambda \in L$, then $Q(\lambda) = 0$ if and only if $Q_i(\lambda) = 0$ for some $i = 1, \ldots, s$. If, for each $i = 1, \ldots, s-1$, we have $Q_{i+1}(X)$ divides $Q_i(X)$, that is, $Q_i(X) = T_i(X)Q_{i+1}(X)$, then it is clear that if $Q_{i+1}(\lambda) = 0$, we have $Q_i(\lambda) = 0$ also. Thus, in this case, if $\lambda \in L$, then $Q(\lambda) = 0$ if and only if $Q_1(\lambda) = 0$.

Corollary 4.4
Let $T: V \to V$ be a linear transformation of the finite-dimensional vector space V over the field K, and let L be any field containing K. If λ is an element of L, then λ is a root of the characteristic polynomial of T if and only if λ is a root of the minimal polynomial of T.

5. CANONICAL FORMS

Having seen that our vector space V is, as a $K[X]$-module, isomorphic to the sum of modules of the form $K[X]/(P(X))$, we are naturally interested in seeing what such $K[X]$-modules look like as vector spaces over K. For simplicity, but with no loss of generality, we shall assume that $P(X)$ is monic so that we have $P(X) = X^d + a_1X^{d-1} + \cdots + a_d$ with $a_i \in K$. We claim that the set $\{\bar{1}, \bar{X}, \bar{X}^1, \ldots, \bar{X}^{d-1}\}$ is a basis for $K[X]/(P(X))$ over K, where $\bar{1}$ and $\bar{X}^i$ are the cosets of, respectively, 1 and X^i in $K[X]/(P(X))$. That this set generates $K[X]/(P(X))$ over K is clear from the fact that if $F(X)$ is any element of $K[X]$, then $F(X) = Q(X)P(X) + R(X)$ where either $R(X)$ is zero or the degree of $R(X)$ is less than d.

Because every polynomial in $K[X]$ of degree less than d is a linear combination, with coefficients in K, of the elements $1, X, \ldots, X^{d-1}$, it follows that $\{\bar{1}, \bar{X}, \ldots, \bar{X}^{d-1}\}$ generates $K[X]/(P(X))$ over K.

Now suppose we have a linear combination $\sum_{i=0}^{d-1} c_i \bar{X}^i$ which is zero. It follows that $\sum_{i=0}^{d-1} c_i X^i$ is divisible by $P(X)$ in $K[X]$. However, $P(X)$ is of degree d and the degree of a product of polynomials in $K[X]$ is the sum of the degrees of the factors. Hence, if $\sum_{i=0}^{d-1} c_i X^i = P(X)Q(X)$, it must be the case that $Q(X) = 0$. Therefore, $\sum_{i=0}^{d-1} c_i X^i = 0$, and thus $c_i = 0$ for $i = 0, \ldots, d-1$. Thus, we have shown that the set $\{\bar{1}, \bar{X}, \ldots, \bar{X}^{d-1}\}$ is linearly independent over K and that this set is a basis for $K[X]/(P(X))$ as a vector space over K.

Because multiplication by X on $K[X]/(P(X))$ is a linear transformation on this vector space, we may see what the matrix associated with this morphism is with respect to the basis just described. Writing $\bar{X}^0$ for $\bar{1}$, we see that $X \cdot \bar{X}^i = \bar{X}^{i+1}$ for $i = 0, \ldots, d-2$. However, $X \cdot \bar{X}^{d-1} = \bar{X}^d = -\sum_{i=0}^{d-1} a_{d-i} \bar{X}^i$ [because $P(X) = X^d + a_1 X^{d-1} + \cdots + a_d$]. Thus, we get the square matrix of order d:

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \\ -a_d & -a_{d-1} & -a_{d-2} & \cdots & -a_1 \end{pmatrix}$$

Returning now to our vector space V, we have $V \approx K[X]/(P_1(X)) \amalg \cdots \amalg K[X]/(P_r(X))$ where $P_1(X), \ldots, P_r(X)$ are the invariant factors of T. Because V is a sum of these modules as a $K[X]$-module, V is certainly a sum of these modules as a vector space. Do not forget, too, that the operation T on V is multiplication by X. Thus, a basis for V may be found which is the union of bases for each of the vector spaces $K[X]/(P_i(X))$, and the matrix for T with respect to this basis will be easy to compute.

To see what we mean by this, let us suppose that we have a vector space V, a transformation $T: V \to V$, and subspaces $V_1, \ldots, V_r$ such that $T(V_i) \subset V_i$ for $i = 1, \ldots, r$. Suppose, moreover, that $V = V_1 \amalg \cdots \amalg V_r$. If we let $\{v_{i1}, \ldots, v_{id_i}\}$ be a basis for $V_i (i = 1, \ldots, r)$, then $\{v_{11}, \ldots, v_{1d_1}, v_{21}, \ldots, v_{rd_r}\}$ is clearly a basis for V. If

$$T(v_{ij}) = \sum_{k=1}^{d_i} a_{jk}^i v_{ik} \qquad \text{for} \qquad 1 \le i \le r, \qquad 1 \le j \le d_i$$

then the matrix of $T|V_i$ (the restriction of T to V_i) is

$$A_i = (a_{jk}^i)$$

and the matrix of T with respect to the basis $\{v_{ij}\}$ is

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & & \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & A_r \end{pmatrix}$$

or, in more detail,

$$\begin{pmatrix} \begin{pmatrix} a^1_{11} & \cdots & a^1_{1d_1} \\ a^1_{21} & \cdots & a^1_{2d_1} \\ \vdots & & \\ a^1_{d_11} & \cdots & a^1_{d_1d_1} \end{pmatrix} & \begin{matrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ & & \\ 0 & \cdots & 0 \end{matrix} & \cdots & \begin{matrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ & & \\ 0 & \cdots & 0 \end{matrix} \\ \begin{matrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{matrix} & \begin{pmatrix} a^2_{11} & \cdots & a^2_{1d_2} \\ a^2_{21} & \cdots & a^2_{2d_2} \\ & & \\ a^2_{d_21} & \cdots & a^2_{d_2d_2} \end{pmatrix} & \cdots & \begin{matrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 \\ & & & & \\ 0 & \cdots & 0 & \cdots & 0 \\ & & \vdots & & \end{matrix} \\ \begin{matrix} 0 \\ \\ \\ 0 \end{matrix} & \cdots & \begin{matrix} 0 \\ \begin{pmatrix} a^r_{11} & \cdots & a^r_{1d_r} \\ & & \\ a^r_{d_r1} & \cdots & a^r_{d_rd_r} \end{pmatrix} \end{matrix} \end{pmatrix}$$

Applying these observations to our vector space $V \approx K[X]/(P_1(X)) \amalg \cdots \amalg K[X]/(P_r(X))$, we let $V_i = K[X]/(P_i(X))$ and, if d_i is the degree of $P_i(X)$, we have the basis $\{\bar{1}, \bar{X}, \ldots, \bar{X}^{d_i-1}\}$ of V_i. Because $T|V_i$ is multiplication by X, we have the matrix

$$A^i = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \vdots & & & \\ 0 & 0 & 0 & 1 \\ -a_{id_i} & -a_{id_i-1} & \cdots & -a_{i1} \end{pmatrix} \qquad (*)$$

where $P_i(X) = X^{d_i} + a_{i1}X^{d_i-1} + \cdots + a_{id_i}$. Thus we have the following.

Proposition 5.1
If $T: V \to V$ is a linear transformation of a finite-dimensional vector space V over a field K, then V has a basis with respect to which T has the matrix of the form

$$\begin{pmatrix} A^1 & 0 & \cdots & 0 \\ 0 & A^2 & \cdots & 0 \\ 0 & 0 & \cdots & A^r \end{pmatrix}$$

where for each $i = 1, \ldots, r$, A^i is a square matrix of order d_i of the form (*), d_i is the degree of the invariant factor $P_i(X)$ of T, and $P_i(X) = X^{d_i} + a_{i1}X^{d_i-1} + \cdots + a_{id_i}$.

Our discussion thus far has made use of the invariant factor decomposition of a module over a PID. Now we shall make use of the decomposition involving elementary divisors. We recall from Chapter 10 that if we take the invariant factors $P_i(X)$, and factor them: $P_i(X) = P^{n_{i1}}_{i1}(X) \cdots P^{n_{is_i}}_{is_i}(X)$ where $P_{ij}(X)$ are

irreducible (hence prime) polynomials, then

$$\frac{K[X]}{(P_i(X))} \approx \coprod_{j=1}^{s_i} \frac{K[X]}{(P_{ij}^{\nu}(X))}$$

and thus

$$V \approx \coprod_{i=1}^{r} \coprod_{j=1}^{s_i} \frac{K[X]}{(P_{ij}^{\nu}(X))} \tag{1}$$

We have already seen how we may choose a basis of $K[X]/P(X)$ when $P(X)$ is any polynomial. We might ask, though, in view of (1), whether there is some nice way of choosing a basis for $K[X]/(P^{\nu}(X))$, given one for $K[X]/(P(X))$. To answer this question, let us look first at the case $\nu = 2$. We have the exact sequence of $K[X]$-modules (hence of vector spaces over K):

$$0 \longrightarrow \frac{(P(X))}{(P^2(X))} \longrightarrow \frac{K[X]}{(P^2(X))} \xrightarrow{k} \frac{K[X]}{(P(X))} \longrightarrow 0 \tag{2}$$

As a sequence of vector spaces over K, this sequence is splittable. Hence, we may find a basis for $K[X]/(P^2(X))$ as follows. Choose a basis $\{w_1, \ldots, w_t\}$ of $(P(X))/(P^2(X))$, and a basis $\{\bar{v}_1, \ldots, \bar{v}_d\}$ of $K[X]/(P(X))$. Letting $\bar{\bar{v}}_i \in K[X]/(P^2(X))$ be such that $k(\bar{\bar{v}}_i) = \bar{v}_i$, then the union $\{w_1, \ldots, w_t\} \cup \{\bar{\bar{v}}, \ldots, \bar{\bar{v}}_d\}$ is a basis for $K[X]/(P^2(X))$.

Now as a $K[X]$-module, $(P(X))/(P^2(X))$ is isomorphic to $K[X]/(P(X))$. To see this, we map $K[X]$ onto $(P(X))/(P^2(X))$ by sending the polynomial $F(X)$ in $K[X]$ to $\overline{F(X)P(X)}$ where $\overline{F(X)P(X)}$ denotes the coset class of $F(X)P(X)$ modulo $P^2(X)$. Clearly, the kernel of this morphism is the ideal generated by $P(X)$, so $K[X]/(P(X)) \approx (P(X))/(P^2(X))$. Thus, if $\{\bar{v}_1, \ldots, \bar{v}_d\}$ is a basis for $K[X]/(P(X))$, and if $v_1, \ldots, v_d$ are elements of $K[X]$ such that $\bar{v}_i$ is the coset of v_i in $K[X]/(P(X))$, then $\{\overline{P(X)v_1}, \ldots, \overline{P(X)v_d}\}$ is a basis for $(P(X))/(P^2(X))$ where $\overline{P(X)v_i}$ is the coset of $P(X)v_i$ in $(P(X))/(P^2(X))$. But it is also clear that if $\bar{\bar{v}}_i$ is the coset of v_i in $K[X]/(P^2(X))$, then $k(\bar{\bar{v}}_i) = \bar{v}_i$. Hence, we see that the set $\{\bar{\bar{v}}_1, \ldots, \bar{\bar{v}}_d, \overline{P(X)v_1}, \ldots, \overline{P(X)v_d}\}$ is a basis for $K[X]/(P^2(X))$. This description of a basis for $K[X]/(P^2(X))$ leads us to the following.

Lemma 5.2
Let $P(X)$ be an element of $K[X]$ of degree d, and let $\{\bar{v}_1, \ldots, \bar{v}_d\}$ be a basis for $K[X]/(P(X))$ as a vector space over K. Let $\{v_1, \ldots, v_d\}$ be elements of $K[X]$ such that $\bar{v}_i$ is the residue class of v_i in $K[X]/(P(X))$. If ν is a positive integer, the set of all products $\{P^j(X)v_i\}$ where $j = 0, \ldots, \nu - 1$, and $i = 1, \ldots, d$ is a set of $d\nu$ elements of $K[X]$. If $P^j(X)v_i$ denotes the coset of $P^j(X)v_i$ in $K[X]/(P^{\nu}(X))$, the set $\{\overline{P^j(X)v_i}\}$ is a basis for $K[X]/(P^{\nu}(X))$ as a vector space over K.

PROOF: Left as an exercise.

The following example not only serves to illustrate the use of Lemma 5.2, but is actually the most important type of application of 5.2.

Example 5.3 Let $P(X) = X - \lambda$, and consider $K[X]/((X-\lambda)^4)$. We know that $K[X]/(X-\lambda)$ is a one-dimensional vector space having, as basis, $\{\bar{1}\}$. Therefore, $\{\bar{\bar{1}}, \overline{(X-\lambda)\cdot 1}, \overline{(X-\lambda)^2 \cdot 1}, \overline{(X-\lambda)^3 \cdot 1}\}$ is a basis for $K[X]/((X-\lambda)^4)$. As before,

we know that multiplication by X is a linear transformation on $K[X]/((X-\lambda)^4)$, and we shall now see what its matrix is with respect to the above basis. We have

$$\begin{aligned}
X\cdot\overline{1} &= (X-\lambda)\cdot\overline{1}+\lambda\cdot\overline{1}\\
X\cdot\overline{(X-\lambda)\cdot 1} &= \overline{(X-\lambda)^2\cdot 1}+\lambda\overline{(X-\lambda)\cdot 1}\\
X\cdot\overline{(X-\lambda)^2\cdot 1} &= \overline{(X-\lambda)^3\cdot 1}+\lambda\overline{(X-\lambda)^2\cdot 1}\\
X\cdot\overline{(X-\lambda)^3\cdot 1} &= \overline{(X-\lambda)^4\cdot 1}+\lambda\overline{(X-\lambda)^3\cdot 1} = 0+\lambda\overline{(X-\lambda)^3\cdot 1}
\end{aligned}$$

Thus, the matrix with respect to the given basis is

$$\begin{pmatrix}\lambda & 1 & 0 & 0\\ 0 & \lambda & 1 & 0\\ 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & \lambda\end{pmatrix}$$

The reader should now find it easy to convince himself that if we had considered the general case $K[X]/((X-\lambda)^\nu)$, then this vector space has a basis with respect to which multiplication by X has the $\nu\times\nu$ matrix

$$\begin{pmatrix}\lambda & 1 & 0 & \cdots & 0 & 0\\ 0 & \lambda & 1 & \cdots & 0 & 0\\ \cdots & & & & & \\ \cdots & & & & & \\ 0 & 0 & 0 & \cdots & \lambda & 1\\ 0 & 0 & 0 & \cdots & & \lambda\end{pmatrix}$$

Example 5.4 Suppose V is a five-dimensional vector space over a field K, $T: V\to V$ is a linear transformation, and T has invariant factors $P_1(X)=(X-2)^2(X-1)$, $P_2(X)=(X-2)(X-1)$. Then we have

$$V\approx\frac{K[X]}{((X-2)^2)}\amalg\frac{K[X]}{((X-1))}\amalg\frac{K[X]}{((X-2))}\amalg\frac{K[X]}{((X-1))}$$

Each of the four modules has a basis with respect to which multiplication by X has a matrix like that of Example 5.3. Taking these modules in order, we have the four matrices

$$\begin{pmatrix}2 & 1\\ 0 & 2\end{pmatrix},\qquad (1),\qquad (2),\qquad \text{and}\qquad (1)$$

The union of these bases gives us a basis for V with respect to which multiplication by X (that is, the transformation T) has the matrix

$$\begin{pmatrix}2 & 1 & 0 & 0 & 0\\ 0 & 2 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 2 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}$$

The fact that $(X-2)^2$ occurs as a factor of $P_1(X)$ accounts for the element 1 off the diagonal.

It is now not much of a jump from the above examples to a more general situation. We assume that V is an n-dimensional vector space over the field K, and that $T: V \to V$ is a linear transformation. Further, we assume that all the irreducible factors of the invariant factors of T are linear. This would be the case, for instance, if K were the field of complex numbers. With these assumptions, we see that

$$V \approx \frac{K[X]}{((X-\lambda_1)^{\nu_1})} \amalg \cdots \amalg \frac{K[X]}{((X-\lambda_s)^{\nu_s})}$$

The reader should not assume here that all the λ_i need be distinct, as Example 5.4 above amply illustrates. For in that example, we have $s=4$ with $\lambda_1=\lambda_3=2$ and $\lambda_2=\lambda_4=1$. Nevertheless, using Lemma 5.2 and the subsequent examples, we know that V has a basis with respect to which T has the matrix

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_s \end{pmatrix}$$

where each A_i is a square matrix of order ν_i of the form

$$A_i = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & & \lambda_i \end{pmatrix}$$

We have therefore almost obtained the following.

Theorem 5.6
Let V be a finite-dimensional vector space over a field K and $T: V \to V$ a linear transformation. If all the irreducible factors of the characteristic polynomial of T are linear, then V has a basis with respect to which T has a matrix

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_s \end{pmatrix}$$

where each matrix A_i is as above.

PROOF: In the statement of this theorem we are assuming that the irreducible factors of the characteristic polynomial of T are all linear, whereas in the preceding discussion we had assumed that all the irreducible factors of the invariant factors of T were linear. But by Proposition 4.3 we know that the characteristic

polynomial of T is the product of the invariant factors of T so that the irreducible factors of T are linear if and only if those of the invariant factors are all linear. With this observation, the theorem is proven.

Definition
The form of the matrix of T given in the above theorem is called the **Jordan Canonical Form** of the transformation T.

EXERCISES

(1) Let F be a free module over the commutative ring R with basis $\{x_1, \ldots, x_n\}$. Let ξ be the element of $\mathrm{Sk}_n(F)$ defined by the property that $\xi(x_1, \ldots, x_n) = 1$. For each x in F, define the map $f_x : \underbrace{F \times \cdots \times F}_{n-1} \to R$ by $f_x(y_1, \ldots, y_{n-1}) = \xi(x, y_1, \ldots, y_{n-1})$.

(a) Show that for each x in F, f_x is in $\mathrm{Sk}_{n-1}(F)$.

(b) Prove that the map $f : F \to \mathrm{Sk}_{n-1}(F)$ defined by $f(x) = f_x$ is an isomorphism of R-modules.

(2) Let R be a commutative ring and M an R-module. If f is in $\mathrm{Sk}_p(M)$ and g is in $\mathrm{Sk}_{p-1}(M)$, define $\phi(f, g) : \underbrace{M \amalg R \times \cdots \times M \amalg R}_{p} \to R$ by

$$\phi(f, g)((m_1, r_1), \ldots, (m_p, r_p)) = f(m_1, \ldots, m_p) + \sum_{i=1}^{p} (-1)^{i+1} r_i g(m_1, \ldots, \hat{m}_i, \ldots, m_p)$$

where, as usual, $(m_1, \ldots, \hat{m}_i, \ldots, m_p)$ means the $(p-1)$-tuple obtained from $(m_1, \ldots, m_p)$ by omitting m_i.

(a) Show that $\phi(f, g)$ is in $\mathrm{Sk}_p(M \amalg R)$.

(b) Show that $\phi : \mathrm{Sk}_p(M) \amalg \mathrm{Sk}_{p-1}(M) \to \mathrm{Sk}_p(M \amalg R)$ sending (f, g) to $\phi(f, g)$ is a monomorphism of R-modules.

(c) If $h : M \amalg R \times \cdots \times M \amalg R \to R$ is an element of $\mathrm{Sk}_p(M \amalg R)$, define f in $\mathrm{Sk}_p(M)$ and g in $\mathrm{Sk}_{p-1}(M)$ as follows: $f(m_1, \ldots, m_p) = h((m_1, 0), \ldots, (m_p, 0))$ and $g(m_1, \ldots, m_{p-1}) = h((0, 1), (m_1, 0), \ldots, (m_{p-1}, 0))$. Prove that $h = \phi(f, g)$ and hence that ϕ is an epimorphism. From this conclude that ϕ is an isomorphism.

(3) Show that if R is an integral domain and I is an ideal of R, then $\mathrm{Sk}_p(I) = 0$ for all $p \geq 2$.

(4) Let K be a subfield of the field L. Let V be an n-dimensional vector space over K and let $f : V \to V$ be a K-morphism of V into V.

(a) Prove that $L \underset{K}{\otimes} V$ is an n-dimensional vector space over L.

(b) Prove that the characteristic polynomial of the L-morphism $L \underset{K}{\otimes} f : L \underset{K}{\otimes} V \to L \underset{K}{\otimes} V$ is in $K[X]$ and is equal to the characteristic polynomial of f.

(c) How do the invariant factors of f and of $L \underset{K}{\otimes} f$ compare?

(5) Let A be a 4×4 matrix over the complex numbers whose characteristic polynomial is $(X-2)(X+1)(X+i)(X+2i)$. Prove that A can be diagonalized.

(6) Prove that if K is a field and $f(X)$ is in $K[X] (f \neq 0)$, then there is a vector space V over K and a linear transformation $T : V \to V$ such that $f(X)$ is the characteristic polynomial of T.

(7) Let M be a finitely generated module over a commutative ring R and let $f: M \to M$ be an R-epimorphism. Prove that f is an isomorphism. [*Hint*: Show first that one may assume that R is a local ring. Next, assume that $F \xrightarrow{\alpha} M \to 0$ is an epimorphism where F is a finitely generated free module such that $F/JF \xrightarrow{\bar{\alpha}} M/JM$ is an isomorphism, where $J = \text{rad}(R)$. Then show that there is an endomorphism $\tilde{f}: F \to F$ such that $\alpha\tilde{f} = f\alpha$, and that $\tilde{f}$ is an epimorphism, hence an isomorphism. This shows that $|\tilde{f}|$ (the determinant of $\tilde{f}$) is a unit in R. Let $\Sigma_{i=0}^{n} a_i X^i$ be the characteristic polynomial of $\tilde{f}$. Then $a_0 = \pm|\tilde{f}|$ and $\tilde{f}^{-1} = (1/a_0)g$ where $g = \Sigma_{i=1}^{n} a_i \tilde{f}^{i-1}$. Use this to show that if x is in $\text{Ker}\,\alpha$, then $x = \tilde{f}(y)$ for some y in $\text{Ker}\,\alpha$. Finally, use this last fact to show that $\text{Ker}\, f = 0$ and hence that f is an isomorphism.]

(8) Let V be an n-dimensional vector space over the rational numbers, and let $v_1, \ldots, v_n$ be a fixed basis of V. If σ is a permutation of the set $\{1, \ldots, n\}$, let $T_\sigma: V \to V$ be the unique vector space endomorphism of V which sends each basis vector v_i to $v_{\sigma(i)}$.

(a) Prove that the matrix of T_σ with respect to the basis $\{v_1, \ldots, v_n\}$ has the property that exactly one entry in each row and each column is 1 while all the others are zero. Any matrix with this property is called a **permutation matrix**.

(b) Prove that the determinant of a permutation matrix is either 1 or -1.

(c) Prove that if M is a permutation matrix, then there is a permutation σ of the set $\{1, \ldots, n\}$ such that $M = T_\sigma$.

(d) Prove that the map which sends each permutation σ to the matrix T_σ is an isomorphism from the permutation group S_n onto the subgroup of permutation matrices.

(e) Prove that a permutation σ is even if and only if $|T_\sigma| = 1$ where $|T_\sigma|$ denotes the determinant of the matrix T_σ.

(9) Let A be a square matrix of order n over a commutative ring R, say $A = (a_{ij})$. Define the trace of A, denoted by $\text{Tr}(A)$, to be the element $\Sigma_{i=1}^{n} a_{ii}$.

(a) Prove that if B is a square matrix of order n over R such that $|B|$ is a unit in R, then $\text{Tr}(BAB^{-1}) = \text{Tr}(A)$.

(b) Let G be a finite group and let $R(G)$ be the group ring of G over R. Then $R(G)$ is a free R-module whose basis consists of the elements $g_1, \ldots, g_n$ of G. $R(G)$ is an $R(G)$-module and multiplication by each element x of $R(G)$ is an endomorphism of $R(G)$ as an R-module. Thus, to each element x in $R(G)$ is associated a square matrix A of order n, corresponding to the R-endomorphism of $R(G)$ induced by multiplication by x. Let $S(x) = \text{Tr}(A(x))$.

  (i) Prove that if g and g' are conjugate, then $S(g) = S(g')$.

  (ii) Prove that $A(g)$ is a permutation matrix for each g in G.

  (iii) Prove that $S(g) = n \cdot 1$ if g is the identity element of G and that $S(g) = 0$ if g is any other element of G.

(10) Let R be a ring and M an R-module. Then $M^* = \text{Hom}_R(M, R)$ is a right R-module and we can form the abelian group $M^* \underset{R}{\otimes} M$.

(a) Prove there exists a unique morphism $\phi: M^* \underset{R}{\otimes} M \to \text{Hom}_R(M, M)$ such that $[\phi(g \otimes m)](m') = g(m')m$. Show also that if R is a commutative ring, then ϕ is an R-morphism.

(b) Prove that if M is a free R-module with basis $\{x_1, \ldots, x_n\}$, then M^* is a free

right R-module with dual basis $\{x_1^*, \ldots, x_n^*\}$ where $x_i^*: M \to R$ is defined by setting $x_i^*(x_j) = \delta_{ij}$ where δ_{ij} is the Kronecker delta.

(c) Prove that if M is a free R-module having a finite basis, then the morphism ϕ defined in part (a) is an isomorphism. [*Hint*: Let $\{x_1, \ldots, x_n\}$ be a basis for M and let $\{x_1^*, \ldots, x_n^*\}$ be the dual basis. Show that if f is in $\mathrm{Hom}_R(M, M)$, then the element $\Sigma x_i^* f \otimes x_i$ in $M^* \underset{R}{\otimes} M$ is such that $\phi(\Sigma x_i^* f \otimes x_i) = f$.]

(d) Let R be a ring and M an R-module. Prove that there exists a unique morphism $T: M^* \underset{R}{\otimes} M \to R$ such that $T(g \otimes m) = g(m)$ for all g in M^* and m in M. Prove that T is an R-morphism when R is a commutative ring.

(e) Let R be a commutative ring, M a free R-module of finite rank, and $f: M \to M$ an R-endomorphism of M. Define $\mathrm{Tr}(f)$ in R to be the element $T(\phi^{-1}(f))$ where ϕ is the morphism defined in part (a). Prove that if $\{x_1, \ldots, x_n\}$ is a basis for M and $A = (a_{ij})$ is the matrix of f with respect to this basis, then $\mathrm{Tr}(f) = \mathrm{Tr}(A)$.

(11) Let V be an n-dimensional vector space over the field K, and let $f: V \to V$ be a K-endomorphism of V. If the characteristic polynomial of f is $X^n + a_1 X^{n-1} + \cdots + a_n$, prove that $\mathrm{Tr}(f) = -a_1$ and that $|f| = (-1)^n a_n$.

(12) Let V be an n-dimensional vector space over the field K, and let $f: V \to V$ be a K-endomorphism of V. An element λ in K is called an *eigenvalue* of f if there is some nonzero vector v in V such that $f(v) = \lambda v$. Such a vector v is called an *eigenvector of f for the eigenvalue λ*.

(a) Let λ be an element of K. Prove that λ is an eigenvalue of f if and only if λ is a root of the characteristic polynomial of f.

(b) Prove that if f has n distinct eigenvalues, then V has a basis consisting of eigenvectors of f. Show that the converse need not be true.

(c) Suppose that for some integer $d > 0$, we have $f^d = \mathrm{id}_V$. Prove that if λ is an eigenvalue of f, then $\lambda^d = 1$.

We shall now apply some of the above exercises to study some elementary facts about group representations.

In general, if K is a field, a $K(G)$-module M which is finite-dimensional as a K-vector space, is called a matrix representation of G, where $K(G)$ denotes the group ring of G over K. For in that case, we may choose a basis $\{x_1, \ldots, x_n\}$ for M as a vector space over K and, for every g in G, the matrix $A_M(g) = (a_{ij}(g))$ corresponding to the K-automorphism of M obtained by operation on M by the element g gives us a map from G to the group of square matrices of order n over K with nonzero determinant. This latter group is denoted by $GL(n, K)$, the general linear group over K. The map $g \to A_M(g)$ is actually a group morphism, as the reader can easily check. Because of this, we say that a $K(G)$-module is a *representation* of the group G. A simple $K(G)$-module is called an irreducible representation. We saw in the exercises to Chapter 7 that if K is a field whose characteristic does not divide the order of G, then $K(G)$ is semisimple so that every representation of G is a sum of irreducible representations. The number of nonisomorphic irreducible representations is finite, and each one is a finite-dimensional vector space over K.

(13) Let $\mathbf{C}(G)$ be the group ring of the finite group G over the complex numbers, and let $M_1, \ldots, M_t$ be all the nonisomorphic irreducible representations of G.

(a) Letting n_i denote the dimension of M_i as a vector space over $\mathbf{C}$, prove that $n = \Sigma_{i=1}^{t} n_i^2$ where n is the order of G.

(b) Let M be a finitely generated $\mathbf{C}(G)$-module. Then for each x in $\mathbf{C}(G)$, multiplication by x on M is an endomorphism f_x of M as a $\mathbf{C}$-vector space. Define $S_M: \mathbf{C}(G) \to \mathbf{C}$ by $S_M(x) = \mathrm{Tr}(f_x)$. Denote the restriction of S_M to G by $\chi_M: G \to \mathbf{C}$.

(i) Prove that $\chi_M(g)$ is a sum of roots of unity, that is, $\chi_M(g) = \Sigma \lambda_i$ where $\lambda_i^{\nu_i} = 1$ for some $\nu_i \in \mathbf{N}$. Use the fact that every polynomial in $\mathbf{C}[X]$ has all of its roots in $\mathbf{C}[X]$.

(ii) Prove that if g and g' are conjugate elements of G, then $\chi_M(g) = \chi_M(g')$.

(iii) Prove that if M' and M'' are finitely generated $\mathbf{C}(G)$-modules and if $M = M' \amalg M''$, then $S_M = S_{M'} + S_{M''}$.

(iv) Let M' and M'' be finitely generated $\mathbf{C}(G)$-modules. Define an operation on $M' \underset{\mathbf{C}}{\otimes} M''$ by setting $g(m_1 \otimes m_2) = gm_1 \otimes gm_2$ for all g in G, m_1 in M', and m_2 in M''. Show that this does make $M' \otimes M''$ into a finitely generated $\mathbf{C}(G)$-module.

(v) With M' and M'' as above, let $M = M' \underset{\mathbf{C}}{\otimes} M''$. Prove that $S_M = S_{M'} S_{M''}$.

(c) We have assumed that the irreducible representations M_i are vector spaces of dimension n_i over $\mathbf{C}$. Let e_i denote the identity automorphism of M_i. Because $\mathbf{C}(G) \approx \prod_{i=1}^{t} \mathrm{End}_{\mathbf{C}}(M_i)$, we may assume that e_i is in $\mathbf{C}(G)$ for each $i = 1, \ldots, t$ and, in fact, $1 = \Sigma e_i$. Prove that $S_{M_i}(e_i) = n_i$ and that $S_{M_i}(e_j) = 0$ for $i \neq j$.

(d) Let M be a finitely generated $\mathbf{C}(G)$-module and suppose that $M = M_1^{u_1} \amalg \cdots \amalg M_t^{u_t}$ where the M_i are the irreducible $\mathbf{C}(G)$-representations and u_i are nonnegative integers.

(i) Prove that the integers $u_1, \ldots, u_t$ are uniquely determined by M.

(ii) Prove that $S_M = u_1 S_{M_1} + \cdots + u_t S_{M_t}$.

(iii) Show that the map S_M determines the $\mathbf{C}(G)$-module M. From this, deduce that χ_M determines the module M.

(e) Let $M = \mathbf{C}(G)$. M is called the **left regular representation of G**. We will denote S_M and χ_M by S_G and χ_G, respectively, in this case.

(i) Prove that $\chi_G = n_1 \chi_{M_1} + \cdots + n_t \chi_{M_t}$. Consequently:

(ii) Prove that $(1/n) \Sigma n_i \chi_{M_i}(1) = 1$ while $\Sigma n_i \chi_{M_i}(g) = 0$ if g is any element of G other than 1.

(14) Let M be a finitely generated $K(G)$-module, where K is a field. Denote $\mathrm{Hom}_K(M, K)$ by M^*. For g in G and f in M^*, define gf in M^* by $(gf)(m) = f(g^{-1}m)$.

(a) Prove that the operation of G on M^* defined above makes M^* a finitely generated left $K(G)$-module. M^* is called the **contragradient representation of M**.

(b) Prove that $M \approx M^{**}$.

(c) Prove that M is a simple $K(G)$-module if and only if M^* is a simple $K(G)$-module.

(d) Let $\{x_1, \ldots, x_s\}$ be a basis for M over K, and let $\{x_1^*, \ldots, x_s^*\}$ be the dual basis for M^*. If $A_M(g) = (a_{ij}(g))$ denotes the matrix corresponding to the operation of g on M with respect to the basis $\{x_1, \ldots, x_s\}$, and $A_{M^*}(g) = (b_{ij}(g))$ denotes the corresponding matrix for M^*, prove that $b_{ij}(g) = a_{ji}(g^{-1})$.

(e) Assume now that $K = \mathbf{C}$. Prove that $\chi_M(g^{-1}) = \chi_{M^*}(g) = \overline{\chi_M(g)}$ for all g in G, where $\overline{\chi_M(g)}$ denotes the complex conjugate of $\chi_M(g)$.

(f) Still assuming that $K = \mathbf{C}$, prove that M^* and M are isomorphic if and only if $\chi_M(g)$ is real for every g in G. [*Hint*: Use Exercise 13 (d) (iii).]

(15) If G is a finite group of order n, we have $\mathbf{C}(G) = \prod_{i=1}^{t} R_i$ where $R_i = \mathrm{End}_{\mathbf{C}}(M_i)$ and $M_1, \ldots, M_t$ are the distinct irreducible representations of G over $\mathbf{C}$. We let $\chi_i = \chi_{M_i}$.

(a) If $e_i = \sum_{g \in G} a_{ig} g$, prove that $a_{ig} = (n_i/n)\chi_i(g^{-1})$ where $e_i = \mathrm{id}_{M_i}$, $n_i =$ dimension of M_i over $\mathbf{C}$. [*Hint*: Use Exercise 11(c), (e).]

(b) If N_1 and N_2 are $\mathbf{C}(G)$-modules, define $\langle \chi_{N_1}, \chi_{N_2} \rangle = (1/n)\sum \chi_{N_1}(g)\chi_{N_2}(g^{-1})$. Prove that $\langle \chi_i, \chi_j \rangle = \delta_{ij}$. [*Hint*: Use part (a) together with the fact that $\chi_i(e_j) = n_i \delta_{ij}$.]

(16) Let K be an algebraically closed field, that is, a field with the property that every polynomial in $K[X]$ has its roots in K (for example, the field $\mathbf{C}$). Let V be a finite-dimensional vector space over K and let G be a finite abelian subgroup of $\mathrm{Aut}_K(V)$. Prove that there is a basis for V such that with respect to this basis, every element of G is a diagonal matrix.

In the next set of exercises we give a brief introduction to the notion of graded rings and graded algebras.

We begin with some preliminary definitions and observations. Let R be a ring. A representation of the underlying abelian group into a sum $\coprod_{i \in \mathbf{N}} R_i$ is called a **grading of R** if for all i and j in $\mathbf{N}$ we have $R_i R_j \subset R_{i+j}$. A ring R together with a grading is called a **graded ring**. If $R = \coprod_{i \in \mathbf{N}} R_i$ is a graded ring, then the subgroup R_i is called the homogeneous component of R of degree i and the elements of R_i are called the homogeneous elements of R of degree i. Obviously, each element r of R can be written in one and only one way as $\sum_{i=0}^{\infty} r_i$ where each r_i is homogeneous of degree i and all but a finite number of the $r_i = 0$. Each r_i is called the **ith homogeneous component** of N.

(17) Let $R = \coprod_{i \in \mathbf{N}} R_i$ be a graded ring. Given two elements $\sum_{i=0}^{\infty} r_i$ and $\sum_{i=0}^{\infty} r'_i$, show:

(a) $\sum_{i=0}^{\infty} r_i + \sum_{i=0}^{\infty} r'_i = \sum_{i=0}^{\infty} (r_i + r'_i)$.

(b) $(\sum_{i=0}^{\infty} r_i)(\sum_{i=0}^{\infty} r'_i) = \sum_{n=0}^{\infty} \sum_{k+l=n} (r_k r'_l)$.

Further, show that:

(c) R_0 is a subring of R.

(d) Each R_i is both a left and right R_0-module by means of the map $R_0 \times R_i \to R_i$ given by $(r_0, r_i) \to r_0 r_i$ and $R_i \times R_0 \to R_i$ given by $(r_i, r_0) \to r_i r_0$.

(18) Suppose $R = \coprod_{i \in \mathbf{N}} R_i$ and $R' = \coprod_{i \in \mathbf{N}} R'_i$ are two graded rings. By a **morphism of graded rings** from R to R' we mean a morphism of rings $f: R \to R'$ such that $f(R_i) \subset R'_i$ for all i in $\mathbf{N}$.

(a) Suppose that $R'' = \coprod_{i \in \mathbf{N}} R''_i$ is a third graded ring and $f: R \to R'$ and $g: R \to R''$ are morphisms of graded rings. Show that the usual composition of ring morphisms $gf: R \to R''$ is also a morphism of graded rings, which is called the composition of the morphisms of the graded morphisms f and g.

(b) Show that the following data define a category, Gr Rings, called the category of graded rings. The objects of Gr Rings are the graded rings, the morphisms

between the objects of Gr Rings are the graded morphisms, and the composition in Gr Rings is the composition of graded morphisms we just defined.

(c) Show that a morphism $f: R \to R'$ of rings is a graded morphism if and only if for each r in R and i in $\mathbf{N}$ we have $f(r_i)$ is the ith homogeneous component of $f(r)$.

(19) Suppose $R = \coprod_{i \in \mathbf{N}} R_i$ and $R' = \coprod_{i \in \mathbf{N}} R'_i$ are graded rings. Show that if $f: R \to R'$ is a morphism of graded rings, then the ideal $\operatorname{Ker} f$ has the property that r is in $\operatorname{Ker} f$ if and only if each homogeneous component r_i of r is also in $\operatorname{Ker} f$.

(20) An ideal I in a graded ring $R = \coprod_{i \in \mathbf{N}} R_i$ is said to be a **homogeneous** or **graded ideal** of R if an element r in R is in I if and only if each homogeneous component r_i of r is in I. Show that for each ideal I of R the following statements are equivalent:

(a) I is a homogeneous ideal in R.

(b) I is generated (as an ideal) by homogeneous elements of R.

(c) $I = \coprod_{i \in \mathbf{N}} R_i \cap I$.

If I is homogeneous ideal in R, we denote $R_i \cap I$ by I_i for all i in $\mathbf{N}$. We call I_i the ith homogeneous component of I.

(d) Suppose I is a homogeneous ideal of the graded ring R.

(i) Show that there is a natural isomorphism of abelian groups $R/I \to \coprod_{i \in \mathbf{N}} R_i/I_i$.

(ii) Show that this sum decomposition of R/I is a grading of R/I. The graded rings consisting of the ring R/I together with the grading $R/I = \coprod_{i \in \mathbf{N}} R_i/I_i$ is called the **factor ring of R by I** in the category of graded rings.

(iii) Show that the canonical surjective ring morphism $k: R \to R/I$ is a graded morphism.

(e) Show that each morphism in the category of graded rings has an analysis.

(21) Let K be a commutative ring. A graded K-algebra is a graded ring $R = \coprod_{i \in \mathbf{N}} R_i$ together with a ring morphism $f: K \to R$ such that: (1) $f: K \to R$ is a K-algebra, (2) f is an injective morphism with $f(K) = R_0$.

(a) Let M be a K-module. Show that the representation of the tensor algebra $T_K(M)$ as the sum $\coprod_{i \in \mathbf{N}} \overset{i}{\otimes} M$ is a grading on $T_K(M)$. This is the only way we ever consider $T_K(M)$ a graded ring. Show that the usual K-algebra structure $K \to T_K(M)$ on $T_K(M)$ makes $T_K(M)$ a graded K-algebra. This is the only way we consider $T_K(M)$ a K-algebra.

(b) Let M be a K-module. Show that the kernel I of the canonical surjective morphism $T_K(M) \to S_K(M)$ is a homogeneous ideal of $T_K(M)$. Hence, $S_K(M)$ is a graded ring because it is a factor ring of a graded ring by a homogeneous ideal. This is the only way we consider the symmetric algebra $S_K(M)$ a graded ring. Show that the usual K-algebra structure $K \to S_K(M)$ makes $S_K(M)$ a graded K-algebra. This is the only K-algebra we consider on $S_K(M)$.

(c) Let $R = K[X_j]_{j \in J}$ be a polynomial ring over K. For each n in $\mathbf{N}$, let R_n be the R-submodule generated by the monomials $\prod_{\sum n_i = n} X_i^{n_i}$ of degree n. Show that the

representation $R = \coprod_{n\in\mathbb{N}} R_n$ is a grading on R which makes R a graded K-algebra.

(d) Let F be the free K-module generated by the set J. Show that $S_K(F)$ and $K[X_j]_{j\in J}$ are isomorphic graded K-algebras.

(22) Let K be a commutative ring. A graded K-algebra $R = \coprod_{i\in\mathbb{N}} R_i$ is said to be **generated by the homogeneous elements of degree 1** if the subring generated by $R_0 \amalg R_1$ is all of R.

(a) Let M be a K-module. Show that the K-algebras $T_K(M)$ and $S_K(M)$ are generated by the homogeneous elements of degree 1.

(b) Let $K \to R$ be a graded K-algebra generated by its homogeneous elements of degree 1. Suppose $K \to \Lambda$ is a K-algebra and $f, g : R \to \Lambda$ are two K-algebra morphisms (not necessarily graded). Then $f = g$ if and only if $f|R_1 = g|R_1$.

(c) Show that if $f : R \to \Lambda$ is a graded surjective morphism of graded K-algebras, then Λ is generated by its homogeneous elements of degree 1 if R is generated by its homogeneous elements of degree 1.

(d) Let $R = \coprod_{i\in\mathbb{N}} R_i$ be a graded K-algebra which is generated by its homogeneous elements of degree 1. If I is a homogeneous ideal such that $I \cap R_0 = 0$, then R/I is a graded K-algebra which is generated by its homogeneous elements of degree 1.

(e) Let $R = \coprod_{i\in\mathbb{N}} R_i$ be a graded K-algebra which is generated by its homogeneous elements of degree 1. Then R_1 is a K-module. Show that the unique K-algebra $f : T_K(M) \to R$, such that $f|R_1 : R_1 \to R$ is the inclusion morphism, is a surjective graded K-algebra morphism.

(23) Let $R = \coprod_{i\in\mathbb{N}} R_i$ and $R' = \coprod_{i\in\mathbb{N}} R'_i$ be two graded K-algebras where K is an arbitrary commutative ring.

(a) Show that the K-module $R \underset{K}{\otimes} R' = \coprod_{n\in\mathbb{N}} \coprod_{i+j=n} (R_i \underset{K}{\otimes} R_j)$.

(b) Denoting $\coprod_{i+j=n} R_i \underset{K}{\otimes} R_j$ by $(R \underset{K}{\otimes} R')_n$, show that for each n in N, the above sum representation $R \underset{K}{\otimes} R' = \coprod_{n\in\mathbb{N}} (R \underset{K}{\otimes} R')_n$ is a grading on the K-algebra $R \underset{K}{\otimes} R'$ which makes $R \underset{K}{\otimes} R'$ a graded K-algebra called the graded tensor product of R and R'.

(c) Show that if R and R' are both generated by their homogeneous elements of degree 1, then their graded tensor product $R \underset{K}{\otimes} R'$ is also generated by its homogeneous elements of degree 1.

(d) Extend the notion of a tensor product of two K-algebras to the tensor product of a finite number of K-algebras.

(e) Extend the notion of the graded tensor product of two graded K-algebras to the graded tensor product of a finite number of graded K-algebras.

(f) Suppose the K-module $M = M_1 \amalg \cdots \amalg M_t$, a finite sum. Then

(i) $(S_K(M_1) \underset{K}{\otimes} \cdots \underset{K}{\otimes} S_K(M_t))_1 \approx M_1 \amalg \cdots \amalg M_t = M$.

(ii) The unique K-algebra morphism $f : S_K(M) \to S_K(M_1) \underset{K}{\otimes} \cdots \underset{K}{\otimes} S_K(M_t)$ with the property $f|M : M \to S_K(M_1) \underset{K}{\otimes} \cdots \underset{K}{\otimes} S_K(M_t)$ is given by the monomorphism $M \to (S_K(M_1) \underset{K}{\otimes} \cdots \underset{K}{\otimes} S_K(M_t))_1$ is a graded K-algebra isomorphism.

In the next set of exercises we give a brief introduction to exterior algebras.

(24) Let M be a module over a commutative ring K. Let I be the ideal in $T_K(M)$ generated by the homogeneous elements of degree 2 of the form $x \otimes x$ for all x in M. Show:

(a) The ideal I is a homogeneous ideal in $T_K(M)$ with the property $I \cap (T_K(M))_0 = (0)$.

(b) The graded ring $T_K(M)/I$ is a graded K-algebra which is generated by its homogeneous elements of degree 1.

The graded K-algebra $T_K(M)/I$ is called the **exterior algebra** of M over K and is usually denoted by ΛM. Also, for each n in $\mathbf{N}$, the nth component $(\Lambda M)_n$ of ΛM is denoted by $\overset{n}{\Lambda} M$. Finally, if $k : T_K(M) \to \Lambda M$ is the canonical surjective K-algebra morphism and $x_1, \ldots, x_s$ are elements of M, then the image in ΛM under k of the element $x_1 \otimes \cdots \otimes x_s$ in the sth component $\overset{s}{\otimes} M$ of $T_K(M)$ is denoted by $x_1 \wedge \cdots \wedge x_s$.

(c) If x is in M, then $x \wedge x = 0$.

(d) If x_1, x_2, x are in M, then:

(i) $(x_1 + x_2) \wedge x = x_1 \wedge x + x_2 \wedge x$.

(ii) $x \wedge (x_1 + x_2) = x \wedge x_1 + x \wedge x_2$.

(iii) $x_1 \wedge x_2 = -x_2 \wedge x_1$. [*Hint*: Use the fact that $(x_1 + x_2) \wedge (x_1 + x_2) = 0$.]

(e) Let σ be a permutation of $[1, \ldots, n]$. Then $x_{\sigma(1)} \wedge \cdots \wedge x_{\sigma(n)} = \operatorname{sgn} \sigma (x_1 \wedge \cdots \wedge x_n)$ where $\operatorname{sgn} \sigma$ is the signature of σ (that is, $\operatorname{sgn} \sigma = 1$ if σ is even and $\operatorname{sgn} \sigma = -1$ if σ is odd).

(f) Let $x_1, \ldots, x_n$ be n elements in M. If there are distinct i and j such that $x_i = x_j$, then $x_1 \wedge \cdots \wedge x_n = 0$.

(g) If k is in K, then

$$x_1 \wedge \cdots \wedge kx_i \wedge \cdots \wedge x_n = k(x_1 \wedge \cdots \wedge x_n)$$

for all $x_1, \ldots, x_n$ in M.

(h) $x_1 \wedge \cdots \wedge x_i \wedge (x + x') \wedge x_{i+2} \wedge \cdots \wedge x_n = x_1 \wedge \cdots \wedge x_i \wedge x \wedge x_{i+2} \wedge \cdots \wedge x_n + x_1 \wedge \cdots \wedge x_i \wedge x' \wedge x_{i+2} \wedge \cdots \wedge x_n$.

(i) $\overset{1}{\Lambda} M = M$.

(25) Let $K \to \Gamma$ be a K-algebra and M a K-module.

(a) Suppose $f : \Lambda M \to \Gamma$ is a K-algebra morphism. Then $f|\overset{1}{\Lambda} M : M \to \Gamma$ is a morphism of K-modules with the property $f(m_1) f(m_2) = 0$ for all m_1, m_2 in M.

(b) If $f, g : \Lambda M \to \Gamma$ are two K-algebra morphisms, then $f = g$ if and only if $f|\overset{1}{\Lambda} M = g|\overset{1}{\Lambda} M$.

(c) Given any K-module morphism $h : M \to \Gamma$ with the property that $h(m_1)h(m_2) = 0$ for all m_1, m_2 in M, then there exists a unique K-algebra morphism $f : \Lambda M \to \Gamma$ such that $f|\overset{1}{\Lambda} M = h$.

(26) Let M be a module over the commutative ring K. Suppose that $K \to \Gamma$ is a K-algebra and $g : M \to \Gamma$ a morphism of K-modules having the properties: (1) $g(m_1)g(m_2) = 0$ for all m_1, m_2 in M and (2) given any K-algebra Σ and any K-module morphism $h : M \to \Sigma$ satisfying $h(m_1)h(m_2) = 0$ for all m_1, m_2, there exists a unique K-algebra morphism $f : \Lambda M \to \Gamma$ with the property that $f|\overset{1}{\Lambda} M = g$ is an isomorphism of K-algebras.

(27) Let Λ_1 and Λ_2 be two K-algebras and Γ an arbitrary K-algebra. Let $f_1 : \Lambda_1 \to \Lambda_1 \underset{K}{\otimes} \Lambda_2$ and $f_2 : \Lambda_2 \to \Lambda_1 \underset{K}{\otimes} \Lambda_2$ be the K-algebra morphisms $f_1(x_1) = x_1 \otimes 1$ for all x_1 in Λ_1 and $f_2(x_2) = 1 \otimes x_2$ for all x_2 in Λ_2.

(a) Let $f : \Lambda_1 \underset{K}{\otimes} \Lambda_2 \to \Gamma$ be a morphism of K-algebras. Show that the compositions $ff_1 : \Lambda_1 \to \Gamma$ and $ff_2 : \Lambda_2 \to \Gamma$ have the property $ff_1(x_1)ff_2(x_2) = ff_2(x_2)ff_1(x_1)$ for all x_1 in Λ_1 and x_2 in Λ_2.

(b) If $f, g : \Lambda_1 \underset{K}{\otimes} \Lambda_2 \to \Gamma$ are two K-algebra morphisms, then $f = g$ if and only if $ff_1 = gf_1$ and $ff_2 = gf_2$.

(c) Given any K-algebra morphisms $g_1 : \Lambda_1 \to \Gamma$ and $g_2 : \Lambda_2 \to \Gamma$, then there is a unique K-algebra morphism $g : \Lambda_1 \underset{K}{\otimes} \Lambda_2 \to \Gamma$ such that $gf_1 = g_1$ and $gf_2 = g_2$.

(d) Extend the results in **(a)**, **(b)**, and **(c)** to tensor products of a finite number of K-algebras.

(28) Let M be a K-module and suppose that $M = M_1 \amalg M_2$. Show that ΛM and $\Lambda M_1 \underset{K}{\otimes} \Lambda M_2$ are isomorphic graded K-algebras, as outlined below:

(a) Show that the map $f : M_1 \amalg M_2 \to \Lambda M_1 \underset{K}{\otimes} \Lambda M_2$ given by $f((m_1, m_2)) = m_1 \otimes 1 + 1 \otimes m_2$ for all m_1 in M_1 and m_2 in M_2 has the following properties:
- **(i)** f is an injective K-module morphism whose image is $(\Lambda M_1 \underset{K}{\otimes} \Lambda M_2)_1$.
- **(ii)** $f((m_1, m_2))f((m_1', m_2')) = 0$ for all m_1, m_1' in M_1 and m_2, m_2' in M_2.

(b) Suppose Γ is a K-algebra and $g : M_1 \amalg M_2 \to \Gamma$ is a K-module morphism with the property that $g((m_1, m_2))g((m_1', m_2')) = 0$ for all m_1, m_1' in M_1 and m_2, m_2' in M_2. Letting $g_1 = g|M_1$ and $g_2 = g|M_2$, show:
- **(i)** There exist K-algebra morphisms $g_1' : \Lambda M_1 \to \Gamma$ and $g_2' : \Lambda M_2 \to \Gamma$ such that $g_1'|\overset{1}{\Lambda} M_1 = g_1$, $g_2'|\overset{1}{\Lambda} M_2 = g_2$ and $g_1'(x)g_2'(y) = 0$ for all x in ΛM_1 and y in ΛM_2.
- **(ii)** There is a unique K-algebra morphism $g' : \Lambda M_1 \underset{K}{\otimes} \Lambda M_2 \to \Gamma$ such that $g'f : M_1 \amalg M_2 \to \Gamma$ is the given $g : M_1 \amalg M_2 \to \Gamma$.

(c) The unique K-algebra morphism $h : \Lambda(M_1 \amalg M_2) \to \Lambda(M_1) \underset{K}{\otimes} \Lambda(M_2)$ with the property $h|\overset{1}{\Lambda}(M_1 \amalg M_2) = f$ is an isomorphism of graded K-algebras. This uniquely determined isomorphism will usually be considered an identification.

(29) Generalize Problem 28 to show that if K is a commutative ring and $M = M_1 \amalg \cdots \amalg M_n$ is a K-module, then the graded K-algebras ΛM and $\Lambda M_1 \underset{K}{\otimes} \cdots \underset{K}{\otimes} \Lambda M_n$ are isomorphic.

(30) Let K be a commutative ring and I an ideal in K. Show that the K-algebra $\Lambda K/I$ has the following properties:

(a) $\overset{2}{\Lambda} K/I = 0$ and so $\overset{n}{\Lambda} K/I = 0$ for all $n \geq 2$.

(b) The graded K-algebra $\Lambda K/I$ is isomorphic to the graded K-algebra $K + K/I$ described as follows:
- **(i)** As a K-module $K + K/I$ is the sum $K \amalg K/I$.
- **(ii)** $(x, y)(x', y') = (xx', xy' + x'y)$.
- **(iii)** The grading of $K + K/I$ is given by the sum decomposition $K \amalg K/I$.

(31) Let $I_1, \ldots, I_n$ be a finite family of proper ideals in the commutative ring K. Let $M = \overset{n}{\underset{j=1}{\amalg}} K/I_j$. Prove that the underlying K-module structure of the K-algebra ΛM

can be described as follows:

(a) $\overset{p}{\Lambda} M = 0$ for all $p > n$.

(b) $\overset{p}{\Lambda} M \neq 0$ for $p \leq n$.

(c) Suppose $1 \leq p \leq n$ and let $V_1, \ldots, V_{\binom{n}{p}}$ be the $\binom{n}{p}$ distinct subsets of $[1, n]$ with p elements each. For each V_j let I_{V_j} be the ideal of K generated by all the ideals I_i with $i \in V_j$. Then $\overset{p}{\Lambda} M$ is isomorphic to $K/I_{V_1} \amalg K/I_{V_2} \amalg \cdots \amalg K/I_{V_{\binom{n}{p}}}$.

(32) Let $I_1 \subset I_2 \subset \cdots \subset I_n \neq K$ be ideals in the commutative ring K, and let $M = \coprod_{i=1}^{n} K/I_i$. Show that the underlying K-module of the K-algebra ΛM has the following properties:

(a) $\overset{p}{\Lambda} M = 0$ for all $p > n$.

(b) $\operatorname{ann}(\overset{p}{\Lambda} M) = I_p$ for each $p = 1, \ldots, n$ and so $\overset{p}{\Lambda} M \neq 0$ for $p = 1, \ldots, n$.

Consequently:

(c) If $I_1' \subset I_2' \subset \cdots \subset I_m' \neq K$ are also ideals in K such that the K-module $M' = \coprod_{i=1}^{m} K/I_i'$ is isomorphic to M, then $m = n$ and $I_i = I_i'$ for all $i = 1, \ldots, n$.

(33) Let M be a free K-module of rank n. Prove that the underlying K-module of the K-algebra ΛM has the following properties:

(a) $\overset{p}{\Lambda} M = 0$ for all $p > n$.

(b) $\overset{p}{\Lambda} M$ is a free K-module of rank $\binom{n}{p}$ for each $p = 1, \ldots, n$.

(34) Let K be a commutative ring.

(a) Suppose $f: M \to N$ is a morphism of K-modules. Show that the uniquely determined K-algebra morphism $f': \Lambda M \to \Lambda N$ such that $f'|M$ is the composition $M \xrightarrow{f} N \to \Lambda N$, is a graded K-algebra morphism. This uniquely determined graded K-algebra morphism f' is usually denoted by Λf and the induced morphisms $\overset{p}{\Lambda} M \to \overset{p}{\Lambda} N$ are usually denoted by $\overset{p}{\Lambda} f: \overset{p}{\Lambda} M \to \overset{p}{\Lambda} N$.

(b) Show that if $f: M \to M'$ and $g: M' \to M''$ are K-module morphisms, then $\Lambda g \Lambda f = \Lambda gf$.

(c) If $f: M \to M$ is the id_M, then Λf is the identity on ΛM.

(d) If $F: M \to N$ is an isomorphism, then Λf is an isomorphism.

(e) If $f: M \to N$ is surjective, then Λf is surjective.

(f) If a K-module M can be generated by n elements, then $\overset{p}{\Lambda} M = 0$ for all $p > n$.

(35) Let M be a module over the commutative ring K. Suppose n is a positive integer and X is a K-module. Let $\overset{n}{\times} M$ be the n-fold Cartesian product of M with itself. A map $f: \overset{n}{\times} M \to X$ is called **n-linear** if for all $m_1, \ldots, m_n, m$ in M and r, s in K we have $f(m_1, \ldots, rm_i + sm, m_{i+1}, \ldots, m_n) = rf(m_1, \ldots, m_i, m_{i+1}, \ldots, m_n) + sf(m_1, \ldots, m, m_{i+1}, \ldots, m_n)$ for each $i = 1, \ldots, n$. The n-linear map is said to be **skew-symmetric** if $f(m_1, \ldots, m_n) = 0$ whenever $m_i = m_{i+1}$ for some $i = 1, \ldots, n-1$.

(a) Show the map $g: \overset{n}{\times} M \to \overset{n}{\otimes} M$ given by $g(m_1, \ldots, m_n) = m_1 \otimes \cdots \otimes m_n$ is n-linear, having the following properties:

(i) If $f: \overset{n}{\otimes} M \to X$ is a morphism of K-modules, then $fg: \overset{n}{\times} M \to X$ is n-linear.

(ii) Suppose that X is a K-module. A map $h : \overset{n}{\times} M \to X$ is n-linear if and only if there is a K-morphism $f : \overset{n}{\otimes} M \to X$ such that $fg = h$.

(iii) If $f_1, f_2 : \overset{n}{\otimes} M \to X$ are two K-morphisms such that $f_1 g = f_2 g$, then $f_1 = f_2$. Hence, we obtain an isomorphism between the set of n-linear maps $\overset{n}{\times} M \to X$ and $\mathrm{Hom}_K(\overset{n}{\otimes} M, X)$.

(b) Show that for each positive integer n the map $\beta : \overset{n}{\times} M \to \overset{n}{\Lambda} M$ given by $\beta(m_1, \ldots, m_n) = m_1 \wedge \cdots \wedge m_n$ is a skew-symmetric map having the following properties:

(i) If $f : \overset{n}{\Lambda} M \to X$ is a morphism of K-modules, then $f\beta : \overset{n}{\times} M \to X$ is a skew-symmetric map.

(ii) For each K-module X, let $\mathrm{Sk}_n(M, X)$ be the set of skew-symmetric maps from M to X. Show that $\mathrm{Sk}_n(M, X)$ has a natural K-module structure such that the map $\mathrm{Hom}_K(\overset{n}{\Lambda} M, X) \to \mathrm{Sk}_n(M, X)$ given by $f \mapsto f\beta$ for all f in $\mathrm{Hom}_K(\overset{n}{\Lambda} M, X)$ is an isomorphism of K-algebras.

(36) Let F be a free K-module of rank n and $f : F \to F$ a morphism of K-modules. Show:

(a) $\overset{n}{\Lambda} F$ is isomorphic to K.

(b) $\overset{n}{\Lambda} f : \overset{n}{\Lambda} F \to \overset{n}{\Lambda} F$ is the K-module morphism multiplication by $|f|$, the determinant of f.

(c) Use (b) to show that if $g : F \to F$ is a K-morphism, then $|gf| = |g||f|$.

PART FOUR

Chapter 12 ALGEBRAIC FIELD EXTENSIONS

In the preceding section, we proved that if the irreducible factors of the characteristic polynomial of a linear transformation were all linear, then that linear transformation had a matrix of a certain canonical form. We will see that the assertion that all the irreducible factors of a polynomial over a field are linear means that all the roots of the polynomial are in the field. This leads us naturally to consider the question of whether a polynomial over a field K has some roots in K or, for that matter, whether it has roots anywhere. This question takes us finally to the study of algebraic extensions of fields.

Although we are primarily interested in fields at this point, we can with little loss of time study polynomials over a commutative ring R and then specialize to the case when R is an integral domain and, in particular, a field.

1. ROOTS OF POLYNOMIALS

Let R be any commutative ring, and let $f \in R[X]$ be a monic polynomial. Then we have the following.

Lemma 1.1

(a) f is a regular element of $R[X]$.
(b) If g is any element of $R[X]$, then there exist elements q and r in $R[X]$ such that $g = qf + r$, with $r = 0$ or $\deg r < \deg f$. Moreover, q and r are unique.

PROOF: The proof of (**a**) is completely trivial. The proof of the existence of the elements q and r in (**b**) follows almost line for line the proof of the fact that if K is a field, then $K[X]$ is a Euclidean ring. The one observation that must be made in transposing this proof to the case of an arbitrary ring R is that because f is assumed to be monic, there is no fudging required on the leading coefficient of f. The uniqueness of the elements q and r follows from the fact that f is a regular element. For, if $g = q_1 f + r_1 = q_2 f + r_2$, we have $(q_1 - q_2)f = r_2 - r_1$. Because $\deg(r_2 - r_1) < \deg f$ and $\deg(q_1 - q_2)f \geq \deg f$ if $q_1 - q_2 \neq 0$, we must have $r_2 - r_1 = 0$ and $(q_1 - q_2)f = 0$. But then $q_1 - q_2 = 0$ because f is regular, and the proof of uniqueness is complete.

The reader should observe that Lemma 1.1 could be used to prove that $K[X]$ is Euclidean when K is a field.

Now let a be an element of the ring R and let f be the monic polynomial $X - a$. If g is any element of $R[X]$ we have $g = q(X - a) + r$ where $r = 0$ or $\deg r = 0$, that is, r is an element of R. If a is a root of g, that is, if $g(a) = 0$, then we have $0 = g(a) = q(a)(a - a) + r$, so that $r = 0$. Conversely, if $r = 0$, we have $g(a) = q(a)(a - a) = q(a) \cdot 0 = 0$ and a is a root of g. Consequently, we have the following.

Proposition 1.2
Let a be an element of R and let g be any element of $R[X]$. Then a is a root of g if and only if g is divisible by $(X - a)$.

Now for any integer $k > 0$, the polynomial $(X - a)^k$ is a monic polynomial of degree k and hence is a regular element in $R[X]$. Therefore, for any nonzero element h of $R[X]$, the product $(X - a)^k h$ is of degree greater than or equal to k. Hence, given a polynomial g in $R[X]$ having a as a root, there is a largest positive integer ν such that $(X - a)^\nu$ divides g.

Definition
The element $a \in R$ is a **root** of the polynomial g of **multiplicity** ν if ν is the largest positive integer such that $(X - a)^\nu$ divides g.

Suppose that a_1 and a_2 are distinct elements of R which are roots of g of multiplicities ν_1 and ν_2, respectively. Then, because $(X - a_i)^{\nu_i}$ divides g for $i = 1, 2$, we have $g = (X - a_1)^{\nu_1} h_1 = (X - a_2)^{\nu_2} h_2$. If $a_2 - a_1$ is not a zero divisor, then we have $0 = g(a_2) = (a_2 - a_1)^{\nu_1} h_1(a_2)$, so that $h_1(a_2) = 0$, and a_2 is therefore a root of h_1 of multiplicity $k > 0$. Clearly, $k \leq \nu_2$, and we have $h_1 = (X - a_2)^k \bar{h}_1$. Hence, $g = (X - a_1)^{\nu_1}(X - a_2)^k \bar{h}_1 = (X - a_2)^{\nu_2} h_2 = (X - a_2)^k (X - a_2)^{\nu_2 - k} h_2$. Because $(X - a_2)^k$ is regular in $R[X]$, we have $(X - a_1)^{\nu_1} \bar{h}_1 = (X - a_2)^{\nu_2 - k} h_2$. Consequently, if $\nu_2 - k > 0$, we have $(a_2 - a_1)^{\nu_1} \bar{h}_1(a_2) = (a_2 - a_2)^{\nu_2 - k} h_2(a_2) = 0$ and thus $\bar{h}_1(a_2) = 0$. But this contradicts the fact that a_2 is a root of h_1 of multiplicity k so that we must have $\nu_2 - k = 0$. Hence, $g = (X - a_1)^{\nu_1}(X - a_2)^{\nu_2} h_1$. Notice, too, that $\nu_1 + \nu_2$ must be less than or equal to the degree of g.

Extending the argument given above, the reader can now prove the following.

Proposition 1.3
Let R be an integral domain, let $a_1, \ldots, a_s$ be distinct elements of R, and let g be an element of $R[X]$ having each of the elements a_i as a root with multiplicity ν_i. Then $g = (X - a_1)^{\nu_1} \cdots (X - a_s)^{\nu_s} h$ where $h \in R[X]$. Moreover, $\Sigma \nu_i \leq \deg g$ and no a_i is a root of h.

Corollary 1.4
If R is an integral domain and g is an element of $R[X]$ of degree n, then g can have no more than n roots.

The statement of Proposition 1.3 can be strengthened by omitting the hypothesis that R be an integral domain and insisting instead that for each pair of distinct indices (i, j), the element $a_i - a_j$ be regular. To see that one cannot do any better than that, consider the following.

Example 1.5 Let R be the ring $\mathbf{Z}/(6)$. The element $3X+3$ in $R[X]$ has the roots 1, 3, and 5 in R, each with multiplicity one (because $3X+3$ is a linear polynomial). Obviously, $3X+3$ is not divisable by $(X-1)(X-3)(X-5)$. Observe that $3-1=5-3=2$, and $5-1=4$, both of which are zero divisors.

So far we have been talking about polynomials in $R[X]$ which have roots in R. However, it need not happen that an element of $R[X]$ have a root in R. For example, if $\mathbf{Q}$ is the field of rational numbers, the element X^2-2 in $\mathbf{Q}[X]$ has no root in $\mathbf{Q}$, since $\sqrt{2}$ is irrational. As another example, if $R = \mathbf{Z}/(2)$, the field of integers modulo 2, the polynomial X^2+X+1 in $R[X]$ has no root in R. However, $\mathbf{Q}$ is contained in another field, namely the field $\mathbf{R}$ of real numbers, and therefore $\mathbf{Q}[X]$ is contained in $\mathbf{R}[X]$. The element X^2-2 of $\mathbf{Q}[X]$, considered as an element of $\mathbf{R}[X]$, has a root (in fact two roots) in $\mathbf{R}$. Can we find a field K containing $\mathbf{Z}/(2)$ such that the element X^2+X+1, considered as an element of $K[X]$, has a root in K?

Example 1.6 Consider a set K having four elements: $\{0, 1, a, b\}$, and make it a field by defining the following addition and multiplication tables (the element 0 will of course be the zero element, so we will not include it in our tables):

Add	1	a	b
1	0	b	a
a	b	0	1
b	a	1	0

Mult	1	a	b
1	1	a	b
a	a	b	1
b	b	1	a

The reader may verify that K is a field with these operations. Because $1+1=0$, we see that $\mathbf{Z}/(2)$ is a subfield of K [here we are of course denoting the elements of $\mathbf{Z}/(2)$ by 0 and 1]. The polynomial X^2+X+1 is an element of $K[X]$ and $a^2+a+1 = b+a+1 = 1+1 = 0$. Hence, X^2+X+1 has a root in K, namely the element a. In fact, the element b is also a root, and so all the roots of X^2+X+1 are in K. Although X^2+X+1 is an irreducible element of $\mathbf{Z}/(2)[X]$, it factors into linear factors in $K[X]$: $X^2+X+1 = (X-a)(X-b)$.

The construction of K in Example 1.6 seems terribly ad hoc but actually it is not. We will show that if K is any field and f any element of $K[X]$, there is a field

L containing K such that the element f, considered as an element of $L[X]$, has at least one root in L.

First of all, because K is a field, $K[X]$ is a PID. Thus, if $f \in K[X]$, f is a product of irreducible polynomials: $f = f_1 \cdots f_r$ with $f_i \in K[X]$. If we can find a field L containing K which contains a root, say, of f_1, then L will also contain a root of f because $f = f_1 \cdots f_i \cdots f_r$ in $L[X]$ as well as in $K[X]$ (although in $L[X]$ the polynomials $f_1, \ldots, f_r$ need no longer be irreducible). Hence, let us suppose that we have an irreducible polynomial f in $K[X]$ and see if we can find a field L containing K which contains a root of f.

Because f is an irreducible polynomial, and because $K[X]$ is a PID, the ideal (f) is a maximal ideal in $K[X]$. Hence, the ring $K[X]/(f)$ is a field which we shall denote by L. If we denote by $k: K[X] \to L$ the canonical ring morphism, and by $i: K \to K[X]$ the canonical "inclusion" morphism, the composition $ki: K \to L$ is a ring morphism. Because $ki(1) = 1$, the morphism ki is not zero, so that ki is a monomorphism (do not forget that K is also a field). Making the usual identification of K with its isomorphic image in L under the monomorphism ki, we may assume that L contains K.

We will now show that L contains a root of f. To this end, let $a \in L$ be the image of X under the morphism k, that is, $a = k(X)$. If $f = c_0 + c_1X + \cdots + c_nX^n$ with $c_i \in K$, then f, considered as an element in $L[X]$, is really $ki(c_0) + ki(c_1)X + \cdots + ki(c_n)X^n$. But then

$$\begin{aligned} f(a) &= ki(c_0) + ki(c_1)k(X) + \cdots + ki(c_n)k(X)^n \\ &= k(i(c_0) + i(c_1)X + \cdots + i(c_n)X^n \\ &= k(c_0 + c_1X + \cdots + c_nX^n) = k(f) = 0 \end{aligned}$$

because $f \in \operatorname{Ker}(k)$. This shows that L contains the root, a, of the polynomial f.

Hence, we see that given a polynomial $f \in K[X]$, we may find a field L_1 containing K and containing at least one root of f whether f is irreducible or not. If the polynomial f has degree n, we know it has no more than n roots in any field containing K. Hence, if $a_1, \ldots, a_s$ are all the roots of f contained in L_1, with multiplicities $\nu_1, \ldots, \nu_s$, we know that in $L_1[X]$ f may be written as a product $(X - a_1)^{\nu_1} \cdots (X - a_s)^{\nu_s} g_1 \cdots g_t$ where the g_i are irreducible polynomials in $L_1[X]$ without roots in L. By applying the same construction as before, we may construct a field L_2 containing L_1 which contains a root, say, of g_1. In $L_2[X]$, then, f has additional roots and we may proceed to construct a sequence of fields $K \subset L_1 \subset L_2 \subset \cdots$ containing more and more roots of f. However, because f has at most n roots, we arrive, after a finite number of steps, at a field L_r in which f factors completely as a product

$$(X - a_1)^{\nu_1} \cdots (X - a_s)^{\nu_s} (X - a_{s+1})^{\nu_{s+1}} \cdots (X - a_m)^{\nu_m} \cdot g$$

where g is a constant and $a_i \in L_r$ for $i = 1, \ldots, m$. In this case, we say that L_r contains all the roots of f.

Notice that our construction has been such that each L_{j+1} is a finite-dimensional vector space over L_j, as well as being a field. Notice, too, that each L_{j+1}, being a vector space over L_j, is also a vector space over K. The following fact tells us that each L_j is, in fact, a finite-dimensional vector space over K.

Proposition 1.7
Let $K \subset L \subset M$ be fields. Suppose that L is a finite-dimensional vector space over K, and that M is a finite-dimensional vector space over L. Then M is a finite-dimensional vector space over K. In fact, if we denote by $[L:K]$ the dimension of the vector space L over K, by $[M:L]$ the dimension of M over L, and by $[M:K]$ the dimension of M over K, then we have $[M:K] = [M:L][L:K]$.

PROOF: Let $\{x_1, \ldots, x_n\}$ be a basis for L over K, and let $\{y_1, \ldots, y_m\}$ be a basis for M over L. Then, because the elements x_i and y_j are all in M, we may consider the elements $x_i y_j$ in M. We shall show that the set $\{x_i y_j\}$ is a basis for M as a vector space over K. First we show that this set generates M over K. This is simple, for if $z \in M$, then $z = \Sigma l_j y_j$ with each $l_j \in L$. But, for each $j = 1, \ldots, m$, we have $l_j = \Sigma_{i=1}^{n} a_{ij} x_i$ with $a_{ij} \in K$. Hence, $z = \Sigma_{j=1}^{m} l_j y_j = \Sigma_{j=1}^{m} (\Sigma_{i=1}^{n} a_{ij} x_i) y_j = \Sigma a_{ij} x_i y_j$, and so $\{x_i y_j\}$ generates M over K.

If we have a linear combination $\Sigma b_{ij} x_i y_j = 0$ with $b_{ij} \in K$ then we have $\Sigma_{j=1}^{m} (\Sigma_{i=1}^{n} b_{ij} x_i) y_j = 0$. Because $\Sigma_{i=1}^{n} b_{ij} x_i \in L$ for each j, this tells us that $\Sigma_{i=1}^{n} b_{ij} x_i = 0$ for each j. However, because $b_{ij} \in K$, we must have $b_{ij} = 0$ for each i and j and so we are done.

To sum up this entire discussion, we may state the following.

Theorem 1.8
Let K be a field and let f be an element of $K[X]$. Then there is a field L containing K which is a finite-dimensional vector space over K, such that in $L[X]$ f factors into the product $c(X - a_1)^{r_1} \cdots (X - a_m)^{r_m}$ where the $a_i \in L$ and $c \in K$.

The only "new" fact is that the element c is in K. However, this follows from the fact that f is, after all, an element of $K[X]$ and the constant c, above, is merely the leading coefficient of f which is, therefore, in K.

In Example 1.6, it turned out that when we found a root of our given polynomial, we found all the roots. To show that this is not always the case, let us consider the following.

Example 1.9 Let $\mathbf{Q}$ be the rational numbers and let f in $\mathbf{Q}[X]$ be the polynomial $X^3 - 2$. Because we know that there is no rational number whose cube is 2, this polynomial is clearly irreducible in $\mathbf{Q}[X]$. However, we know from elementary function theory that the function $X^3 - 2$ has a real zero, because $X^3 - 2$ is a continuous function which takes on the value -2 at 0 and the value 6 at $X = 2$. Thus, the function has a real zero somewhere between 0 and 2. Let us denote the real zero of this function by $\sqrt[3]{2}$ and let us denote the smallest subfield of the reals which contains both $\mathbf{Q}$ and $\sqrt[3]{2}$ by L_1. The field L_1 then, is a subset of the reals, and in particular, L_1 contains no complex numbers of the form $a + bi$ with $b \neq 0$. Now L_1 is isomorphic to $\mathbf{Q}[x]/(x^3 - 2)$ for we may map $\mathbf{Q}[X]$ into L_1 by sending any polynomial $\Sigma a_i X^i$ in $\mathbf{Q}[X]$ to the element $\Sigma a_i (\sqrt[3]{2})^i$ in L_1. That $\Sigma a_i (\sqrt[3]{2})^i$ is indeed an element of L_1 comes from the fact that because the a_i are in $\mathbf{Q}$ they are in L_1; because $\sqrt[3]{2}$ is L_1, so is $(\sqrt[3]{2})^i$ for any $i \geq 0$; and because $a_i (\sqrt[3]{2})^i$ is in L_1, so also is $\Sigma a_i (\sqrt[3]{2})^i$. Because $X^3 - 2$ gets sent to zero in L_1, the kernel of the morphism $\mathbf{Q}[X] \to L_1$ contains the ideal generated by $X^3 - 2$ which is prime, because $X^3 - 2$ is irreducible. Therefore, the ideal $(X^3 - 2)$ in $\mathbf{Q}[X]$ is maximal, because $\mathbf{Q}[X]$ is a

PID. If the kernel of the morphism were bigger than the ideal generated by X^3-2, it would have to be all of $\mathbf{Q}[X]$ which is impossible because 1 is sent to 1 in L_1. Thus, the image of $\mathbf{Q}[X]$ in L_1 is isomorphic to the field $\mathbf{Q}[X]/(X^3-2)$. Because this image contains both $\mathbf{Q}$ and $\sqrt[3]{2}$ it must be all of L_1 because L_1 is the smallest subfield of the reals containing both $\mathbf{Q}$ and $\sqrt[3]{2}$.

Note that L_1 is a three-dimensional vector space over $\mathbf{Q}$ and also that every element of L_1 is uniquely expressible as a sum: $a_0+a_1\sqrt[3]{2}+a_2\sqrt[3]{4}$ with $a_i\in\mathbf{Q}$. More importantly, observe that L_1 cannot contain any other root of X^3-2 besides $\sqrt[3]{2}$. For if it did, the function X^3-2 would have more than one real zero (because every element of L_1 is a real number), and we know that the graph of X^3-2 crosses the x-axis at only one point. In fact, in $L_1[X]$ the polynomial X^3-2 factors into $(X-\sqrt[3]{2})(X^2+\sqrt[3]{2}X+\sqrt[3]{4})$ and this second factor is irreducible in $L_1[X]$.

Now we could find a field L_2 containing a root of $X^2+\sqrt[3]{2}X+\sqrt[3]{4}$ and containing L_1 simply by considering $L_1[X]/(X^2+\sqrt[3]{2}X+\sqrt[3]{4})$. However, let us consider the complex number $z=-1/2+(\sqrt{3}/2)i$ (where $i=\sqrt{-1}$). Then $z^2=-1/2-(\sqrt{3}/2)i$ as the reader can easily verify, and $z^3=1$. If we take the complex number $z\sqrt[3]{2}$, we have $(z\sqrt[3]{2})^3=z^3(\sqrt[3]{2})^3=1.2=2$. Therefore, $z\sqrt[3]{2}$ is a complex root of X^3-2. Similarly, $z^2\sqrt[3]{2}$ is another complex root of X^3-2. (This, of course, was to be expected since $z^2\sqrt[3]{2}$ is the complex conjugate of $z\sqrt[3]{2}$.) Both $z\sqrt[3]{2}$ and $z^2\sqrt[3]{2}$ are roots of the polynomial $X^2+\sqrt[3]{2}X+\sqrt[3]{4}$ as can be easily verified using the fact that $z^2+z+1=0$.

Let L_2 be the smallest subfield of the complex numbers containing L_1 and $z\sqrt[3]{2}$. Then an argument, similar to the preceding one which showed that L_1 and $\mathbf{Q}[X]/(X^3-2)$ were isomorphic, shows that L_2 is isomorphic to $L_1[X]/(X^2+\sqrt[3]{2}X+\sqrt[3]{4})$. The field L_2 contains $z\sqrt[3]{2}$ and $\sqrt[3]{2}$. Therefore, it contains $(z\sqrt[3]{2})\times\sqrt[3]{2})^{-1}=z$ and hence also z^2 and $z^2\sqrt[3]{2}$. Thus, L_2 now contains all the roots of X^3-2; that is, in $L_2[X]$ the polynomial X^3-2 factors into the product $(X-\sqrt[3]{2})(X-z\sqrt[3]{2})(X-z^2\sqrt[3]{2})$. Because L_2 is a two-dimensional vector space over L_1 and L_1 is a three-dimensional vector space over $\mathbf{Q}$, we know by Proposition 1.7 that L_2 is a six-dimensional vector space over $\mathbf{Q}$. Therefore, in order to factor the cubic X^3-2 into a product of linear factors, it seems to require going to an extension whose dimension over $\mathbf{Q}$ is 6.

The field L_2 was described as the smallest subfield of the complex numbers containing L_1 and $z\sqrt[3]{2}$. We saw that it must also contain z and $\sqrt[3]{2}$. The reader should show that L_2 may be described as the smallest subfield of the complex numbers containing z and $\sqrt[3]{2}$, as well as $\mathbf{Q}$. Also, because $z^3=1$, z is a root of $X^3-1=(X-1)(X^2+X+1)$ and hence of X^2+X+1. The reader should show that L_2 is isomorphic to $\mathbf{Q}[X,Y]/(X^2+X+1, Y^3-2)$. The isomorphism is set up by mapping $\mathbf{Q}[X. Y]$ into L sending X to z and Y to $\sqrt[3]{2}$.

2. ALGEBRAIC ELEMENTS

So far, we have been concentrating on a field K, a polynomial $f\in K[X]$, and we have been looking for fields L containing K having roots of f. Let us change our point of view now and consider a field K contained in a field L. Because L contains K, L is a vector space over K, but we do not insist now that L be a

finite-dimensional vector space over K. If a is any element of L, we may consider the smallest subring of L containing both K and a and denote it by $K[a]$. Clearly, this subring consists precisely of those elements of L of the form $\Sigma_{i=0}^{n} c_i a^i$ where $c_i \in K$. In other words, $K[a]$ is the image of the ring morphism $h: K[X] \to L$ defined by setting $h(\Sigma c_i X^i) = \Sigma c_i a^i$. Now there are two possibilities for h: Either it is a monomorphism or it has a nonzero prime ideal for a kernel. That the kernel has to be prime (zero or not) comes from the fact that the image of h, being a subring of a field, is an integral domain.

Definition
We say that the element $a \in L$ is **algebraic** over K if the kernel of the morphism h above is not zero. Otherwise, we say that a is a **transcendental over K.**

If a is algebraic over K, then the ring $K[a]$ is isomorphic to $K[X]/(f)$ where f is an irreducible polynomial. (We may, of course, choose f to be the unique **monic** polynomial that generates the kernel of h.) In this case, we see that $K[a]$ is not only a subring, but actually a subfield of L, because (f), being a nonzero prime ideal in $K[X]$, is maximal.

Definition
The unique monic irreducible polynomial f such that $K[a] \approx K[X]/(f)$ is called the **minimal polynomial** of a.

Basic Property 2.1
If $a \in L$ is algebraic over K and f the minimal polynomial of a in $K[X]$, then $K[a]$ is a vector space over K of dimension n, where $n = \deg f$. The proof of this is left to the reader.

If a is transcendental over K, then $K[a]$ is isomorphic to $K[X]$ and the smallest subfield of L containing $K[a]$ is the field of quotients of $K[a]$ which is obviously isomorphic to $K(X)$, that is, the field of rational functions of X. In this book, however, we will not be dealing much with transcendental elements.

Definition
If L is a field containing the field K, we say that L is an **algebraic extension of K**, if every element of L is algebraic over K. If L is an extension of K, we denote by $[L:K]$ the dimension of L as a vector space over K (finite or not), and call $[L:K]$ the **degree of the extension L over K.**

Basic Properties 2.2

(a) If K is a subfield of L and $a \in L$, then a is algebraic over K if and only if there is a polynomial $f \neq 0$ in $K[X]$ such that $f(a) = 0$.
(b) If $K \subset L \subset M$ are fields and $a \in M$ is algebraic over K, then a is algebraic over L.
(c) If $K \subset L$ are fields and L is a finite-dimensional vector space over K, then L is an algebraic extension of K.

(d) If $K \subset L$ are fields and $a, b \in L$ are algebraic over K, then $a \pm b$, ab, and ab^{-1} are algebraic over K.

(e) If $K \subset L$ are fields, then the set of elements of L which are algebraic over K is a subfield of L.

(f) If $K \subset L \subset M$ are fields and M is algebraic over K, then M is algebraic over L.

(g) If $K \subset L \subset M$ are fields with L an algebraic extension of K and M an algebraic extension of L, then M is an algebraic extension of K.

PROOF: **(a)** and **(b)** left as exercises.

(c) If L is a finite-dimensional vector space over K, say of dimension n, consider for any element a of L the set of $n+1$ elements: $\{1, a, a^2, \ldots, a^n\}$. Then this set is linearly dependent, so that we may find elements $c_0, c_1, \ldots, c_n \in K$, not all zero, such that $\sum_{i=0}^{n} c_i a^i = 0$. If we let $f \in K[X]$ be the polynomial $\sum_{i=0}^{n} c_i X^i$, then $f(a) = 0$, so a is algebraic over K.

(d) If a is algebraic over K, then $K \subset K[a] \subset L$, and $K[a]$ is a field with $[K[a]:K] < \infty$ [because $K[a] \approx K[X]/(f)$]. Because $b \in L$ is algebraic over K, b is algebraic over $K[a]$ so that $K[a][b]$ is finite-dimensional over $K[a]$. Thus, by Proposition 1.7 $K[a][b]$ is finite-dimensional over K and, by **(c)**, is algebraic over K. But $K[a][b]$ is a subfield of L containing a and b. Hence, it contains $a \pm b$, ab, and ab^{-1} all of which are therefore algebraic over K.

(e) and **(f)** are left as exercises.

(g) To see that M is algebraic over K, choose any element $b \in M$. Because M is algebraic over L, there is a polynomial $f = \sum_{i=0}^{n} a_i X^i \in L[X]$ such that $f(b) = 0$.

Because $a_i \in L$ and L is algebraic over K, we have $K[a_0]$ is algebraic over K and $[K[a_0]:K] < \infty$. Similarly, $[K[a_0, \ldots, a_{i+1}]:K[a_0, \ldots, a_i]] < \infty$ for $i = 1, \ldots, n-1$, where $K[a_0, \ldots, a_{i+1}]$ is defined inductively to be $K[a_0, \ldots, a_i][a_{i+1}]$. Thus, $[K[a_0, \ldots, a_n]:K] < \infty$ by repeated application of Proposition 1.7. But b is algebraic over $K[a_0, \ldots, a_n]$ because the polynomial $f = \sum_{i=0}^{n} a_i X^i$ is an element of $K[a_0, \ldots, a_n][X]$, and $f(b) = 0$. Thus, $[K[a_0, \ldots, a_n][b]:K[a_0, \ldots, a_n]] < \infty$ which, again by 1.7, tells us that $[K[a_0, \ldots, a_n][b]:K] < \infty$. Hence, $K[a_0, \ldots, a_n, b]$ is an algebraic extension of K. Because it contains b, b is algebraic over K, and we are done.

In the proofs of **(d)** and **(g)** we got involved with $K[a][b]$ and $K[a_0, \ldots, a_n]$. Let us introduce some terminology that will make it easier to talk about these fields.

Definition

Let $K \subset L$ be fields. We say that L is a **finitely generated extension of K** if L contains a finite number of elements $a_1, \ldots, a_n$ such that L is the smallest subfield of L containing K and the elements $a_1, \ldots, a_n$. In this case, we write $L = K(a_1, \ldots, a_n)$. L is called a **simple extension of K** if $L = K(a)$ for some $a \in L$.

If $L = K(a_1, \ldots, a_n)$, it is easy to see that L is the field of quotients of the image of the ring morphism $h: K[X_1, \ldots, X_n] \to L$ defined by setting $h(c) = c$ for $c \in K$ and $h(X_i) = a_i$ for $i = 1, \ldots, n$. It is clear that for any $a_1, \ldots, a_n \in L$, the smallest subring of L containing K and $a_1, \ldots, a_n$ is the image of h, and is denoted by $K[a_1, \ldots, a_n]$. This conforms with the notation we used in the proof of **(g)**

above. Moreover, when we said that $K[a_0, \ldots, a_{i+1}] = K[a_0, \ldots, a_i][a_{i+1}]$, we were merely using the fact that $K[X_0, \ldots, X_i][X_{i+1}]$ is isomorphic to $K[X_0, \ldots, X_{i+1}]$, which implies the above equality. If $a_1, \ldots, a_n$ are all algebraic over K, then $K[a_1, \ldots, a_n] = K(a_1, \ldots, a_n)$. For when $n = 1$, we have already observed this to be the case. Using induction on n, we have $K[a_1, \ldots, a_n] = K[a_1, \ldots, a_{n-1}][a_n] = K(a_1, \ldots, a_{n-1})[a_n] = K(a_1, \ldots, a_{n-1})(a_n)$ and this last field is easily seen to be $K(a_1, \ldots, a_n)$.

From this last observation we may conclude that if L is a finitely generated extension of K with $L = K(a_1, \ldots, a_n)$ and a_i algebraic over K for $i = 1, \ldots, n$, then $[L : K] < \infty$ and L is an algebraic extension of K. It is obvious, too, that if L is an extension of K of finite degree, then L is a finitely generated algebraic extension of K.

The reader may wonder if and when finitely generated extensions are simple extensions. In the next example we show this is not always the case.

Example 2.3 Let K be any field, let $K[X, Y]$ be the polynomial ring in two indeterminates over K, and let $L = K(X, Y)$ be the field of quotients of $K[X, Y]$. Then L is finitely generated (by two elements) over K. We claim that L is not a simple extension of K. For suppose there was an element $z \in L$ such that $L = K(z)$. Then, because $X \in L$, $X = f(z)/g(z)$ where f and g are polynomials over K in one variable. Letting $h(U, V) = Ug(V) - f(V)$, we see that h is a polynomial over K in two variables with $h(X, z) = 0$. Then $h(X, V)$ is a polynomial in $K(X)[V]$ of which z is a root.

Thus, z is algebraic over $K(X)$ and therefore L is algebraic over $K(X)$ because $L = K(z)$. But $Y \in L$, so Y is also algebraic over $K(X)$. This says that there is a nonzero polynomial $P \in K(X)[V]$ such that $P(Y) = 0$. However, this implies that the zero element of $K[X, Y]$ can be expressed as a polynomial over K, not all of whose coefficients are zero. Because this is absurd, we see that L is not a simple extension of K. (This example, by the way, is basic in understanding the notion of transcendence degree of transcendental extensions.)

Proposition 2.4
Let K be an infinite field and L be a finitely generated extension of K having the property that only finitely many extensions L' of K exist such that $K \subset L' \subset L$. Then L is a simple algebraic extension of K.

PROOF: Consider any two elements $a_1, a_2 \in L$. For each $c \in K$, we have the element $a_1 + ca_2 \in L$ and clearly $K \subset K(a_1 + ca_2) \subset K(a_1, a_2) \subset L$. Because K has infinitely many elements and because there are only finitely many fields between K and L, there must be two distinct elements $c_1, c_2 \in K$ such that $K(a_1 + c_1a_2) = K(a_1 + c_2a_2) = L'$. The field L' contains both $a_1 + c_1a_2$ and $a_1 + c_2a_2$. Therefore, it contains $a_1 + c_1a_2 - (a_1 + c_2a_2) = (c_1 - c_2)a_2$. However, $c_1 - c_2$ is a nonzero element of K, so that L' also contains a_2. But then $c_1a_2 \in L'$, so that $a_1 + c_1a_2 - c_1a_2 = a_1 \in L'$. Thus, L' contains both a_1 and a_2; so L' contains $K(a_1, a_2)$.

Therefore, we have shown that $K(a_1, a_2) = K(a_1 + c_1a_2) = L'$ and $K(a_1, a_2)$ is a simple extension of K. By induction, we can show that for any finite set of elements $\{a_1, \ldots, a_n\}$ of L, the field $K(a_1, \ldots, a_n)$ is a simple extension of K.

Hence, the field $K(a_1, \ldots, a_n)$ is a simple extension of K. Hence, the fact that L is a finitely generated extension of K implies that L is a simple extension of K.

That L must be an algebraic extension of K follows from the fact that a simple transcendental extension of K (whether K is finite or infinite) contains infinitely many subfields containing K (prove this!).

The reader will see in the exercises that the assumption that K be infinite is not necessary. However, the proof for finite K is quite different.

Proposition 2.4 naturally raises the question as to whether a converse statement is true. Let us therefore prove the following.

Proposition 2.5
If L is a simple algebraic extension of a field K, then there are only finitely many fields L' such that $K \subset L' \subset L$.

PROOF: Let $L = K(a)$ where a is algebraic over K. Then a is the root of an irreducible monic polynomial f in $K[X]$ of degree $n = [L : K]$. If $K \subset L' \subset L$, then a is the root of an irreducible monic polynomial g in $L'[X]$ and g divides f in $L'[X]$. If $b_0, \ldots, b_m$ are the coefficients of g in L', then $L' = K(b_0, \ldots, b_m)$ because obviously $[L : L'] = [L : K(b_0, \ldots, b_m)]$ and $K(b_0, \ldots, b_m) \subset L'$. This shows the intermediate fields L' (that is, those containing K but contained in L) are determined by the various possible irreducible factors of f in fields between K and L. Because f has only finitely many irreducible factors in $L[X]$, and because any irreducible factor of f in $L'[X]$ (where $K \subset L' \subset L$) must be a product of irreducible factors of f in $L[X]$, there are only finitely many fields L' between K and L.

Before giving our next example we introduce the following.

Definition
Let R be an arbitrary not necessarily commutative ring and $h : \mathbf{Z} \to R$ the unique ring morphism given by $h(n) = n \cdot 1$ for all n in $\mathbf{Z}$. The **characteristic** of R is $|c|$ where $(c) = \operatorname{Ker} h$.

Basic Properties 2.6
Suppose R is a ring of characteristic c.

(a) $cr = 0$ for all r in R.
(b) If c is a prime and r_1 and r_2 are elements of R such that $r_1 r_2 = r_2 r_1$, then $(r_1 + r_2)^{c^n} = r_1^{c^n} + r_2^{c^n}$ for all n in $\mathbf{N}$.
(c) If R is an integral domain, then c is a prime.

PROOF: (a) Left as an exercise.

(b) We prove this for $n = 1$ and leave the obvious induction to the reader. Because r_1 and r_2 commute we have $(r_1 + r_2)^c = \sum_{k=0}^{c} \binom{c}{k} r_1^k r_2^{c-k}$ where $\binom{c}{k}$ is the binomial coefficient. Because c is prime it is easy to show that c divides $\binom{c}{k}$ for $0 < k < c$. Hence, $(r_1 + r_2)^c = r_1^c + r_2^c$.

(c) Left as an exercise.

We now give an example of a finitely generated algebraic extension which is not simple.

Example 2.7 Let K be the field of rational functions in two variables X and Y over the field $\mathbf{Z}/2\mathbf{Z}$. K is then an infinite field of characteristic two. Let L' be the field $K[U]/(U^2 - X)$ and let $L = L'[V]/(V^2 - Y)$ where U and V are indeterminates. The reader should prove that $U^2 - X$ is irreducible in $K[U]$ and that $V^2 - Y$ is irreducible in $L'[V]$. It can also be readily checked that L is an algebraic extension of K of degree 4, that L contains $\sqrt{X}$ and $\sqrt{Y}$, and that $L = K(\sqrt{X}, \sqrt{Y})$. If L were a simple extension of K, it would have to contain just a finite number of intermediate fields L' between K and L. Because K is infinite, this would mean that for two distinct $c_1, c_2 \in K$, $K(\sqrt{X} + c_1\sqrt{Y}) = K(\sqrt{X} + c_2\sqrt{Y})$ which would imply (as we saw in Proposition 2.4) that $K(\sqrt{X}, \sqrt{Y}) = K(\sqrt{X} + c_1\sqrt{Y})$. We show, however, that $K(\sqrt{X} + c\sqrt{Y}) \neq K(\sqrt{X}, \sqrt{Y})$ for every $c \in K$. We do this by showing that $[K(\sqrt{X} + c\sqrt{Y}) : K] \leq 2$ for all $c \in K$. Because $[K(\sqrt{X}, \sqrt{Y}) : K] = 4$, we cannot have $K(\sqrt{X} + c\sqrt{Y}) = K(\sqrt{X}, \sqrt{Y})$. Because the characteristic of K is two, $(\sqrt{X} + c\sqrt{Y})^2 = X + c^2Y \in K$. Thus, $\sqrt{X} + c\sqrt{Y}$ is a root of a quadratic polynomial over K, and therefore we have the desired inequality.

We now put together some of this preliminary material that we have developed. In Theorem 1.8, we started with a field K, a monic polynomial f in $K[X]$, and we found a field L containing K such that $[L : K] < \infty$ and such that f factors into a product of linear polynomials in $L[X]$, say $f = (X - a_1)^{r_1} \cdots (X - a_m)^{r_m}$. Because $[L : K] < \infty$, we know that L is an algebraic extension of K. More important, though, the field $L' = K(a_1, \ldots, a_m)$ is an algebraic extension of K (contained in L) and L' is the smallest subfield of L containing K such that f is a product of linear factors in $L[X]$. L' is called a splitting field for f and, in fact, we have the following.

Definition
Let K be a field, and f an element of $K[X]$. **A splitting field of f** is a field L containing K having the property that f splits completely into linear factors in $L[X]$ and L contains no proper subfield with this property.

The reader should be able to prove the following with no difficulty.

Basic Properties 2.8

(a) For any field K and any $f \in K[X]$, there exists a splitting field for f.
(b) If $f \in K[X]$ and L is a splitting field of f, then L is a finite algebraic extension of K. In fact, if $a_1, \ldots, a_m$ are all the roots of f in L, then $L = K(a_1, \ldots, a_m)$.

3. MORPHISMS OF FIELDS

We have avoided listing one other basic property of splitting fields because we would first like to emphasize a basic question. The question is: How unique is a splitting field of a polynomial in $K[X]$? What we shall show is that if L_1 and L_2 are both splitting fields of a polynomial $f \in K[X]$, then there is a field

isomorphism $\sigma: L_1 \to L_2$ such that $\sigma(c) = c$ for all $c \in K$. To expect that there be only one such isomorphism would be a little too much. For, if $\tau: L_2 \to L_2$ is an automorphism of L_2 leaving K fixed, that is, such that $\tau(c) = c$ for all $c \in K$, then clearly $\tau\sigma: L_1 \to L_2$ would be another isomorphism of L_1 into L_2 such that $\tau\sigma(c) = c$ for all $c \in K$. Of course, one might well ask if automorphisms $\tau: L_2 \to L_2$, other than the identity, can exist if they must leave K fixed. These questions lead us naturally and immediately to the study of isomorphisms and automorphisms of fields.

Definition

Let L_1 and L_2 be fields containing a field K. A field morphism $\sigma: L_1 \to L_2$ is said to be a **morphism over K** if $\sigma(c) = c$ for all $c \in K$. Such a morphism is also said to be one which **leaves K fixed**.

Basic Properties 3.1

(a) If $\sigma: L_1 \to L_2$ is a morphism over K, then σ is a K-morphism of the K-vector space L_1 into the vector space L_2.

(b) If $\sigma: L_1 \to L_2$ is a morphism over K and $a \in L_1$ is a root of the polynomial $f \in K[X]$, then $\sigma(a) \in L_2$ is also a root of f.

The proofs of these two facts are left to the reader.

Example 3.2 Let $\mathbf{Q}$ be the field of rationals and let $L = \mathbf{Q}(\sqrt{2})$, that is, L is the smallest subfield of the reals containing $\sqrt{2}$ (and also $\mathbf{Q}$). Every element of L is uniquely of the form $a + b\sqrt{2}$ where $a, b \in \mathbf{Q}$. Define $\sigma: L \to L$ by $\sigma(a + b\sqrt{2}) = a - b\sqrt{2}$.

The reader should verify that σ is an automorphism of L over $\mathbf{Q}$. In fact, this automorphism and the identity on L are the only automorphisms of L leaving $\mathbf{Q}$ fixed. (Actually, any automorphism of L would automatically leave $\mathbf{Q}$ fixed.) The reason for this is quite clear. For suppose that $\tau: L \to L$ is an automorphism over $\mathbf{Q}$ other than the identity. Because τ is an automorphism over $\mathbf{Q}$ and because $\sqrt{2}$ is a root of $X^2 - 2$, $\tau(\sqrt{2})$ must also be a root of $X^2 - 2$ by Basic Properties 3.1 above. But the only roots of $X^2 - 2$ in L are $\sqrt{2}$ and $-\sqrt{2}$. If $\tau(\sqrt{2}) = \sqrt{2}$, then we would have (because τ leaves $\mathbf{Q}$ fixed) $\tau(a + b\sqrt{2}) = a + b\sqrt{2}$ and so τ would be the identity. If $\tau(\sqrt{2}) = -\sqrt{2}$, then $\tau(a + b\sqrt{2}) = a + b\tau(\sqrt{2}) = a - b\sqrt{2}$ and $\tau = \sigma$. Thus, we have only the two possibilities mentioned above. This example shows, by the way, that the splitting field L of $X^2 - 2$ does admit an automorphism over $\mathbf{Q}$ other than the identity.

Now let us suppose that we have two fields L_1 and L_2 containing K, and that $\sigma: L_1 \to L_2$ is a field morphism over K. Then we know that σ is a monomorphism because σ is not the zero morphism. The morphism σ may be extended to a morphism $\tilde{\sigma}: L_1[X] \to L_2[X]$ by defining $\tilde{\sigma}(a_0 + a_1X + \cdots + a_nX^n) = \sigma(a_0) + \sigma(a_1)X + \cdots + \sigma(a_n)X^n$ for every element $a_0 + \cdots + a_nX^n \in L_1[X]$. The morphism $\tilde{\sigma}$ leaves every element of $K[X]$ fixed. Thus, if a polynomial f in $K[X]$ is a product $f = f_1f_2$ in $L_1[X]$, we have $\tilde{\sigma}(f) = f = \tilde{\sigma}(f_1f_2) = \tilde{\sigma}(f_1)\tilde{\sigma}(f_2)$ in $L_2[X]$.

We can apply this to the case when L_1 and L_2 are both splitting fields of a polynomial $f \in K[X]$. For, if $\sigma: L_1 \to L_2$ is a morphism over K, then, because $f =$

$(X-a_1)^{r_1}\cdots(X-a_m)^{r_m}$ in $L_1[X]$, we have $f=\sigma(X-a_i)^{r_1}\cdots\sigma(X-a_m)^{r_m}=(X-\sigma(a_1))^{r_1}\cdots(X-\sigma(a_m))^{r_m}$ in $L_2(X)$. Thus, $K(\sigma(a_1),\ldots,\sigma(a_m))$ is a subfield of L_2 containing K in which f splits completely into linear factors and because L_2 was a splitting field of f, we must have $L_2=K(\sigma(a_1),\ldots,\sigma(a_m))$. But clearly $K(\sigma(a_1),\ldots,\sigma(a_m))$ is contained in $\operatorname{Im}\sigma\approx L_1$, so σ is an isomorphism of L_1 onto L_2. This shows that if we have a morphism $\sigma:L_1\to L_2$ over K, then σ must be an isomorphism when L_1 and L_2 are splitting fields of the same $f\in K[X]$. The trick then is to show that there is some morphism of L_1 into L_2 over K when L_1 and L_2 are splitting fields for a polynomial $f\in K[X]$.

To find a morphism $\sigma:L_1\to L_2$ over K, we first observe that $L_1=K(a_1,\ldots,a_m)$ where $f=(X-a_1)^{r_1}\cdots(X-a_m)^{r_m}$. We know that $K(a_1)=K[a_1]$ and that for each i we have $K(a_1,\ldots,a_{i+1})=K[a_1,\ldots,a_i][a_{i+1}]$. We shall show that if $\sigma_i:K[a_1,\ldots,a_i]\to L_2$ is a morphism over K, then σ_i may be extended to a morphism $\sigma_{i+1}:K[a_1,\ldots,a_{i+1}]\to L_2$ if $i<n$. Clearly, if σ_i is a morphism over K and if σ_{i+1} extends σ_i, then σ_{i+1} is also a morphism over K. Note that $\sigma_0:K\to L_2$ may be taken to be the inclusion. Hence, if we show how to go from σ_i to σ_{i+1}, we are done.

If $a_{i+1}\in K(a_1,\ldots,a_i)$, then $K(a_1,\ldots,a_{i+1})=K(a_1,\ldots,a_i)$ so we simply define $\sigma_{i+1}=\sigma_i$. (This would be the case, say, if $a_1=\sqrt{2}$ and $a_2=-\sqrt{2}$, as in Example 3.2.) Thus, we may assume that $a_{i+1}\notin K(a_1,\ldots,a_i)$. Letting $K_i=K(a_1,\ldots,a_i)$, we have $f=(X-a_1)^{r_1}\ldots(X-a_i)^{r_i}f_1\ldots f_t$ where $f_1,\ldots,f_t$ are irreducible polynomials in $K_i[X]$. Because $f(a_{i+1})=0$, it follows that a_{i+1} must be a root of one of the polynomials $f_1,\ldots,f_t$, and we may as well assume it is a root of f_1. Using the notation we introduced before, we have the morphism $\tilde{\sigma}_i:K_i[X]\to L_2[X]$ and thus we have, because σ_i is a morphism over K,

$$f=\tilde{\sigma}_i(f)=(X-\sigma_i(a_1))^{r_1}\cdots(X-\sigma_i(a_i))^{r_i}\tilde{\sigma}_i(f_1)\cdots\tilde{\sigma}_i(f_t)$$

This is a factorization of f in $L_2[X]$. However, because L_2 is a splitting field of f, we know that there is an element $b\in L_2$ which is a root of $\tilde{\sigma}_i(f_1)$ (because $L_2[X]$ is a UFD and the linear polynomials in $L_2[X]$ are irreducible elements of $L_2[X]$). Define a morphism $\sigma'_{i+1}:K_i[X]\to L_2$ by $\sigma'_{i+1}(g)=[\tilde{\sigma}_i(g)](b)$ for any $g\in K_i[X]$. The kernel of σ'_{i+1} contains f_1 because b is a root of $\tilde{\sigma}_i(f_1)$, and therefore $\operatorname{Ker}\sigma'_{i+1}$ contains the maximal ideal (f_1). However, $\sigma'_{i+1}(1)=1$ so $\operatorname{Ker}\sigma'_{i+1}\neq K_i[X]$, and thus $\operatorname{Ker}\sigma'_{i+1}$ is precisely the ideal (f_1). This tells us that the induced morphism $\sigma'_{i+1}:K_i[X]/(f_1)\to L_2$ is a monomorphism. On the other hand, from all of our previous discussion about finding roots of polynomials, we know that we have an isomorphism $\tau:K_i[X]/(f_1)\to K_i(a_{i+1})=K(a_1,\ldots,a_{i+1})$. The morphism $\sigma'_{i+1}\tau^{-1}:K(a_1,\ldots,a_{i+1})\to L_2$ is now easily seen to be a morphism of $K(a_1,\ldots,a_{i+1})\to L_2$ which extends σ_i, and we are done. We therefore have the following.

Theorem 3.3

If L_1 and L_2 are splitting fields of $f\in K[X]$, then there is an isomorphism $\sigma:L_1\to L_2$ over K. Thus, any two splitting fields of f are isomorphic over K.

If we look at the crucial step of the proof of Theorem 3.3 we see that we can guarantee the existence of an extension of $\sigma:K'\to L$ to $K'(a)\to L$ if the minimal

polynomial of a in $K'[X]$ goes over by $\sigma : K'[X] \to L[X]$ into a polynomial which has a root in L. Suppose, then, that we have a field L having the property that every element of $L[X]$ has a root in L. Then such a field should be a sort of "universal receiver" of morphisms. Let us try to make this a bit more precise.

Definition
A field L is said to be **algebraically closed** if every element of $L[X]$ has a root in L. A field L containing a field K is said to be an **algebraic closure of K** if L is an algebraic extension of K which is also algebraically closed.

Basic Properties 3.4

(a) A field L is algebraically closed if and only if the only irreducible elements of $L[X]$ are linear.
(b) If L is an algebraically closed field containing a field K, then the set of elements in L algebraic over K is an algebraic closure of K.
(c) If $K \subset K' \subset L$ are fields with K' an algebraic extension of K and L an algebraic closure of K', then L is an algebraic closure of K.

PROOF: (a) It is very easy and is left to the reader.

(b) We know by Basic Property 2.2 of algebraic extensions that the set $\bar{K}$ of elements of L algebraic over K is a field. It is also clearly an algebraic extension of K. What must be shown is that $\bar{K}$ is an algebraically closed field.

Suppose, then, that $f \in \bar{K}[X]$. Being an element of $\bar{K}[X]$, f is also an element of $L[X]$ and therefore f has a root, a, in L. This tells us that a is algebraic over $\bar{K}$ and therefore $\bar{K}(a)$ is algebraic over $\bar{K}$ which in turn is algebraic over K. By Basic Properties 2.2, $\bar{K}(a)$ is algebraic over K, so a itself is algebraic over K. Because a is an element of L which is algebraic over K, a is in $\bar{K}$. This shows that f actually has a root in $\bar{K}$, and hence $\bar{K}$ is algebraically closed.

(c) Left to the reader.

Example 3.5 Let $\mathbf{Q}$ be the field of rational numbers and let $\mathbf{C}$ be the field of complex numbers. We assume that the reader knows that $\mathbf{C}$ is algebraically closed. This is the Fundamental Theorem of Algebra, first proved by Gauss. We assume also that the reader knows that the numbers e and π are not algebraic over $\mathbf{Q}$. Thus, $\mathbf{C}$ is an algebraically closed field containing $\mathbf{Q}$ which is not an algebraic extension of $\mathbf{Q}$.

The set $\bar{\mathbf{Q}}$ of elements of $\mathbf{C}$ which are algebraic over $\mathbf{Q}$ provides us with an algebraic closure of $\mathbf{Q}$. If, in place of $\mathbf{Q}$, we were to consider the field $\mathbf{R}$ of real numbers, the field $\mathbf{C}$ would be an example of an algebraic closure of $\mathbf{R}$. For $\mathbf{C}$ is algebraically closed and is an algebraic extension of $\mathbf{R}$: $\mathbf{C} \approx \mathbf{R}[X]/(X^2+1)$.

We shall assume throughout the rest of this book that an algebraic closure exists for any field K. The reader may supply his own proof of this fact by doing the exercises at the end of this chapter.

The following proposition makes clear what we meant when we said that an algebraically closed field was a "universal receiver."

Proposition 3.6
Let $K \subset K_0 \subset K'$ be fields with K' an algebraic extension of K. Let L be an algebraically closed field containing K and let $\sigma_0 : K_0 \to L$ be a morphism over K. Then there is a morphism $\sigma : K' \to L$ such that σ restricted to K_0 is σ_0.

PROOF: Let $\mathcal{S}$ be the set of all pairs (E, τ) where E is a subfield of K' containing K_0 and $\tau : E \to L$ is a morphism extending σ_0. The set $\mathcal{S}$ is not empty because it contains (K_0, σ_0). We order $\mathcal{S}$ by setting $(E_1, \tau_1) \leq (E_2, \tau_2)$ if $E_1 \subset E_2$ and τ_2 is an extension of τ_1 to E_2. If $\{(E_i, \tau_i)\}$ is a totally ordered subset of $\mathcal{S}$, we let $E = UE_i$ and define $\tau : E \to L$ by setting $\tau(e) = \tau_i(e)$ if $e \in E_i$. The pair (E, τ) is in $\mathcal{S}$ and is an upper bound for $\{(E_i, \tau_i)\}$. Hence, by Zorn's lemma, we know that $\mathcal{S}$ contains a maximal element which we denote by (E_0, τ_0).

We claim that $E_0 = K'$. If not, we could find an element $a \in K'$, $a \notin E_0$. Because K' is algebraic over K, it is algebraic over E_0 and so a has a minimal polynomial $f \in E_0[X]$. Because L is algebraically closed, the polynomial $\tau_0(f) \in L[X]$ has a root in L, and therefore, by the argument used in the proof of Theorem 3.3, we know that τ_0 may be extended to a morphism $\tau_1 : E_0(a) \to L$. This contradicts the maximality of (E_0, τ_0), so that we must have $E_0 = K'$, and the proof of the proposition is complete.

Corollary 3.7
If L_1 and L_2 are algebraic closures of the field K, then there is an isomorphism $\sigma : L_1 \to L_2$ over K.

PROOF: The existence of a morphism $\sigma : L_1 \to L_2$ over K follows from Proposition 3.6 because L_1 is an algebraic extension of K and L_2 is algebraically closed. Because L_1 is algebraically closed and σ is a monomorphism, the image of σ is also algebraically closed. But the image of σ is contained in L_2 and therefore L_2 is algebraic over $\operatorname{Im} \sigma$ (because it is algebraic over K). This implies that $L_2 = \operatorname{Im} \sigma$, and we have our result.

Corollary 3.7 shows us that any two algebraic closures of K are isomorphic over K. We therefore usually talk about the algebraic closure of a field K and denote it by $\bar{K}$. Also, because every algebraic extension of K admits a monomorphism into $\bar{K}$ over K, we generally assume that all our algebraic extensions of our field K are contained in $\bar{K}$. Thus, if $K \subset L \subset \bar{K}$, we may talk about the isomorphisms of L into $\bar{K}$ over K and included among them is the inclusion of L into $\bar{K}$.

Proposition 3.8
Let L be an algebraic extension of K and let $\sigma : L \to L$ be an endomorphism over K. Then σ is an automorphism.

PROOF: Because σ is a monomorphism, we must simply prove that it is surjective. If $a \in L$, let f be the minimal polynomial of a in $K[X]$. To show that $a = \sigma(b)$ for some b we restrict our attention to the finitely generated extension L' of K which is the subfield generated by all roots of f which are contained in L.

Clearly, $a \in L'$ and $[L' : K] < \infty$. Because the endomorphism σ must carry roots of f into roots of f, it follows that σ restricted to L' carries L' into L'. But σ restricted to L' is still a monomorphism and, because L' is a finite-dimensional vector space over K, σ must be surjective. Thus, $a = \sigma(b)$ for some $b \in L'$ and σ is therefore surjective.

We use Proposition 3.8 to prove the following.

Proposition 3.9
Let $K \subset L \subset \bar{K}$ be fields and let $\sigma : L \to \bar{K}$ be a morphism over K. Then there exists an automorphism $\bar{\sigma} : \bar{K} \to \bar{K}$ which extends σ.

PROOF: By Proposition 3.6 we know that there is an endomorphism $\bar{\sigma} : \bar{K} \to \bar{K}$ extending the morphism $\sigma : L \to \bar{K}$. But Proposition 3.8 tells us that σ must be an automorphism.

We are now in a position to prove one of the most basic facts about isomorphisms of fields.

Theorem 3.10
Let $K \subset L \subset M \subset \bar{K}$ be fields where $\bar{K}$ is the algebraic closure of K. Let $\{\sigma_i\}$ be the set of distinct morphisms of L into $\bar{K}$ over K, and let $\{\tau_j\}$ be the set of distinct morphisms of M into $\bar{K}$ over L. For each index i, let $\bar{\sigma}_i$ be an automorphism of $\bar{K}$ which extends σ_i. (Although there may be many which extend σ_i choose only one for each i.) Define $\omega_{ij} : M \to \bar{K}$ to be the composition $\bar{\sigma}_i \tau_j$. Then $\{\omega_{ij}\}$ is the set of all distinct morphisms of M into $\bar{K}$ over K.

PROOF: First we show that if $\omega_{ij} = \omega_{i'j'}$, then $i = i'$ and $j = j'$. If $\omega_{ij} = \omega_{i'j'}$, then $\bar{\sigma}_{i'} \tau_{j'} = \bar{\sigma}_i \tau_j$. Hence, for any $b \in L$ we have $\sigma_i(b) = \bar{\sigma}_i(b) = \bar{\sigma}_i(\tau_j(b)) = \bar{\sigma}_i \tau_j(b) = \bar{\sigma}_{i'} \tau_{j'}(b) = \bar{\sigma}_{i'}(\tau_{j'}(b)) = \bar{\sigma}_{i'}(b) = \sigma_{i'}(b)$, so that $\sigma_i = \sigma_{i'}$ and hence $i = i'$. From the fact that $\bar{\sigma}_i \tau_j = \bar{\sigma}_i \tau_{j'}$, we may conclude that $\tau_j = \tau_{j'}$ and so $j = j'$.

Because it is clear that each ω_{ij} is a morphism of M into $\bar{K}$ over K, all that remains to be shown is that if $\omega : M \to \bar{K}$ is any morphism of M into $\bar{K}$ over K, then $\omega = \omega_{ij}$ for some pair (i, j). Given our $\omega : M \to \bar{K}$, the restriction of ω to L gives a morphism of L into $\bar{K}$ over K. Hence, $\omega | L = \sigma_i$ for some i. Consider the morphism $\bar{\sigma}_i^{-1} \omega : M \to \bar{K}$. It is easy to see that this morphism is τ_j for some j. This gives us $\omega = \bar{\sigma}_i \tau_j = \omega_{ij}$ and we are done.

4. SEPARABILITY

Definition
If $K \subset L \subset \bar{K}$, we denote by $[L : K]_s$ the cardinality of the set of morphisms of L into $\bar{K}$ over K. $[L : K]_s$ is called the **separable degree** of L over K.

We will study the separable degree of L over K when $[L : K] < \infty$. If $[L : K] < \infty$, then we know that L is a finitely generated extension of K, so that $L = K(a_1, \ldots, a_m)$. In order to determine all the isomorphisms of L into $\bar{K}$ it is sufficient (by Theorem 3.10) to study the morphisms of $K(a_1, \ldots, a_{i+1})$ into $\bar{K}$ over $K(a_1, \ldots, a_i)$. Because $\bar{K}$ is an algebraic closure of each of the fields

$K(a_1, \ldots, a_i)$, our problem boils down to studying the morphisms of a simple extension $K(a)$ of K into $\bar{K}$ over K.

We have seen that if $\sigma: K(a) \to \bar{K}$ is a morphism over K, then $\sigma(a)$ must be a root of the minimal polynomial f of a over K. Conversely, if $b \in \bar{K}$ is a root of f, then there exists a unique morphism $\sigma: K(a) \to \bar{K}$ over K which carries a into b. For, if $b \in \bar{K}$ is a root of f, then $K(b) \approx K[X]/(f)$. Because $K(a) \approx K[X]/(f)$, it is trivial to verify the above statement (and the reader is strongly urged to do so). As a result of this we see that the number of distinct morphisms of $K(a)$ into $\bar{K}$ over K is precisely equal to the number of distinct roots of f. Because the number of distinct roots of f is less than or equal to $\deg f$, while $[K(a):K] = \deg f$, we immediately have $[K(a):K]_s \leq [K(a):K]$.

Proposition 4.1
If $K \subset L \subset \bar{K}$ are fields and $[L:K] < \infty$, then $[L:K]_s \leq [L:K]$.

PROOF: Follows easily from previous discussion.

If a polynomial f in $K[X]$ of degree n does not have n distinct roots, then in $\bar{K}[X]$ we have $f = (X-a_1)^{\nu_1} \cdots (X-a_m)^{\nu_m}$ with at least one of the integers ν_i larger than 1, say $\nu_1 > 1$. Let us write $f = (X-a_1)^{\nu_1} g$ where g is an element of $\bar{K}[X]$.

Now if h is a polynomial over any field (in fact, over any commutative ring), say $h = c_0 + c_1 X + \cdots + C_n X^n$, then the derivative of h may be defined formally as $h' = c_1 + 2c_2 X + \cdots + nc_n X^{n-1}$. The usual rules for derivatives of sums and products hold, that is, $(h_1+h_2)' = h_1' + h_2'$ and $(h_1 h_2)' = h_1' h_2 + h_1 h_2'$. The reader may check these facts for himself if he has not already verified them. Applying differentiation to our polynomial f, we find that $f' = \nu_1 (X-a_1)^{\nu_1 - 1} g + (X-a_1)^{\nu_1} g' = (X-a_1)^{\nu_1 - 1}[\nu_1 g + (X-a_1)g']$. Because we are assuming $\nu_1 > 1$, we see that $\nu_1 - 1 > 0$ and so f' also has a_1 as a root. We therefore see that if f has a root a of multiplicity bigger than one, then f' also has a as a root. The reader should show that the converse is true: If f and f' have a common root a, then f has a as a root with multiplicity greater than 1.

Now let us suppose that f is an irreducible (monic) polynomial in $K[X]$ of degree n having fewer than n distinct roots. (What we are trying to do, after all, is find out under what conditions we may have $[K(a):K]_s < [K(a):K]$.) Then f has a root a of multiplicity $\nu > 1$, and f' also has a as a root. Because $K(a) \approx K[X]/(f)$ (with X being sent to a), we see that (f) is the ideal of polynomials in $K[X]$ having a as a root. Because a is a root of f', f' is in the ideal generated by f and hence is a multiple of f. But if $f' \neq 0$, this is absurd because $\deg f' \leq n-1$ while $\deg f = n$. The only way out of this absurdity is for f' to be the zero polynomial. Although this may seem absurd, the reader should consider the polynomial $f = X^2 - b$ in $K[X]$ where K is a field of characteristic two. Then $f' = 2X = 0$. Of course, if K has characteristic zero, it is impossible for a polynomial f of positive degree to have a zero derivative. So we see that for characteristic zero, we must always have $[K(a):K]_s = [K(a):K]$. Hence, if $K \subset L \subset \bar{K}$ and $[L:K] < \infty$, we have $[L:K]_s = [L:K]$ if K has characteristic zero.

Let us assume, then, that we are in characteristic $p > 0$, and that $f' = 0$. If $f = c_0 + c_1 X + c_2 X^2 + \cdots + X^n$, we have $f' = c_1 + 2c_2 X + \cdots + nX^{n-1} = 0$ which means that $ic_i = 0$ for $i = 1, \ldots, n (c_n = 1)$. If p does not divide i, this means that $c_i = 0$, so

in order that $f' = 0$ it is necessary (and sufficient) that $c_i = 0$ when $p \neq i$. Thus, $f = c_0 + c_p X^p + c_{2p} X^{2p} + \cdots + c_{up} X^{up}$ with $c_{up} = 1$, and we may write $f = g_1(X^p)$ where $g_1 \in K[X]$ is defined to be $g_1 = c_0 + c_p X + c_{2p} X^2 + \cdots + c_{up} X^u$. Clearly, g_1 is irreducible because f is, and it may or may not be the case that $g_1' = 0$.

If $g_1' = 0$, then we may write $g_1 = g_2(X^p)$ for some irreducible g_2, and so forth. If $g_1 = g_2(X^p)$, then $f = g_1(X^p) = g_2(X^{p^2})$ so, continuing in this way we finally arrive at a situation where $f = g(X^{p^u})$, g is an irreducible monic polynomial in $K[X]$, and $g' \neq 0$. In $\bar{K}[X]$, $g = (X - b_1) \cdots (X - b_m)$ with the b_i distinct, because we are assuming that $g' \neq 0$ and hence g cannot have multiple roots. Then, because $f = g(X^{p^u})$, we have $f = (X^{p^u} - b_1) \cdots (X^{p^u} - b_m)$. Because $b_i \in \bar{K}$, we can find unique a_i in $\bar{K}$ such that $a_i^{p^u} = b_i$ (namely the root of $X^{p^u} - b_i$). Observe that if $a_i^{p^u} = a_i'^{p^u}$, then $0 = a_i^{p^u} - a_i'^{p^u} = (a_i - a_i')^{p^u}$ so that $a_i = a_i'$. We therefore have $f = (X^{p^u} - a_1^{p^u}) \cdots (X^{p^u} - a_m^{p^u}) = (X - a_1)^{p^u} \cdots (X - a_m)^{p^u}$. This shows that f has m distinct roots each occurring with the same multiplicity p^u, and $\deg f = n = mp^u$. To sum up this entire discussion, we have the following.

Theorem 4.2
Let K be a field and let f be an irreducible monic polynomial in $K[X]$ of degree n. If the characteristic of K is zero, then f has n distinct roots $a_1, \ldots, a_n$ and $[K(a_i):K]_s = [K(a):K] = n$. If the characteristic of K is $p > 0$, then f has m distinct roots $a_1, \ldots, a_m$ each of multiplicity p^u where μ is a nonnegative integer (it is possible that $\mu = 0$), and we have $n = mp^u$. Hence, for $i = 1, \ldots, m$, we have $[K(a_i):K]_s = m$ and $p^u[K(a_i):K]_s = [K(a_i):K]$. In particular, it is always true that $[K(a_i):K]_s$ divides $[K(a_i):K]$.

Corollary 4.3
If $K \subset L \subset \bar{K}$ are fields and $[L:K] < \infty$, then $[L:K]_s$ divides $[L:K]$.

Example 4.4 Let $K = (\mathbf{Z}/2\mathbf{Z})(X)$, that is, K is the field of quotients of the polynomial ring in one indeterminate over the field $\mathbf{Z}/2\mathbf{Z}$. Then K has characteristic two. In $K[Y]$, we have the irreducible polynomial $f = Y^2 - X$. To show that $Y^2 - X$ is irreducible, the reader need only show that $\sqrt{X}$ is not in K. But because (X) is a prime ideal in $\mathbf{Z}/2\mathbf{Z}[X]$, the proof that $\sqrt{X}$ is not in K is exactly analogous to the proof that $\sqrt{p}$ is not a rational number if p is a prime number. Now $f' = 0$, so $f = g(Y^2)$ where $g = Y - X$. Thus, f has one root which we denote by $\sqrt{X}$ and this root has multiplicity 2. Consequently, $[K(\sqrt{X}):K]_s = 1$ while $[K(\sqrt{X}):K] = 2$.

Definition
If K is a field and $a \in \bar{K}$, we say that a **is separable over** K if $[K(a):K]_s = [K(a):K]$. If $K \subset L \subset \bar{K}$, we say that L is a **separable extension of** K if for every $a \in L$, a is separable over K. If $f \in K[X]$, we say that f is a **separable polynomial** if every irreducible factor of f in $K[X]$ has no multiple roots. The field K is **perfect** if every element of $K[X]$ is separable.

Basic Properties 4.5

(a) An element $a \in \bar{K}$ is separable over K if and only if its minimal polynomial in $K[X]$ is separable.

(b) If $K \subset L \subset \bar{K}$ and $a \in \bar{K}$ is separable over K, then a is separable over L.
(c) If $K \subset L \subset \bar{K}$ and $[L:K] < \infty$, then L is a separable extension of K if and only if $[L:K]_s = [L:K]$.
(d) If $\{a_i\}$ is a set of elements in K, then each a_i is separable over K if and only if $K(\{a_i\})$ is separable over K, where $K(\{a_i\})$ denotes the smallest subfield of $\bar{K}$ containing K and all the elements a_i.
(e) If $K \subset L \subset \bar{K}$, then the set L' of all elements of L which are separable over K is a separable extension of K.
(f) If $K \subset L \subset M \subset \bar{K}$, if L is a separable extension of K, and if M is a separable extension of L, then M is a separable extension of K.
(g) If $K \subset L \subset M \subset \bar{K}$ and if M is a separable extension of K, then M is a separable extension of L and L is a separable extension of K.
(h) If K is a field of characteristic zero, then K is perfect.
(i) If K is a perfect field, then every algebraic extension of K is perfect.
(j) Every finite field is perfect.

PROOF: The proof of **(a)** is trivial and is left to the reader, while the proof of **(b)** follows immediately from **(a)**.

(c) If $[L:K] < \infty$ and $[L:K]_s = [L:K]$, we must show that if $a \in L$, then $[K(a):K]_s = [K(a):K]$. But if $[K(a):K]_s < [K(a):K]$, we would have $[L:K]_s = [L:K(a)]_s[K(a):K]_s < [L:K(a)][K(a):K] = [L:K]$, which is a contradiction. Hence, every element of L is separable over K, and L is separable over K.

Conversely, suppose $[L:K] < \infty$ and that L is a separable extension of K. We want to show that $[L:K]_s = [L:K]$. Because $[L:K] < \infty$, $L = K(a_1, \ldots, a_m)$. Moreover, because each $a_i \in L$ is separable over K, each a_i is separable over $K(a_1, \ldots, a_{i-1})$ [where $K(a_1, \ldots, a_{i-1}) = K$ when $i = 1$]. Thus, $[L:K]_s = \prod_{i=1}^{m} [K(a_1, \ldots, a_i):K(a_1, \ldots, a_{i-1})]_s = \prod_{i=1}^{m} [K(a_1, \ldots, a_i):K(a_1, \ldots, a_{i-1})] = [L:K]$, and **(c)** is proven.

(d) We suppose that each $a_i \in \bar{K}$ is separable over K, and prove that $K(\{a_i\})$ is separable over K. Let $a \in K(\{a_i\})$. Then there is an integer m such that $a \in K(a_1, \ldots, a_m)$. If we know that $K(a_1, \ldots, a_m)$ is a separable extension of K, we will be done. But, as in the proof of **(c)**, we know that because each a_i is separable over K, we have

$$[K(a_1, \ldots, a_i):K(a_1, \ldots, a_{i-1})]_s = [K(a_1, \ldots, a_i):K(a_1, \ldots, a_{i-1})]$$

Thus, $[K(a_1, \ldots, a_m):K]_s = [K(a_1, \ldots, a_m):K]$ and by **(c)**, $K(a_1, \ldots, a_m)$ is a separable extension of K.

(e) Follows immediately from **(d)**. If we let $\{a_i\}$ be the set L' of all elements of L which are separable over K, then by **(d)** we know that $K(\{a_i\})$ is a separable extension of K. Because $K(\{a_i\}) \subset L$ and every element of $K(\{a_i\})$ is separable over K, we have $L' = K(\{a_i\})$.

(f) Let us first prove **(f)** when $[M:K] < \infty$. In that case, we know that $[M:L] < \infty$, $[L:K] < \infty$, and $[M:K] = [M:L][L:K]$. But our hypothesis tells us that $[M:L]_s = [M:L]$ and $[L:K]_s = [L:K]$ so that, because $[M:K]_s = [M:L]_s[L:K]_s$, we have $[M:K]_s = [M:K]$ and M is separable over K.

Now, without assuming $[M:K]<\infty$, let us take $a \in M$ and show that a is separable over K. Let f be the minimal polynomial of a in $L[X]$, with $f = b_0 + b_1X + \cdots + X^n$, $b_i \in L$. Then f is a separable polynomial and therefore a is also separable over $K(b_0, \ldots, b_{n-1})$. Because $b_i \in L$, each b_i is separable over K so that, by (d) we know that $K(b_0, \ldots, b_{n-1})$ is separable over K. We are now in the situation $K \subset K(b_0, \ldots, b_{n-1}) \subset K(b_0, \ldots, b_{n-1}, a)$ with $[K(b_0, \ldots, b_{n-1}, a):K] < \infty$, and each field separable over the preceding one. Thus, $K(b_0, \ldots, b_{n-1}, a)$ is separable over K and, because it contains a, a is separable over K. This proves (f).

The proof of (g) follows trivially from (b). The proof of (h) is trivial.

(i) First observe that a field K is perfect if and only if every monic irreducible polynomial in $K[X]$ is separable. Thus, if a field K is perfect, every algebraic extension of K is a separable extension of K and, if every simple algebraic extension of K is a separable extension, K is perfect.

Now let L be an algebraic extension of K, and assume that K is perfect. To show that L is perfect we must show that if M is a simple algebraic extension of L, then M is a separable extension of L. But if $M = L(a)$, then $M = K(L, a)$ is separable over K because K is perfect. But then by (g), M is a separable extension of L and we are done.

(j) If K is a finite field, then K contains $\mathbf{Z}/p\mathbf{Z}$ for some prime p (p = characteristic of K) and is a finite, hence algebraic extension of $\mathbf{Z}/p\mathbf{Z}$. Thus, if we show that $\mathbf{Z}/p\mathbf{Z}$ is perfect for every prime p, we will have our result by applying (i) above.

To see that $\mathbf{Z}/p\mathbf{Z}$ is perfect, we want to show that every monic irreducible polynomial in $\mathbf{Z}/p\mathbf{Z}[X]$ is separable. If we had a monic irreducible polynomial f that was not separable, we would have $f = g(X^p)$ for some $g \in \mathbf{Z}/p\mathbf{Z}[X]$. Let us therefore show that any polynomial $f \in \mathbf{Z}/p\mathbf{Z}[X]$ which is such that $f = g(X^p)$ cannot be irreducible.

Let $g = c_0 + c_1X + \cdots + c_nX^n$ and $f = c_0 + c_1X^p + \cdots + c_nX^{np}$. Now we know that $a = a^p$ for every $a \in \mathbf{Z}/p\mathbf{Z}$. (We may know this for several reasons. One is that the nonzero elements of $\mathbf{Z}/p\mathbf{Z}$ form a multiplicative group of order $p-1$. Thus, $a^{p-1} = 1$ for all $a \neq 0$. Hence, $a^p = a$ for $a \neq 0$. Clearly, though, $0^p = 0$, so $a^p = a$ for all $a \in \mathbf{Z}/p\mathbf{Z}$.) Thus, $c_i = c_i^p$ for each i, and we have $f = c_0^p + c_1^pX^p + \cdots + c_n^pX^{np} = (c_0 + c_1X + \cdots + c_nX^n)^p = g^p$. This shows that every irreducible f in $\mathbf{Z}/p\mathbf{Z}[X]$ must be separable and $\mathbf{Z}/p\mathbf{Z}$ is therefore perfect.

This last proof actually shows that any algebraic extension of $\mathbf{Z}/p\mathbf{Z}$ is perfect. This was why, in giving an example of an inseparable extension in Example 4.4, we had to go to a transcendental extension $K = \mathbf{Z}/2\mathbf{Z}(X)$ of $\mathbf{Z}/2\mathbf{Z}$ before we could find an inseparable polynomial.

5. GALOIS EXTENSIONS

Our discussion of separable extensions revolved around the problem of determining those algebraic extensions L of K which admitted many morphisms into $\bar{K}$ over K. If $L = K(a_1, \ldots, a_m)$, we saw that a morphism $\sigma: L \to \bar{K}$ over K was

completely determined by what it did to the elements a_i. In particular, a_i and $\sigma(a_i)$ had to be roots of the same irreducible polynomial in $K[X]$. Suppose the elements $a_i \in L$ had the property that all the roots of the minimal polynomial f_i of a_i were in L. Then for any $\sigma: L \to \bar{K}$ over K, $\sigma(a_i)$ would have to be an element of L so that σ would actually be an endomorphism, hence, an automorphism of L. Fields of this type are given a special name.

Definition
Let $K \subset L \subset \bar{K}$ be fields. L is called a **normal extension of K** if every morphism $\sigma: L \to \bar{K}$ over K is an automorphism of L. L is a **galois extension** of K if L is both a normal and separable extension of K.

Basic Properties 5.1

(a) If $K \subset L \subset M \subset \bar{K}$ are fields and if M is a normal extension of K, then M is a normal extension of L.
(b) If $K \subset L \subset \bar{K}$ are fields, then L is a normal extension of K if and only if every irreducible element of $K[X]$ that has a root in L splits into a product of linear factors in $L[X]$.

PROOF: (a) is trivial. The proof of (b) is not completely trivial, but we leave it to the reader.

Normal extensions are not as well behaved as algebraic extensions or separable extensions. For instance, if $K \subset L \subset M$ are fields, if L is an algebraic (separable) extension of K, and if M is an algebraic (separable) extension of L, then M is an algebraic (separable) extension of K. This is not so for normal extensions.

Example 5.2 Let $\mathbf{Q}$ be the field of rational numbers, and consider the polynomial $X^4 - 2$ in $\mathbf{Q}[X]$. The proof that this polynomial is irreducible in $\mathbf{Q}[X]$ is outlined in the exercises. Now $\mathbf{Q}(\sqrt[4]{2})$, that is, the smallest subfield of the reals containing $\mathbf{Q}$ and a root of $X^4 - 2$, is an extension of degree 4 over $\mathbf{Q}$. However, $\mathbf{Q}(\sqrt[4]{2})$, being a subfield of the reals, does not contain all the roots of $X^4 - 2$. For instance, it does not contain $\pm i\sqrt[4]{2}$ where $i^2 = -1$. Therefore, $\mathbf{Q}(\sqrt[4]{2})$ is not a normal extension of $\mathbf{Q}$. However, $\mathbf{Q}(\sqrt[4]{2})$ contains $\mathbf{Q}(\sqrt{2})$ and we have the inclusions: $\mathbf{Q} \subset \mathbf{Q}(\sqrt{2}) \subset \mathbf{Q}(\sqrt[4]{2})$. Because $[\mathbf{Q}(\sqrt{2}) : \mathbf{Q}] = 2$ and $[\mathbf{Q}(\sqrt[4]{2}) : \mathbf{Q}(\sqrt{2})] = 2$, each of the fields is a normal extension of the preceding one [all quadratic extensions are necessarily normal (why?)]. But this does not force $\mathbf{Q}(\sqrt[4]{2})$ to be normal over $\mathbf{Q}$.

If L is a normal extension of K, then the set of automorphisms of L over K has $[L:K]_s$ elements, and is a group. If L is a galois extension of K and $[L:K] < \infty$, we then have $[L:K]_s = [L:K]$, so that the group of automorphisms of L over K (or K-automorphisms of L) is a group whose order is the degree of the extension L over K.

Definition
If L is a galois extension of K, the group of K-automorphisms of L is called the **galois group of L over K** and is denoted by $G(L/K)$.

Proposition 5.3
Let L be an algebraic extension of K, with $L \subset \bar{K}$. The intersection M of all normal extensions of K in $\bar{K}$ containing L is a normal extension of K. If $[L:K] < \infty$, then $[M:K] < \infty$. Finally, if L is a separable extension of K, then M is a galois extension of K.

PROOF: The set of normal extensions of K in $\bar{K}$ containing L is not empty because $\bar{K}$ is such an extension. To see that M is a normal extension of K, we must show that if $\sigma : M \to \bar{K}$ is a morphism over K, then $\operatorname{Im} \sigma$ is contained in M. Extend σ to an automorphism $\bar{\sigma} : \bar{K} \to \bar{K}$ over K. If M' is a normal extension of K in $\bar{K}$ containing L, then $\bar{\sigma}|M' : M' \to \bar{K}$ is a morphism over K, so its image is contained in M'. Because $(\bar{\sigma}|M')|M = \sigma$, we see that $\operatorname{Im} \sigma$ is contained in M'. Hence, $\operatorname{Im} \sigma \cap M' = M$ and M is therefore a normal extension of K.

Before proceeding to the proof of the remaining two statements, we should get a good look at what M is really like.

Suppose that $L = K(\{a_i\})$ where the indexing set may or may not be finite, and let $\{\sigma_i\}$ be the set of all morphisms of L into $\bar{K}$ over K. We then have the set of elements $\{\sigma_i(a_j)\}$ in $\bar{K}$, and we may consider the field $L' = K(\{\sigma_i(a_j)\})$. Because the inclusion $L \to \bar{K}$ is among the set $\{\sigma_i\}$, we have $K \subset L \subset L' \subset \bar{K}$. What we shall show is that $M = L'$. If we do this, our remaining assertions about M will follow easily. For, if $[L:K] < \infty$, then we may assume that the set $\{a_i\}$ is finite. Moreover, because $[L:K]$, would then be finite, the set $\{\sigma_i(a_j)\}$ would be finite, and thus M would be a finitely generated algebraic extension of K. Hence, we would have $[M:K] < \infty$.

Finally, if we were given that L is a separable extension of K, we would know that each a_i is separable over K. Because the minimal polynomial $f_i \in K[X]$ of a_i is also the minimal polynomial of $\sigma_i(a_j)$ if $\sigma_i : L \to \bar{K}$ is a morphism over K, it follows that each of the elements $\sigma_i(a_j)$ is separable over K. Thus, $L' = K(\{\sigma_i(a_j)\})$ is a separable extension of K and so, therefore, is M.

In order to show that $L' = M$, where $L' = K(\{\sigma_i(a_j)\})$, it will suffice to show that L' is contained in M and that L' is a normal extension of K. Obviously, L' is contained in M.

To see that L' is a normal extension of K, we let $\tau : L' \to \bar{K}$ be a morphism over K, and we show that $\tau(\sigma_i(a_j)) \in L'$ for all i, j. This suffices to show that $\operatorname{Im} \tau \subset L'$ and hence that L' is normal over K. Now if f_j is the minimal polynomial of a_j over K, then $\sigma_i(a_j)$ is also a root of f_j. Because τ is a morphism over K, $\tau(\sigma_i(a_j))$ is a root of f_j. Thus, there is a morphism of $K(a_j)$ into $\bar{K}$ over K which carries a_j to $\tau(\sigma_i(a_j))$, and this morphism may be extended to a morphism over L into $\bar{K}$ over K. The morphism, then, must be one of our σ's, say it is σ_k. Hence, we have shown that $\tau(\sigma_i(a_j)) = \sigma_k(a_j)$ for some k, and therefore $\operatorname{Im} \tau$ is contained in L'. This proves that $L' = M$ and the proof of Proposition 5.3 is complete.

Consequently, we have the following.

Proposition 5.4
If K is a field and $f \in K[X]$ is an irreducible separable polynomial, then the splitting field of f is a galois extension of K of finite degree.

We will shortly be able to prove a converse to this proposition.

Suppose we consider the following situation: $K \subset L \subset M$ are fields, $[M:K] < \infty$, and M is a galois extension of K. We may let $G = G(M/K)$ be the galois group of M over K. Because M is a normal extension of K, it is a normal extension of L. Also, because M is a separable extension of K, it is a separable extension of L, and thus M is a galois extension of L. If we set $H = G(M/L)$ = the galois group of M over L, we see that H is a subgroup of G because H is the set of automorphisms of M over L. Because the order of H is $[M:L]$ and the order of G is $[M:K]$, we see that the index of H in G is $[L:K]$. We are now in a position to state and start proving the main theorem of galois theory. The theorem is the following.

Theorem 5.5
Let M be a galois extension of K, with $[M:K] < \infty$. Let $\mathcal{S}$ be the set of subgroups of $G = G(M/K)$, and let $\mathcal{T}$ be the set of fields L such that $K \subset L \subset M$. Define $\theta: \mathcal{T} \to \mathcal{S}$ by setting $\theta(L) = G(M/L)$ for $L \in \mathcal{T}$. Then:

(a) θ is a bijective map and $\theta(L_1) \subset \theta(L_2)$ if and only if $L_2 \subset L_1$.
(b) $\theta(L)$ is a normal subgroup of G if and only if L is a normal extension of K.
(c) If L is a normal extension of K, then L is a galois extension of K and the map $\rho: G \to G(L/K)$ defined by $\rho(\sigma) = \sigma|L$ induces an isomorphism $G/\theta(L) \approx G(L/K)$.
(d) If L_1 and L_2 are in $\mathcal{T}$, there exists a morphism $\sigma: L_1 \to L_2$ over K if and only if there exists an element $\tau \in G$ such that $\tau x \tau^{-1} \in \theta(L_2)$ for all $x \in \theta(L_1)$.

The real heart of the theorem is part (a). The rest will be fairly straightforward and parts will be left as exercises for the reader.

In order to show that $\theta: \mathcal{T} \to \mathcal{S}$ is bijective, we will construct a map $\theta': \mathcal{S} \to \mathcal{T}$ and show that it is the inverse of θ. To construct a map θ', we have to associate to a subgroup H of G an intermediate field $\theta'(H): K \subset \theta'(H) \subset M$. Let us therefore study the following general problem.

Suppose we have a field M and a finite set H of automorphisms of M which form a group under composition. That is, if we let $\mathrm{Aut}(M)$ be the group of all automorphisms of the field M, we consider a finite subgroup H of $\mathrm{Aut}(M)$. Denote by M^H the subset of M consisting of all $x \in M$ such that $\sigma(x) = x$ for all $\sigma \in H$. The subset M^H of M is actually a subfield of M as can easily be verified by the reader.

Definition
The subfield M^H of M described above is called the **fixed field of H.**

Basic Property 5.6
If H is a group of automorphisms of a field M and if H' is a subgroup of H, then $M^H \subset M^{H'}$. The proof of this is obvious.

Retaining the notation that we used in Theorem 5.5, we are now in a position to define a map $\theta: \mathcal{S} \to \mathcal{T}$. If H is a subgroup of $G = G(M/K)$, then H is a finite group of automorphisms of M, and we may define $\theta'(H) = M^H$. Clearly, $M^H \subset M$. Equally clearly, $K \subset M^H$ for, because $H \subset G$, every element in H leaves every element of K fixed. Thus, $K \subset M^H \subset M$ and so $\theta'(H) \in \mathcal{T}$. Having defined θ',

what we would like to do is show that $\theta\theta'(H) = H$ for all $H \in \mathcal{S}$, and $\theta'\theta(L) = L$ for all $L \in \mathcal{T}$.

Clearly, we have $\theta\theta'(H) \supset H$ and $\theta'\theta(L) \supset L$. Without too much difficulty, we can prove that $\theta'\theta(L) = L$. To do this we prove the following.

Lemma 5.7
Let M be a separable extension of a field L, and let a be an element in M which is not in L. Then there is a morphism $\sigma: M \to \bar{L}$ over L such that $\sigma(a) \neq a$.

PROOF: Because M is separable over L, the element a is separable over L. Moreover, because $a \notin L$, $[L(a):L]_s \neq 1$, so there is a morphism $\sigma_0: L(a) \to \bar{L}$ over L which is not the inclusion, that is, $\sigma_0(a) \neq a$. Extending σ_0 to $\sigma: M \to \bar{L}$ we have the result.

As a result of Lemma 5.7 we see that if M is a galois extension of L and if H is a galois group of M over L, then $M^H = L$. For otherwise there would be some element $a \in M$ but not in L that was left fixed by every morphism $\sigma: M \to \bar{L}$ over L, and that is precisely what Lemma 5.7 says cannot happen. Consequently, we may conclude that $\theta'\theta(L) = L$ for all $L \in \mathcal{T}$, because $\theta(L)$ is the galois group of M over L and $\theta'(\theta(L))$ is the fixed field of this group, which must be L.

Because $\theta'\theta$ is the identity on $\mathcal{T}$, we see that θ' is injective. Thus, $\mathcal{T}$ must be a finite set because $\mathcal{S}$, being the set of subgroups of a finite group, is finite. Hence, we have the following interesting fact.

Proposition 5.8
Let K be an infinite field and L a separable extension of K of finite degree. Then L is a simple extension of K.

PROOF: By Proposition 5.3, we have $K \subset L \subset M$ where M is a galois extension of K of finite degree. From the fact that $\theta: \mathcal{T} \to \mathcal{S}$ is injective, we know that the set of fields containing K and contained in M is finite. Hence, the set of fields containing K and contained in L is certainly finite. By Proposition 2.4, then, L must be a simple extension of K.

The preceding proposition is true even when K is finite. However, as in 2.4, one has to provide a different (but still easy) proof. See the exercises.

Returning to Theorem 5.5 how can we prove that $\theta\theta'(H) = H$? We know that $\theta\theta'(H) \supset H$, so that the number of elements of $\theta\theta'(H)$ is greater than or equal to that of H. Because $\theta\theta'(H)$ is the galois group of M over $\theta'(H)$, we have $[M:\theta'(H)]$ is equal to the number of elements in $\theta\theta'(H)$. Therefore, if we can show that $[M:\theta'(H)]$ is less than or equal to the number of elements of H, we will be done.

PROOF OF THEOREM 5.5: Let $L = \theta'(H)$ and let $H = \{\sigma_1, \ldots, \sigma_n\}$, where $\sigma_1 = \mathrm{id}_M$. Notice that the morphisms $\tilde{\sigma}_j: M[X] \to M[X]$ are automorphisms of $M[X]$ and that the subring of $M[X]$ left fixed by all the $\tilde{\sigma}_i$ is just $L[X]$. Thus, if $a \in M$, and if we consider the polynomial $g = \prod_{i=1}^{n} (X - \sigma_i(a))$ in $M[X]$, we have $g \in L[X]$. For if we take any σ_j, we have $\tilde{\sigma}_j(g) = \Pi(X - \sigma_j\sigma_i(a))$ and, because the set

$\{\sigma_j\sigma_1, \ldots, \sigma_j\sigma_n\}$ is equal to the set $\{\sigma_1, \ldots, \sigma_n\}$, $\tilde{\sigma}_j(g) = g$ for $j = 1, \ldots, n$. As a result we see that if $a \in M$ and if $f \in L[X]$ is the minimal polynomial for a over L, then f divides $g = \prod_{i=1}^{n} (X - \sigma_i a)$ because $g(a) = 0$ and $g \in L[X]$. Because $\deg g = n$, we must have $\deg f \leq n$. Now suppose that K is infinite. Then L, too, is infinite and because M is a separable extension of L, we know that $M = L(a)$ for some $a \in M$.

But we have seen that the degree of the minimal polynomial of a over L cannot exceed n so that $[M : L] \leq n$. Thus, when K is infinite, we have shown that $[M : \theta'(H)]$ is less than or equal to the number of elements of H and therefore $\theta\theta'(H) = H$ and we are done.

The only place that we have used the fact that K was infinite was in asserting that M was a simple extension of L. Because the reader will prove for himself that this is still true when the fields are finite, we may consider the theorem to be proved in general, that is, θ is an isomorphism whose inverse is θ'.

The fact that $\theta(L_1) \subset \theta(L_2)$ if and only if $L_2 \subset L_1$ follows immediately from the bijectivity of θ. It is certainly clear that if $L_2 \subset L_1$, then $\theta(L_1) \subset (L_2)$. Now, if $\theta(L_1) \subset \theta(L_2)$, we have $L_1 = \theta'\theta(L_1) \supset \theta'\theta(L_2) = L_2$ and Theorem 5.5(a) is established.

To prove part (b), suppose first that H is a normal subgroup of G and let $L = \theta'(H)$. We want to show that L is a normal extension of K. If $a \in L$ and $\sigma' : L \to \bar{K}$ is a morphism over K, we want to show that $\sigma'(a) \in L$. This will show that every morphism $\sigma' : L \to \bar{K}$ over K is an endomorphism (hence, an automorphism) and that L is normal over K. Given $\sigma' : L \to \bar{K}$, we know that there is an automorphism $\bar{\sigma}' : \bar{K} \to \bar{K}$ over K that extends σ'. Restricting $\bar{\sigma}'$ to M, we obtain an automorphism σ of M over K which is an extension of σ'. Thus, for any $a \in L$, we have $\sigma(a) = \sigma'(a) \in M$. To show that $\sigma'(a) \in L$ it suffices to show that $\tau(\sigma'(a)) = \sigma'(a)$ for all $\tau \in H$. For this will show that $\sigma'(a) \in M^H = \theta'(H) = L$. Because H is assumed to be normal, we know that for every $\sigma \in G(M/K)$ and every $\tau \in H$, we have $\sigma^{-1}\tau\sigma = \tau'$ where $\tau' \in H$. In particular, then, for $a \in L$ we have $\sigma^{-1}\tau\sigma(a) = \tau'(a) = a$, so that $\tau\sigma'(a) = \sigma(a)$ for all $a \in L$. But $\sigma'(a) = \sigma(a)$ for $a \in L$, so $\tau\sigma'(a) = \sigma'(a)$ and thus $\sigma'(a) \in L$.

Conversely, suppose L is a normal extension of K, with $H = G(M/L)$. We want to show that H is a normal subgroup of $G = G(M/K)$ or that $\sigma^{-1}\tau\sigma \in H$ for all $\sigma \in G$ and all $\tau \in H$. To see that $\sigma^{-1}\tau\sigma \in H$, we must show that $\sigma^{-1}\tau\sigma(a) = a$ for all $a \in L$. Because L is assumed to be normal over K, $\sigma(a) \in L$ for all $\sigma \in G$. Thus, $\tau(\sigma(a)) = \sigma(a)$ for $\tau \in H$ and hence $\sigma^{-1}(\tau(\sigma(a))) = \sigma^{-1}(\sigma(a)) = a$. This proves (b).

To prove (c) we know that if $K \subset L \subset M$, then L is a separable extension of K. Hence, if L is also a normal extension of K, L is a galois extension and there is a galois group $G(L/K)$ consisting of all automorphisms of L over K. Thus the map $\rho : G \to G(L/K)$ defined by setting $\rho(\sigma) = \sigma|L$ is well defined and is clearly a group morphism.

We have already seen that if $\sigma' : L \to L$ is an automorphism of L, then there is an automorphism σ of M such that $\sigma|L = \sigma'$. Hence, ρ is surjective. Also, it is clear that $H \subset \operatorname{Ker} \rho$ because $\tau|L = \mathrm{id}_L$ for all $\tau \in H$. On the other hand, if $\sigma \in G$ is such that $\rho(\sigma) = \sigma|L$ is the identity on L, then $\sigma \in G(M/L) = H$. Hence, $H = \operatorname{Ker} \rho$ and the induced morphism $G/H \to G(L/K)$ is an isomorphism.

We leave the proof of (d) for the reader.

EXERCISES

(1) Let F be a finite field.
(a) Prove that F has characteristic $p \neq 0$, where p is a prime.
(b) Prove that F has p^n elements for some positive integer n.
(c) Prove that F is a splitting field for the polynomial $X^{p^n} - X$ in $\mathbf{Z}/p\mathbf{Z}[X]$.
(d) Prove that for any prime integer p and any positive integer n, there is a finite field of order p^n.
(e) Prove that two finite fields are isomorphic if and only if they have the same number of elements.
(2) Prove that if $K \subset L$ are finite fields, then L is a galois extension of K. Prove, therefore, that L is a simple extension of K.
(3) Let K be a finite field of order q and let L be a finite extension of K, say $[L:K] = n$.
(a) Prove that L has q^n elements.
(b) Prove that the galois group of L over K is cyclic and that it is generated by the automorphism $\sigma: L \to L$ over K defined by $\sigma(x) = x^q$.
(4) Let G be a group and K a field. Let $\sigma_1, \ldots, \sigma_n$ be distinct group morphisms from G to the multiplicative group of nonzero elements of K. Let $a_1, \ldots, a_n$ be elements of K such that $a_1\sigma_1(x) + \cdots + a_n\sigma_n(x) = 0$ for all x in G. Prove that $a_1 = \cdots = a_n = 0$. [*Hint*: Use induction on n.]
(5) Let $K \subset L$ be finite fields, and let σ be a generator of the galois group of L over K. Prove that there is an element a in L such that $\{a, \sigma(a), \ldots, \sigma^{n-1}(a)\}$ is a basis of L over K. [*Hint*: Using preceding exercise, show that the minimal polynomial $P(X)$ of the linear transformation $\sigma: L \to L$ must be equal to the characteristic polynomial of σ. Hence, show that, as a vector space, $L \approx K[X]/P(X)$.]
(6) Let K be an infinite field and let $f(X_1, \ldots, X_n)$ be an element of $K[X_1, \ldots, X_n]$. Prove that there exist elements $a_1, \ldots, a_n$ in K such that $f(a_1, \ldots, a_n) \neq 0$.
(7) Let $K \subset L$ be fields. Assume that there are only finitely many fields between K and L. Prove that L is a finitely generated extension of K. This shows that the assumption in Proposition 2.4 that L be a finitely generated extension of K is redundant.
(8) In this exercise, we outline a proof (following Artin) of the existence of an algebraic closure. Let K be a field.
(a) To each f in $K[X]$, with degree $(f) \geq 1$, associate the symbol X_f and let $R = K[\{X_f\}]$. Prove that the ideal in R generated by all elements of the form $f(X_f)$, with f in $K[X]$, is not the unit ideal. [*Hint*: Suppose $\Sigma_{i=1}^{n} g_i f_i(X_{f_i}) = 1$. Let L be an extension of K containing at least one root of each of the f_i. Show that this leads to the contradiction: $0 = 1$.]
(b) Let M be a maximal ideal of R containing the ideal generated by the elements $f(X_f)$, and let $E_1 = R/M$. Identify K with its image in E_1 via the morphism $K \to K[\{X_f\}] = R \to R/M = E_1$. Prove that every polynomial in $K[X]$ of degree at least one has a root in the field E_1.
(c) Let $E_1 \subset E_2 \subset \cdots$ be an increasing sequence of fields with the property that every polynomial in $E_k[X]$ of degree at least one, has a root in E_{k+1}. Prove that $E = \cup E_n$ is an algebraically closed field containing K.

(d) Prove that the set of elements in E which are algebraic over K is an algebraic closure of K.

(9) Let $\mathbf{Z}$ be a ring of integers and let p be a prime number. Prove that the polynomial $X^{p-1}+X^{p-2}+\cdots+1$ is an irreducible polynomial in $\mathbf{Z}[X]$. [*Hint*: Let $X=Y+1$ and show that $f(Y+1)=\Sigma_{k=0}^{p-1}\binom{p}{k}Y^{p-k-1}$.

(10) Let $\mathbf{C}^*$ be the multiplicative group of nonzero complex numbers, and let $\mathbf{Z}_m$ be the subset of $\mathbf{C}^*$ consisting of all roots of the polynomial X^m-1.

(a) Show that $\mathbf{Z}_m$ is a subgroup of $\mathbf{C}^*$ of order m.

(b) Prove that $\mathbf{Z}_m$ is cyclic. The elements of $\mathbf{Z}_m$ which generate $\mathbf{Z}_m$ are called **primitive mth roots of unity**.

(c) Show that if ζ is a primitive mth root of unity, then $\mathbf{Q}(\zeta)$ contains all mth roots of unity, where $\mathbf{Q}$ is the field of rational numbers.

(d) Prove that $\mathbf{Q}(\zeta)$ is a galois extension of $\mathbf{Q}$ whose galois group is isomorphic to a subgroup of the group of units of $\mathbf{Z}/m\mathbf{Z}$. [*Hint*: Show that if σ is an automorphism of $\mathbf{Q}(\zeta)$, then $\sigma(\zeta)=\zeta^t$ where t is an integer relatively prime to m.]

(11) Let $\zeta_1,\ldots,\zeta_t$ be the t distinct primitive mth roots of unity, where $t=\phi(m)$.

(a) Prove that $\mathbf{Q}(\zeta_i)=\mathbf{Q}(\zeta_j)$ for all $i,j=1,\ldots,t$.

(b) Letting K denote the field $\mathbf{Q}(\zeta_i)$ for any $i=1,\ldots,t$, show that the polynomial $\Phi_m(X)=\prod_{i=1}^{t}(X-\zeta_i)$ in $K[X]$ is actually in $\mathbf{Q}[X]$. [*Hint*: Show that every automorphism σ of K effects a permutation of the set $\{\zeta_1,\ldots,\zeta_t\}$, and thus $\bar{\sigma}:K[X]\to K[X]$ leaves $\Phi_m(X)$ invariant. The polynomial $\Phi_m(X)$ is called the **mth cyclotomic polynomial**. It has degree $\phi(m)$.]

(c) Prove that $X^m-1=\prod_{d/m}\Phi_d(X)$. Note that because each polynomial $\Phi_d(X)$ has degree $\phi(d)$, we retrieve our old formula: $m=\Sigma_{d/m}\phi(d)$. See the exercises of Chapter 13 for more information about the cyclotomic polynomials.

(12) (a) Assuming that the field of complex numbers is algebraically closed, prove that the only irreducible polynomials over the real numbers are linear or quadratic.

(b) Factor the polynomial x^4+1 over the real numbers.

(13) (a) Give several examples of nonnormal field extensions of the rational numbers. Find their least normal extensions and compute their galois groups.

(b) Give at least two examples of a field and an inseparable extension of that field.

(14) Let $K\subset L$ be fields with L an algebraic extension of K. The **separable closure** of K in L is the subset of L consisting of all elements which are separable over K. K is said to be **separately closed** in L if K is its separable closure in L.

(a) Show that the separable closure of K in L which we denote by K_S is a field extension which is separably closed in L.

(b) Show that K is separably closed in L if and only if either:

(i) $L=K$ or

(ii) characteristic of K is $p\neq 0$ and each element a in L has a minimal polynomial over K of the form $X^{p^n}-b$ for some nonnegative integer n, and b in K.

(c) Show that K is separably closed in L if and only if $L=K$ or characteristic of K is $p\neq 0$ and given a in L there is a nonnegative n such that a^{p^n} is in K.

(d) If $[L:K]<\infty$, then K is separably closed in L if and only if either:
  (i) $L=K$ or
  (ii) the characteristic of K is $p\neq 0$ and there is a positive integer n such that $L^{p^n}\subset K$ where L^{p^n} is the image of the ring morphism $L\to L$ given by $x\to x^{p^n}$ for all x in L.

(e) An algebraic field extension L of K is said to be **purely inseparable** if K is separably closed in L. Show that if $K\subset L\subset M$ are algebraic field extensions of K, then M is a purely inseparable extension of K if and only if L is a purely inseparable extension of K and M is a purely inseparable extension of L.

(f) Suppose $[L:K]<\infty$ and L is a purely inseparable extension of K. If $[L:K]\neq 1$, then $[L:K]=p^n$ for some positive integer where p is the characteristic of K.

(15) Let $L\subset K$ be an algebraic field extension of the field K. Let a be an element of L.

(a) Let $f(X)$ in $K[X]$ be the minimal polynomial of a over K. Show that a is separable over K if and only if the prime decomposition $P_1(X)^{n_1}\cdots P_s(X)^{n_s}$ of $f(X)$ in $L[X]$ has the property that all the $n_i=1$.

(b) Show that the following statements are equivalent.
  (i) a is separable over K.
  (ii) If M is a field extension of K, then $K[a]\underset{K}{\otimes}M$ is a semisimple ring.
  (iii) $K[a]\underset{K}{\otimes}K[a]$ is a semisimple ring. [*Hint*: Use the fact that if $f(X)$ is a polynomial in $K[X]$ and M is a commutative K-algebra, then the K-algebra $K[X]/f(X)K[X]\underset{K}{\otimes}M$ is isomorphic to the K-algebra $M[X]/f(X)M[X]$.]

(c) If $[L:K]<\infty$, then the following are equivalent:
  (i) L is a separable extension of K.
  (ii) If M is a field extension of K, then $L\underset{K}{\otimes}M$ is a semisimple ring.
  (iii) $L\underset{K}{\otimes}L$ is a semisimple ring. [*Hint*: Use the fact that if $f_1:\Lambda_1\to\Gamma_1$ and $f_2:\Lambda_2\to\Gamma_2$ are injective morphisms of K-algebras, then there is a unique morphism of K-algebras $f:\Lambda_1\underset{K}{\otimes}\Lambda_2\to\Gamma_1\underset{K}{\otimes}\Gamma_2$ with the property $f(x_1\otimes\lambda_2)=f(x_1)\otimes f(\lambda_2)$ and this unique morphism is injective.]

(d) Show that if $K(a)$ is a purely inseparable extension of K, then $K(a)\underset{K}{\otimes}K(a)$ is a local ring. Is the converse true?

(16) Let G be a group of automorphisms of the commutative ring L. Let $F(G)$ be the free L-module generated by G. Show that there is a unique map $F(G)\times F(G)\to F(G)$ satisfying $(xg, x'g')\mapsto xg(x')gg'$ for all x, x' in L and g, g' in G, which is a law of composition making the underlying abelian group of $F(G)$ a ring. We denote this ring, which is called the **twisted group ring of G over L**, by $L\{G\}$.

(a) Show that the map $L\to L\{G\}$ given by $x\to\Sigma_{g\in G}x_g g$ where $x_1=x$ and $x_g=0$ for $g\neq 1$ is an injective ring morphism. L is usually considered a subring of $L\{G\}$ by identifying each x in L with its image under the ring morphism $L\to L\{G\}$ just described.

(b) Show that L^G, the set of all x in L such that $g(x)=x$ for all g in G, is a subring of L which is contained in the center of $L\{G\}$.

(c) Let K be a subring of L^G.

(i) Show that for each $\Sigma_{g\in G} x_g g$ in $L\{G\}$ the map $f_{\Sigma_{g\in G} x_g g}: L \to L$ given by $f_{\Sigma_{g\in G} x_g g}(y) = \Sigma_{g\in G} x_g g(y)$ for all y in L, is a K-module morphism of L.

(ii) Show that the map $f: L\{G\} \to \mathrm{End}_K(L)$ given by $f(\Sigma_{g\in G} x_g g) = f_{\Sigma_{g\in G} x_g g}$ is a morphism of rings where $\mathrm{End}_K(L)$ is the endomorphism ring of the K-module L.

(d) Show that $f: L\{G\} \to \mathrm{End}_K(L)$ is also a K-algebra morphism where $L\{G\}$ and $\mathrm{End}_K(L)$ are considered K-algebras in the obvious way.

(e) Show that $f: L\{G\} \to \mathrm{End}_K(L)$ is injective if L is a field. [*Hint*: Use Exercise 4.]

(f) Suppose L and K are fields and $[L:K] < \infty$. Show that:

(i) G is a finite group of order at most $[L:K]$.

(ii) The following are equivalent.

(1) Order of G equals $[L:K]$.

(2) $f: L\{G\} \to \mathrm{End}_K(L)$ is an isomorphism.

(3) L is a galois extension of K with galois group G.

(17) Let L be an algebraic field extension of the field K. Suppose K_1 and K_2 are subfields of L which are finite galois extensions of K and K_1K_2 is the subfield of L generated by K_1 and K_2.

(a) Show that K_1K_2 is a finite field extension of K.

(b) Show that K_1K_2 is a galois extension of K.

(c) Let G be the galois group of K_1K_2 over K and G_1 and G_2 the galois groups of K_1 and K_2 over K, respectively. Then:

(i) The map $G \to G_1 \times G_2$ given by $\sigma \to (\sigma|K_1, \sigma|K_2)$ is an injective group morphism.

(ii) If $K_1 \cap K_2 = K$, then the injective group morphism $G \to G_1 \times G_2$ is an isomorphism.

(18) Let L be a purely inseparable algebraic extension of the field K.

(a) Show that there is a unique K-algebra morphism $f: L \underset{K}{\otimes} L \to L$ with the property $f(x \otimes y) = xy$ for all x, y in L.

(b) Show that f is surjective and Ker f is the ideal generated by $x \otimes 1 - 1 \otimes x$.

(c) Show that Ker f is a nilideal, that is, every element of Ker f is nilpotent.

(d) Show that $L \underset{K}{\otimes} L$ is a local ring with Ker f as its unique maximal ideal.

(19) Suppose L is a field extension of the field K with $[L:K] < \infty$ such that $L \underset{K}{\otimes} L$ is a local ring. Is L necessarily a purely inseparable extension of K?

(20) Let L be a field extension of the field K. Suppose we are given a family $\{x_i\}_{i\in I}$ of elements in L. Then there exists a unique K-algebra morphism $f: K[X_i]_{i\in I} \to L$ satisfying $f(X_i) = x_i$. The family of elements $\{x_i\}_{i\in I}$ are said to be **algebraically independent** over K if $f: K[\bar{X}_i]_{i\in I} \to L$ is injective.

(a) Show that an element x in L is algebraically independent over K if and only if it is transcendental over K.

(b) If $\{x_i\}_{i\in I}$ is a family of elements of L, we denote by $K[x_i]_{i\in I}$ the subring of L generated by K and the elements $\{x_i\}_{i\in I}$, and by $K(x_i)_{i\in I}$, the subfield of L generated by K and the elements $\{x_i\}_{i\in I}$.

(i) Show that $K(x_i)_{i\in I}$ is the field of quotients of $K[x_i]_{i\in I}$.

(ii) Show that if $\{x_i\}_{i\in I}$ is algebraically independent over K, then $K(x_i)_{i\in I}$ is isomorphic as a K-algebra to the field of quotients $K(X_i)_{i\in I}$ of $K[X_i]_{i\in I}$ which is a K-algebra in the obvious way.

(c) Suppose $\{x_i\}_{i\in I}$ is a family of elements in L. Show that the following statements are equivalent:
  (i) $\{x_i\}_{i\in I}$ is algebraically independent over K.
  (ii) If $I = I_1 \cup I_2$ is a partition of I, then $\{x_j\}_{j\in I_1}$ is algebraically independent over K and $\{x_j\}_{j\in I_2}$ is algebraically independent over $K(x_i)_{i\in I_1}$.
  (iii) There exists a partition $I = I_1 \cup I_2$ of I such that $\{x_j\}_{j\in I_1}$ is algebraically independent over K and $\{x_j\}_{j\in I_2}$ is algebraically independent over $K(x_j)_{j\in I_1}$.

(d) Suppose that $x_1, \ldots, x_n$ is a finite family of elements in L. Show that this family of elements is algebraically independent over K if and only if x_{i+1} is transcendental over $K(x_{i\in i})_{\{1,\ldots,i\}}$ for all $i = 0, \ldots, n-1$.

(21) Let L be a field extension of the field K. A family $\{x_i\}_{i\in I}$ is said to be a **transcendence basis** for L over K if $\{x_i\}_{i\in I}$ is algebraically independent over K and L is algebraic over $K(x_i)_{i\in I}$.

(a) Show that if $\{x_i\}_{i\in I}$ is algebraically independent over K, then there is a transcendence basis of K containing $\{x_i\}_{i\in I}$.

(b) Show that L has a transcendence basis over K.

(c) Suppose $\{x_i\}_{i\in I}$ is a family of elements of L such that $K(x_i)_{i\in I} = L$. Show that $\{x_i\}_{i\in I}$ contains a transcendence basis for L over K.

(22) Let L be a field extension of the field K which has a finite transcendence basis $\{x_1, \ldots, x_n\}$ with $n \geq 1$.

(a) Show that if ω is transcendental over K, then there are $n-1$ distinct elements, say $x_2, \ldots, x_n$, such that $\{\omega, x_2, \ldots, x_n\}$ is a transcendence basis for L over K.

(b) Proceed by induction on n to show that if $\{\omega_1, \ldots, \omega_m\}$ is an algebraically independent set over K, then $m \leq n$ and there are $n-m$ distinct elements say $x_{m+1}, \ldots, x_n$ such that $\{\omega_1, \ldots, \omega_m, x_{m+1}, \ldots, x_n\}$ is a transcendence basis of L over K.

(c) Show that all transcendence bases of L over K have the same number of elements, namely n, which is called the **degree of transcendence of L over K**.

(23) Let L be a field extension of the field K. Show that any two transcendence bases for L over K have the same cardinality.

Chapter 13 DEDEKIND DOMAINS

The commutative rings we have discussed until now have been fields and principal ideal domains. These rings are precisely the class of commutative rings with the property that submodules of free modules are free. We have already seen that projective modules are a natural generalization of free modules. Hence, the class of commutative rings with the property that every submodule of a projective module is projective is a natural generalization of the class of PID's. Such rings are called hereditary rings. We remind the reader that we have already shown in Chapter 10, Theorem 1.1 that a commutative ring is hereditary if and only if every ideal of the ring is projective.

In this chapter we study the most classical example of hereditary commutative rings, namely, hereditary integral domains. These rings are better known as Dedekind domains.

Historically, Dedekind domains arose in number theory as the ring of all integral elements in some finite algebraic extension of the rational numbers. Therefore, in this chapter, we also discuss integral extensions of PID's and, more generally, integral domains.

1. DEDEKIND DOMAINS

Definition
A ring R is called a **Dedekind domain** if all nonzero submodules of free R-modules are faithful and projective.

Basic Properties 1.1

(a) If R is a Dedekind domain, it is an integral domain.
(b) If R is a Dedekind domain, every ideal of R is projective, and every projective R-module is a sum of ideals of R.
(c) A commutative ring R is a Dedekind domain if and only if every nonzero submodule of a projective R-module is a sum of faithful projective ideals of R.
(d) R is a Dedekind domain if and only if R is an integral domain and every ideal of R is projective.
(e) A Dedekind domain is a noetherian integral domain.
(f) If R is a Dedekind domain and S is a multiplicative subset of R, then R_S is a Dedekind domain.
(g) If R is a Dedekind domain and $\mathfrak{P}$ is a prime ideal of R, then $R_{\mathfrak{P}}$ is a PID.

PROOF: (a) Left as an exercise.

(b) is just a restatement of Chapter 10, Theorem 1.1, as are Properties (c) and (d). Before proving (e) we establish the following.

Lemma 1.2

Let M be an R-module over an arbitrary ring R. Then M is a projective R-module if and only if there is a family $\{m_j\}_{j\in J}$ of elements in M and a family of R-morphisms $\{f_j : M \to R\}_{j\in J}$ such that for each m in M we have:

(a) $f_j(m) = 0$ for all but a finite number of $j \in J$.
(b) $m = \Sigma_{j\in J} f_j(m)m_j$.

PROOF: Suppose $h : F \to M$ is an epimorphism with F a free R-module with basis J. For each j in J we have the morphism $p_j : F \to R$ defined by $p_j(\Sigma_{k\in J} r_k k) = r_j$.

If M is projective, there is a splitting $s : M \to F$ for the epimorphism h. Let $m_j = h(j)$ and let $f_j : M \to R$ be the composition $p_j s$ for each j in J. The reader can check that the families $\{m_j\}_{j\in J}$ and $\{f_j : M \to R\}_{j\in J}$ satisfy (a) and (b).

Suppose that $\{m_j\}$ and $\{f_j : M \to R\}_{j\in J}$ satisfy (a) and (b). Clearly, $\{m_j\}_{j\in J}$ generates M. Hence, there is an epimorphism $h : F \to M$, where F is a free R-module generated by J, given by $h(j) = m_j$ for all j in J. Define $s : M \to F$ by $s(m) = \Sigma_{j\in J} f_j(m)j$ for all m in M. It is not hard to check that s is a splitting for h, and so M is projective because it is a summand of F.

PROOF OF BASIC PROPERTIES 1.1: Because R is a Dedekind domain, we know that every nonzero ideal of R is faithful and projective. Let I be a nonzero ideal of R. Then by Lemma 1.2 we know that there are elements $\{a_j\}$ in I and morphisms $\{f_j : I \to R\}$ such that for each $a \in I$, $a = \Sigma_j f_j(a)a_j$ with only finitely many of the elements $f_j(a)$ different from 0.

Notice that because each $f_j : I \to R$ is an R-morphism, we have $af_j(a') = f_j(aa') = f_j(a'a) = a'f_j(a)$ for every f_j and every pair of elements a, a' in I.

Because I is a nonzero ideal of R, I contains an element $a \neq 0$ and a is regular because R is an integral domain. Then $a = \Sigma f_j(a)a_j$ with only finitely many of the $f_j(a)$, say $f_{j_1}(a), \ldots, f_{j_n}(a)$ different from zero. Hence, we have $a = f_{j_1}(a)a_{j_1} + \cdots + f_{j_n}(a)a_{j_n}$. But because $f_j(a)a_j = af_j(a_j)$, we have $a = f_{j_1}(a)a_{j_1} + \cdots + f_{j_n}(a)a_{j_n} = a[f_{j_1}(a_{j_1}) + \cdots + f_{j_n}(a_{j_n})]$ so that $1 = f_{j_1}(a_{j_1}) + \cdots + f_{j_n}(a_{j_n})$.

From this it follows easily that the set $\{a_1, \ldots, a_n\}$ generates I. Thus, I is finitely generated, and this shows that R is noetherian.

(f) If R is a Dedekind domain and S is a multiplicative subset of R, then we know that R_S is an integral domain. Hence, to show that R_S is Dedekind, it suffices by (d) to show that every ideal of R_S is projective. By the results of Chapter 9, we know that if I' is an ideal of R_S, then $I' = I_S$ where I is an ideal of R. But then, because R is Dedekind, I is a projective R-module so that again by Chapter 9 we know that $I' = I_S$ is a projective R_S-module, and we are done.

(g) By (f), we know that $R_{\mathfrak{P}}$ is a Dedekind domain. Hence, $R_{\mathfrak{P}}$ is a local ring and every ideal is finitely generated and projective. Because every finitely projective module over a local ring is free, we know that every ideal of $R_{\mathfrak{P}}$ is free. This implies that every ideal of $R_{\mathfrak{P}}$ is principal and hence $R_{\mathfrak{P}}$ is a PID.

We have shown that Dedekind domains are pretty close to being PID's, because $R_{\mathfrak{P}}$ is a PID for every prime ideal $\mathfrak{P}$ of R, if R is Dedekind. In fact, if R is a local Dedekind domain, then R is a PID.

Another important class of Dedekind domains which are PID's are those with only a finite number of maximal ideals.

Definition
A **semilocal ring** is a commutative ring having only a finite number of maximal ideals.

Proposition 1.3
Let R be a semilocal integral domain. If M is a finitely generated projective R-module, then M is a free R-module.

PROOF: Let $\mathfrak{P}_1, \ldots, \mathfrak{P}_t$ be the distinct maximal ideals of R. Then $J = \mathfrak{P}_1 \cap \cdots \cap \mathfrak{P}_t$ is rad(R) and by the Chinese Remainder Theorem, $R/J \approx \coprod_{i=1}^{t} R/\mathfrak{P}_i$. In order to show that the finitely generated projective module M is free, it suffices, by Chapter 8, Basic Properties 3.6, to show that M/JM is a free R/J-module.

But $M/JM \cong R/J \underset{R}{\otimes} M \cong (\coprod_{i=1}^{t} R/\mathfrak{P}_i) \underset{R}{\otimes} M \cong \coprod_{i=1}^{t} (R/\mathfrak{P}_i \underset{R}{\otimes} M)$. Because $R/\mathfrak{P}_i$ is a field, $R/\mathfrak{P}_i \underset{R}{\otimes} M$ is a finite-dimensional vector space over $R/\mathfrak{P}_i$ for each $i = 1, \ldots, t$. If we show that all of these vector spaces have the same dimension s, it will follow, as the reader can show, that M/JM is a free R/J-module of rank s.

Let s be the dimension of the K-vector space $K \underset{R}{\otimes} M$ where K is the field of quotients of R. For each maximal ideal $\mathfrak{P}_i$ we know that $M_{\mathfrak{P}_i}$ is a free $R_{\mathfrak{P}_i}$-module. Because $K \underset{R_{\mathfrak{P}_i}}{\otimes} (R_{\mathfrak{P}_i} \underset{R}{\otimes} M) = K \underset{R}{\otimes} M$, we have that $s = \text{rank } M_{\mathfrak{P}_i}$. It is now easy to see that $s = \dim R/\mathfrak{P}_i \underset{R}{\otimes} M$ as a vector space over $R/\mathfrak{P}_i$ for each i.

As an immediate consequence of this proposition we have the following.

Corollary 1.4
If R is a semilocal Dedekind domain, then R is a PID.

So far, it looks as though we are heading toward a statement that Dedekind

domains are PID's. This is far from being the case, but we must develop a bit more machinery and prove a few more results before we can be fully convinced. In any event, we are in a position to prove the following.

Proposition 1.5
Let R be an integral domain. Then R is a Dedekind domain if and only if R is noetherian and $R_{\mathfrak{P}}$ is a PID for every prime ideal $\mathfrak{P}$ of R.

PROOF: If R is Dedekind, we know that R is noetherian and $R_{\mathfrak{P}}$ is a PID for every prime. Hence, we work on the converse, and we need only show that if I is any ideal of R, then I is projective. However, because we are assuming that R is noetherian, it suffices by Chapter 9, Theorem 6.4 to show that $R_{\mathfrak{P}} \underset{R}{\otimes} I$ is $R_{\mathfrak{P}}$-projective for every prime ideal $\mathfrak{P}$ of R. But because $R_{\mathfrak{P}}$ is a PID for every prime ideal $\mathfrak{P}$, and because $R_{\mathfrak{P}} \underset{R}{\otimes} I$ is an ideal of $R_{\mathfrak{P}}$, it follows that $R_{\mathfrak{P}} \underset{R}{\otimes} I$, being a principal ideal of $R_{\mathfrak{P}}$, is $R_{\mathfrak{P}}$-free. Thus I is projective and R is Dedekind.

From Proposition 1.5 we can deduce some results similar to those we know about PID's. For instance, suppose R is a PID, Q its field of quotients, and x an element of Q such that $x^2 \in R$. Then we can prove easily that x itself must be in R. For we may write $x = a/b$ with a and b relatively prime, and then we have $a^2/b^2 = c \in R$ or $a^2 = b^2 c$ so that a^2 divides c (because a and b are relatively prime). Thus, $c = a^2 c'$, $1 = b^2 c'$, and b is a unit. This shows that $a/b \in R$, so $x \in R$. As a special case, we know from this that $\sqrt{2}$ is not rational. For, by the above result, if it were rational it would have to be an integer, and we know that there is no integer whose square is 2. To see that similar results can be proven for Dedekind domains, consider the following.

Example 1.6 Let R be a Dedekind domain, Q its field of quotients, and $x \in Q$ such that $x^2 \in R$. Then $x \in R$. To see why this is so, we use what we have just seen about PID's. Because R is a domain, $R_{\mathfrak{P}}$ is contained in Q for every prime ideal $\mathfrak{P}$, and R is contained in $R_{\mathfrak{P}}$. Also, Q is the field of quotients of $R_{\mathfrak{P}}$. Thus, if $x \in Q$ and $x^2 \in R$, then $x^2 \in R_{\mathfrak{P}}$ for each $\mathfrak{P}$. Hence, $x \in R_{\mathfrak{P}}$ for each $\mathfrak{P}$ because each $R_{\mathfrak{P}}$ is a PID. Suppose $x = a/b$ with $a, b \in R$, $b \neq 0$. What we would like to show is that $a = rb$ for some $r \in R$, for then we would have $x = a/b = rb/b = r \in R$.

Consider, then, the set I of elements $y \in R$ such that $ya = rb$ for some $r \in R$, that is, $I = \{y \in R \mid ya \in (b)\}$. It can be easily shown that I is an ideal in R [in fact, it is usually denoted by $(b):(a)$]. If we can show that $I = R$, then $1 \in I$ and we have our result. But, if $I \neq R$, then I is contained in some prime ideal $\mathfrak{P}$. Because $b \in I$, we also have $b \in \mathfrak{P}$. However, we know that $x \in R_{\mathfrak{P}}$ also; thus, $x = a'/b'$ with $b' \in \mathfrak{P}$. This is absurd, for $a'/b' = a/b$ implies $a'b = ab'$ and hence, $b' \in I \subset \mathfrak{P}$. Therefore, $I = R$ and we have our result.

The reader might well ask whether what we have just done works for square roots only. That is, we have shown that if $x \in Q$ is a root of $x^2 - c$, where $c \in R$, then $x \in R$ provided R is a Dedekind domain. Could we equally well have done it for a cubic, a quartic, in fact for any polynomial with coefficients in R? When we say *any* polynomial, we have to be careful. For instance, if we consider the polynomial $4x^2 - 1$, the rational number $\frac{1}{2}$ is a root of it but $\frac{1}{2}$ is not an integer. However, the roots of $4x^2 - 1$ and of $x^2 - \frac{1}{4}$ are the same, but the latter is not a

polynomial all of whose coefficients are integers, while the former is not a monic polynomial. This suggests that we might restrict our inquiry to roots of monic polynomials all of whose coefficients are in our given ring R. Let us take a look at what this means if R is a PID.

Proposition 1.7
Let R be a PID with field of quotients Q, let $f \in R[X]$ be a monic polynomial, and let $a \in Q$ be such that $f(a) = 0$. Then $a \in R$.

PROOF: Suppose $f = X^n + c_1X^{n-1} + \cdots + c_n$ with $c_i \in R$, and suppose $a = u/v$ with $u, v \in R$, $v \neq 0$. Because $f(a) = 0$, we have $(u/v)^n + c_1(u/v)^{n-1} + \cdots + c_n = 0$. If we multiply through by v^n we get $u^n + (c_1v)u^{n-1} + \cdots + c_nv^n = 0$ or $v(-c_1u^{n-1} - cvu^{n-2} - \cdots - c_nv^{n-1}) = u^n$.

Assuming that u and v are relatively prime (which we may do because R is a PID and hence a UFD), we see that v divides u^n and therefore v is a unit.

The reader should observe that we really have used the property that R is a UFD rather than its being a PID to prove the proposition.

Example 1.6 suggests that we prove the following.

Corollary 1.8
Let R be a Dedekind domain, and Q its field of quotients. If $a \in Q$ is an element which is a root of a monic polynomial $f \in R[X]$, then $a \in R$.

PROOF: As in Example 1.6, we observe that for every prime $\mathfrak{P}$, we have $R \subset R_{\mathfrak{P}} \subset Q$ and thus a is a root of $f \in R_{\mathfrak{P}}[X]$ for every P. Because $R_{\mathfrak{P}}$ is a PID, Proposition 1.7 tells us that $a \in R_{\mathfrak{P}}$ for every P. Letting $a = u/v$ and $I = \{y \subset R | yu \in (v)\}$, we prove as in Example 1.6 that $I = R$ and thus $a \in R$.

Notice that the crucial step in both Examples 1.6 and 1.8 was the one that showed that if $x \in Q$ and $x \in R_P$ for every prime p, then $x \in R$. This generalizes to arbitrary integral domains as we point out in the following.

Proposition 1.9
If R is an arbitrary integral domain with quotient field Q, then $R = \cap R_{\mathfrak{P}}$ where $\mathfrak{P}$ ranges over all prime ideals of R.

2. INTEGRAL EXTENSIONS

So far we have been talking about elements a in the field of quotients Q of an integral domain R which are roots of monic polynomials in $R[X]$. However, there is no reason why we cannot consider a more general situation of, say, a ring R contained in another ring S, and talk about elements of S which are roots of monic polynomials in $R[X]$.

Definition
Let $R \subset S$ be commutative rings. An element $a \in S$ is said to be **integral over** R if a is the root of a monic polynomial in $R[X]$. We say that S is **integral over** R if every element of S is integral over R. We say that R is **integrally closed** in S if the

only elements of S which are integral over R are the elements of R. We say that an integral domain R is **integrally closed** if it is integrally closed in its field of quotients.

Basic Properties 2.1

(a) If R is a UFD, R is integrally closed.
(b) If R is a Dedekind domain, R is integrally closed.
(c) If $R \subset S$ and $a \in S$, then a is integral over R if and only if the subring $R[a]$ of S is a finitely generated R-module.
(d) If $R \subset S$ and $a \in S$, then a is integral over R if and only if there is a faithful $R[a]$-module M which is finitely generated as an R-module.
(e) If $R \subset S$ and $a_1, a_2 \in S$ are integral over R, then $a_1 \pm a_2$ and $a_1 a_2$ are integral over R.
(f) If $R \subset S$, then the set of elements in S which are integral over R is a subring of S containing R.
(g) If $R \subset S \subset T$ are rings with S integral over R and T integral over S, then T is integral over R.

PROOF: (a) and (b) have already been proven.

(c) and (d) We will prove that if a is integral over R, then $R[a]$ is a finitely generated R-module. Then we will show that if $R[a]$ is a finitely generated R-module, there is a faithful $R[a]$-module M which is finitely generated as an R-module. Finally, we will show that if there is a faithful $R[a]$-module which is finitely generated as an R-module, then a is integral over R. This will prove the equivalence of conditions (c) and (d) with the condition that a is integral over R.

Suppose that a is integral over R, and let $f \in R[X]$ be a monic polynomial such that $f(a) = 0$. $R[a]$ is the image of the morphism $R[X] \to S$ defined by sending $g \in R[X]$ to $g(a) \in S$. Consider any element $g \in R[X]$. From the fact that f is monic, we may write $g = qf + r$ with $\deg r < \deg f$ where q and r are in $R[X]$. Because $f(a) = 0$, we have $g(a) = q(a)f(a) + r(a) = r(a)$, and thus the image of g in S is equal to the image of r in S. This shows that $R[a]$ is the image of the set of polynomials in $R[X]$ whose degrees are less than the degree of f. This is clearly a finitely generated R-module.

Now suppose that $R[a]$ is a finitely generated R-module. Then $R[a]$ is certainly a faithful $R[a]$-module which is a finitely generated R-module, so we may choose M to be $R[a]$ itself.

Finally, suppose that we have a faithful $R[a]$-module M which is finitely generated as an R-module, and let $\{m_1, \ldots, m_n\}$ be a set of generators over R. Because M is an $R[a]$-module, multiplication by a on M makes sense and we have $am_i = \sum_{j=1}^{n} c_{ij} m_j$ with $c_{ij} \in R$ for $i = 1, \ldots, n$, $j = 1, \ldots, n$. We therefore have $\sum_{j=1}^{n} (a\delta_{ij} - c_{ij}) m_j = 0$ for $i = 1, \ldots, n$. Let F be the free R-module with basis $\{x_1, \ldots, x_n\}$, let $k : F \to M$ be the epimorphism defined by sending x_i to m_i for $i = 1, \ldots, n$, and let $g : F \to F$ be the endomorphism whose matrix with respect to the basis $\{x_1, \ldots, x_n\}$ is $(a\delta_{ij} - c_{ij})$. If we denote by L the cokernel of g, we have a unique morphism $k' : L \to M$ such that $k = k'k''$ where $k'' : F \to L$ is the canonical epimorphism. This is due to the fact that $kg = 0$. Hence, k' is an epimorphism and, because $|g|L = 0$ where $|g|$ is the determinant of g, we also have $|g|M = 0$. How-

ever, the determinant of g is an element of $R[a]$ and, because M is assumed to be a faithful $R[a]$-module, this determinant must be zero. If we let $f \in R[X]$ be the characteristic polynomial of the matrix (c_{ij}), that is, if we let $f = |X\delta_{ij} - c_{ij}|$, then we know that f is a monic polynomial in $R[X]$ and clearly $f(a) = |g| = 0$. Thus, we produced a monic polynomial in $R[X]$ of which a is a root, and a is thus integral over R.

(e) If a_1 and a_2 are in S and integral over R, then $R[a_1]$ is a finitely generated R-module and, because a_2 is also integral over $R[a_1]$, $R[a_1, a_2]$ is a finitely generated $R[a_1]$-module. Thus, $R[a_1, a_2]$ is a finitely generated R-module. Clearly, $R[a_1, a_2]$ is a faithful T-module for any subring T of $R[a_1, a_2]$. Hence, if b is any element of $R[a_1, a_2]$, then $R[a_1, a_2]$ is a faithful $R[b]$-module and is finitely generated as an R-module. Thus, every $b \in R[a_1, a_2]$ is integral over R. In particular, $a_1 \pm a_2$ and $a_1 a_2$ are integral over R.

(f) is an immediate corollary of (e).

(g) See comparable theorems for algebraic extensions (see Chapter 12, Basic Properties 2.2).

Example 2.2 Let $R = \mathbf{Z}$ and let $S = \mathbf{Q}(\sqrt{d})$ where d is a square free integer, that is, it has no square factors. Then every element x of S can be written uniquely as $x = a + b\sqrt{d}$. If $2a$ and $a^2 - b^2 d$ are integers, then x is a root of the monic polynomial $X^2 - 2aX + a^2 - b^2 d$. Hence, x is integral over $\mathbf{Z}$. We want to show now that if x is integral over $\mathbf{Z}$, then $2a$ and $a^2 - b^2 d$ are in $\mathbf{Z}$.

Because $[\mathbf{Q}(\sqrt{d}) : \mathbf{Q}] = 2$ we know that S is a galois extension of $\mathbf{Q}$ whose galois group consists of the automorphisms id, and $\sigma : S \to S$ where $\sigma(a + b\sqrt{d}) = a - b\sqrt{d}$. If $x = a + b\sqrt{d}$ in S is integral over $\mathbf{Z}$, then x is the root of a monic polynomial f in $\mathbf{Z}[X]$. Because $\sigma(x)$ is also a root of f, we have that $\sigma(x)$ is integral over $\mathbf{Z}$. Consequently, $x + \sigma(x)$ and $x\sigma(x)$ are integral over $\mathbf{Z}$. But $x + \sigma(x) = 2a$ and $x\sigma(x) = a^2 - b^2 d$. Therefore, $2a$ and $a^2 - b^2 d$ are elements of $\mathbf{Q}$ which are integral over Z if and only if $2a$ and $a^2 - b^2 d$ are integers. In connection with this example, we make the following.

Definition
If $R \subset S$ are commutative rings, the subring of S consisting of all elements in S integral over R is called the **integral closure of R in S.**

Basic Property 2.3
If $R \subset S \subset T$ are rings, if S' is the integral closure of R in S, and if T' is the integral closure of S' in T, then T' is the integral closure of R in T.

Theorem 2.4
Let R be an integrally closed noetherian integral domain, K its field of quotients, and L a separable extension of K with $[L : K] < \infty$. If S is the integral closure of R in L, then S is a finitely generated R-module. Hence, S is also a noetherian integrally closed integral domain.

PROOF: Because L is a separable extension of K, we know that L is a simple extension of K, that is, $L = K(a)$ for some $a \in L$. We claim that $a = a'/v$ with

$a' \in S$ and $v \in R$. To see this, let $f \in K[X]$ be the minimal polynomial for a so that

$$f(a) = a^n + \frac{u_1}{v} a^{n-1} + \cdots + \frac{u_n}{v} = 0$$

where $u_1, \ldots, u_n$, $v \in R$. Multiplying through by v^n we get $(va)^n + u_1(va)^{n-1} + (va)^n + u_1(va)^{n-1} + (u_2 v)(va)^{n-2} + \cdots + v^{n-1}u_n = 0$, so that va is a root of a monic polynomial in $R[X]$ and is therefore in S. Hence, $va = a' \in S$ and $a = a'/v$.

We next claim that $L = K(a')$. But this is clear because if $\{1, a, a^2, \ldots, a^{n-1}\}$ is a basis for L over K, then so is $\{1, a', \ldots, a'^{n-1}\}$. Therefore, we may assume that $L = K(a)$ with $a \in S$.

Next, we may assume that L is a galois extension of K. Because L is separable over K, we know that there is a galois extension M of K containing L, with $[M : K] < \infty$. Thus, we have $R \subset K \subset L \subset M$, and if we let T be the integral closure of R in M, we know that $R \subset S \subset T$. If we show that, for galois extensions, the integral closure is a finitely generated R-module, then we will know that T is finitely generated over R. However, because S is an R-submodule of T, it follows (because R is noetherian) that S, too, is a finitely generated R-module. Thus, we shall assume that L is a galois extension of K, and we let $G = \{\sigma_0, \ldots, \sigma_{n-1}\}$ be the galois group of L over K, with $\sigma_0 =$ identity.

Next, we define a map Tr: $L \to K$ by setting $\mathrm{Tr}(a) = \sum_{i=0}^{n-1} \sigma_i(a)$ for $a \in L$. Observe first, that for all $a \in L$, $\mathrm{Tr}(a)$ is indeed an element of K. For if we take any $\sigma_j \in G$ we have $\sigma_j(\mathrm{Tr}(a)) = \sigma_j(\sum_{i=0}^{n-1} \sigma_i(a)) = \sum_{i=0}^{n-1} \sigma_j\sigma_i(a) = \mathrm{Tr}(a)$ because $\{\sigma_j\sigma_i\}$ runs through G as i runs from 0 to $n-1$. Thus, $\mathrm{Tr}(a)$ is left fixed by each $\sigma_j \in G$ and must therefore be an element of K. Observe next that if $a \in S$, $\mathrm{Tr}(a) \in R$. For, if a is integral over R, then so is $\sigma_i(a)$ for $i = 0, \ldots, n-1$, and hence $\mathrm{Tr}(a) = \sum_{i=0}^{n-1} \sigma_i(a)$ is integral over R. But because $\mathrm{Tr}(a) \in K$ and R is integrally closed, we must have $\mathrm{Tr}(a) \in R$.

It is clear that $\mathrm{Tr}(a_1 + a_2) = \mathrm{Tr}(a_1) + \mathrm{Tr}(a_2)$ and that $\mathrm{Tr}(ca) = c\ \mathrm{Tr}(a)$ for all $c \in K$ and all $a \in L$. Thus, $\mathrm{Tr}: L \to K$ is a morphism of L into K as K-vector spaces.

Suppose that we can now show that the morphism Tr is not the zero morphism. Assuming this, we claim that we may establish an isomorphism (of K-vector spaces) $\delta : L \to (L, K)$ as follows: For each $a \in L$, define $\delta(a): L \to K$ by $\delta(a)(b) = \mathrm{Tr}(ab)$. The reader can check easily that $\delta(a): L \to K$ is indeed a K-morphism of L into K and that δ is also a morphism of K-vector spaces. Because L and (L, K) have the same finite dimension over K, it suffices to show that $\mathrm{Ker}\ \delta = 0$ if we want to show that δ is an isomorphism.

Suppose, then, that $\beta(a) = 0$ for some $a \in L$. This means that $\mathrm{Tr}(ab) = 0$ for all $b \in L$. Because we are assuming that Tr is not the zero morphism, we know there is some element $b_0 \in L$ such that $\mathrm{Tr}(b_0) \neq 0$. If $a \neq 0$, consider the element $b = a^{-1}b_0$. Then $\mathrm{Tr}(ab) = \mathrm{Tr}(a(a^{-1}b_0)) = \mathrm{Tr}(b_0) \neq 0$ which is a contradiction. Hence, $a = 0$, and we have shown that δ is an isomorphism because $\mathrm{Ker}\ \delta = 0$.

Using the fact that δ is an isomorphism, we can finish the proof of the theorem. For, given our basis $\{1, a, a^2, \ldots, a^{n-1}\}$ of L (with a, and hence a^i, in S), we have the dual basis $(\beta_0, \ldots, \beta_{n-1})$ for (L, K). If we let $b_i = \delta^{-1}(\beta_i)$ for $i = 0, \ldots, n-1$, then $\{b_0, \ldots, b_{n-1}\}$ is a basis for L over K. Because $\beta_i(a^j) = \delta_{ij}$ (by definition of dual basis), we have $\mathrm{Tr}(b_i a^j) = \delta(b_i)(a^j) = \beta_i(a^j) = \delta_{ij}$. If, now, x is any element of S, then we have seen that, because xa^i is also in S, $\mathrm{Tr}(xa^i) \in R$. Writing

$x = \Sigma_{j=0}^{n-1} c_j b_j$, we have $xa^i = \Sigma\, c_j b_j a^i$ and $\mathrm{Tr}(xa^i) = \mathrm{Tr}(\Sigma\, c_j b_j a^i) = \Sigma\, c_j \mathrm{Tr}(b_j a^i) = \Sigma\, c_j \delta_{ij} = c_i$. Thus, if $x \in S$, we have shown that $c_i \in R$ for $i = 0, \ldots, n-1$. This shows that S is contained in the R-module generated by $\{b_0, \ldots, b_{n-1}\}$. Because this is a finitely generated R-module, S is also a finitely generated R-module because R is noetherian.

Our whole proof now rests on the assumption we made that $\mathrm{Tr}: L \to K$ is not the zero morphism. Hence, we must show that $\mathrm{Tr}(b) \neq 0$ for some $b \in L$. We know that $\{1, a, \ldots, a^{n-1}\}$ is a basis for L over K, so to show that $\mathrm{Tr} \neq 0$, we must show that $\mathrm{Tr}(a^i) \neq 0$ for some $i = 0, \ldots, n-1$. Suppose this were not the case. Then for each $j = 0, \ldots, n-1$ we would have

$$\sigma_0(a^j) + \sigma_1(a^j) + \cdots + \sigma_{n-1}(a^j) = 0$$

or

$$\sigma_0(a)^j + \sigma_1(a)^j + \cdots + \sigma_{n-1}(a)^j = 0$$

If we take $L^n = L \times \cdots \times L$ (n times) and consider it as an L-vector space with the usual basis: $\{(1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}$, we have the endomorphism $g: L \to L$ whose corresponding matrix is $(\sigma_i(a)^j)$ where $i, j = 0, \ldots, n-1$. From the fact that $\Sigma_{i=0}^{n-1} \sigma_i(a)^j = 0$ for $j = 1, \ldots, n-1$, we see that $\mathrm{Ker}\, g$ contains the element $(1, 1, \ldots, 1)$ and thus $\mathrm{Ker}\, g \neq 0$. Hence, $|g| = 0$ because $|g|\,\mathrm{Ker}\, g = 0$. Therefore, assuming that $\mathrm{Tr}(a^i) = 0$ for $i = 0, \ldots, n-1$, we have arrived at the conclusion that $|g| = 0$. We must therefore show that we cannot have $|g| = 0$ to come up with a contradiction. For this we prove the following.

Lemma 2.5

Let R be any commutative ring, and let $\{a_0, \ldots, a_{n-1}\}$ be a set of elements of R. The determinant of the matrix $(b_{ij})(i = 0, \ldots, n-1;\ j = 0, \ldots, n-1)$ with $b_{ij} = (a_i)^j$ is the product $\prod_{n-1 \geq i > j \geq 0} (a_i - a_j)$. Thus, if R is an integral domain, this determinant is not zero if all elements a_i are distinct.

PROOF: We want to compute the determinant of the matrix

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_0^2 & a_1^2 & a_2^2 & \cdots & a_{n-1}^2 \\ \vdots & \vdots & \vdots & & \\ a_0^{n-1} & a_1^{n-1} & a_2^{n-1} & \cdots & a_{n-1}^{n-1} \end{pmatrix}$$

When $n = 2$, we have our result. Therefore, we use induction to prove Lemma 2.5 in general. Succesively multiplying each row of the matrix by a_0 and subtracting the result from the row below it, we get a matrix

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & a_1 - a_0 & \cdots & a_{n-1} - a_0 \\ 0 & a_1(a_1 - a_0) & \cdots & a_{n-1}(a_{n-1} - a_0) \\ \vdots & \vdots & & \vdots \\ 0 & a_1^{n-2}(a_1 - a_0) & \cdots & a_{n-1}^{n-2}(a_{n-1} - a_0) \end{pmatrix}$$

whose determinant is equal to that of the original matrix. Because the first column is zero from the second row on, the determinant of the above matrix is the determinant of the matrix

$$\begin{pmatrix} a_1-a_0 & a_2-a_0 & \cdots & a_{n-1}-a_0 \\ a_1(a_1-a_0) & a_2(a_2-a_0) & \cdots & a_{n-1}(a_{n-1}-a_0) \\ \vdots & \vdots & & \vdots \\ a_1^{n-2}(a_1-a_0) & a_2^{n-2}(a_2-a_0) & \cdots & a_{n-1}^{n-2}(a_{n-1}-a_0) \end{pmatrix}$$

Finally, the determinant of this matrix is easily seen to be

$$(a_1-a_0)\cdots(a_{n-1}-a_0)\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_{n-1} \\ \vdots & & & \\ a_1^{n-2} & a_2^{n-2} & \cdots & a_{n-1}^{n-2} \end{vmatrix}$$

and our induction hypothesis finishes the proof.

PROOF OF THEOREM 2.4, CONTINUED: Returning to our main theorem, we had concluded that if $\mathrm{Tr}=0$, then $|g|=0$ where g was a morphism whose matrix was $(\sigma_i(a)^j)$ for $i, j=0,\ldots,n-1$. Letting $a_i=\sigma_i(a)$, we see that we are precisely in the situation of Lemma 2.5. Thus, $|g|=\prod_{n-1\geq i>j\geq 0}(\sigma_i(a)-\sigma_j(a))$. But because L is galois over K, we know that $\sigma_i(a)\neq\sigma_j(a)$ for $i\neq j$ because the minimal polynomial of a is separable, that is, it has distinct roots. Thus, we cannot have $|g|=0$ and hence $\mathrm{Tr}: L\to K$ is not the zero morphism. This completes the proof that S is a finitely generated R-module.

Being a finitely generated module over the noetherian ring, it is obviously a noetherian module and hence a noetherian ring.

Finally, we show that S is integrally closed. We have already seen that if $a\in L$ then $a=a'/v$ with $v\in K$ and $a'\in S$. Thus, L is the field of quotients of S. If S were not integrally closed in L, then the integral closure of R in L would properly contain S which would be a contradiction. Thus, we have shown that S is an integrally closed noetherian integral domain, finitely generated as an R-module. This completes the proof of Theorem 2.4.

In the next section we will use this theorem to construct Dedekind domains which are not necessarily PID's.

3. CHARACTERIZATIONS OF DEDEKIND DOMAINS

Proposition 3.1

An integral domain R is a Dedekind domain if and only if R is noetherian, integrally closed, and has the property that every prime ideal other than (0) is maximal.

PROOF: If R is a Dedekind domain, we know it is noetherian and integrally closed. Also, we know that every $R_{\mathfrak{P}}$ is a PID for every prime ideal $\mathfrak{P}\neq 0$. Because of the correspondence between the prime ideals in a ring and its localizations, the fact that $R_{\mathfrak{P}}$ is a PID for all nonzero prime ideals $\mathfrak{P}$ of R, implies that every nonzero prime ideal of R is maximal.

Conversely, suppose R is noetherian, integrally closed, and that every prime ideal $\mathfrak{P}$ other than (0) is maximal. By Proposition 1.5 it suffices to prove that $R_{\mathfrak{P}}$ is a PID for every prime $\mathfrak{P}$. Now $R_{\mathfrak{P}}$ is a noetherian local domain, because R is a noetherian domain. We claim that $R_{\mathfrak{P}}$ is integrally closed. For if $a \in Q$ where Q is the field of quotients of R, and also of R_P, and if a is integral over R_P, we have

$$a^n + \frac{c_1}{s} a^{n-1} + \frac{c_2}{s} a^{n-2} + \cdots + \frac{c_n}{s} = 0$$

with c_i and s in R, and $s \notin P$. Multiplying through by s^n, we get $(sa)^n + sc_1(sa)^{n-1} + \cdots + s^{n-1}c_n = 0$ so that sa is integral over R. Because R is integrally closed, we must have $sa = a' \in R$, so $a = a'/s$ with $a' \in R$ and $s \notin P$. Hence, $a \in R_{\mathfrak{P}}$.

In addition to being integrally closed, $R_{\mathfrak{P}}$ contains no prime ideal other than (0) and $PR_{\mathfrak{P}}$. For if P' were a prime ideal of $R_{\mathfrak{P}}$ distinct from (0) and $PR_{\mathfrak{P}}$, then $P' \cap R$ would be a prime ideal of R different from (0) and properly contained in P. This would contradict the fact that nonzero prime ideals of R are maximal. Hence, $PR_{\mathfrak{P}}$ is the only prime ideal of $R_{\mathfrak{P}}$ other than (0). If we can now prove the following fact, we will be done.

Lemma 3.2
Let R be a local noetherian integrally closed domain whose only prime ideals are (0) and M where M is the maximal ideal of R. Then R is a PID.

PROOF: If $M = (0)$, then R is a field and we need not bother any more. We may therefore assume that $M \neq (0)$. What we propose to do is show that M is a principal ideal. From this it follows that every ideal of R is principal. For suppose that I is an ideal of R generated minimally by $\{u_1, \ldots, u_n\}$. Recall that this implies that if $\Sigma r_i u_i = 0$, then $r_i \in M$ for $i = 1, \ldots, n$. Now let F be a free R-module with basis $\{x_1, \ldots, x_n\}$ and let $g: F \to I$ be the epimorphism defined by setting $g(x_i) = u_i$. If $N = \operatorname{Ker} g$, we have the exact sequence $0 \to N \to F \xrightarrow{g} I \to 0$. If we can show that $N/MN = 0$, we will have $N = 0$ for since R is noetherian and F is finitely generated, N is finitely generated. Thus, $N/MN = 0$ implies $N = 0$.

To this end, consider an element $z = \Sigma r_i x_i \in N$. Then $0 = g(z) = \Sigma r_i g(x_i) = \Sigma r_i u_i$ so that $r_i \in M$. Because we are assuming that M is principal, we have $M = (a)$. Hence, $r_i = s_i a$ with $s_i \in R$, $i = 1, \ldots, n$. Then $z = a \Sigma s_i x_i$. But $0 = g(z) = g(a \Sigma s_i x_i) = ag(\Sigma s_i x_i)$ and because $g(\Sigma s_i x_i) \in I$ we must have either $a = 0$ or $g(\Sigma s_i x_i) = 0$. Our assumption that $M \neq 0$ implies that $g(\Sigma s_i x_i) = 0$ or that $\Sigma s_i x_i \in N$. Hence, $z = a \Sigma s_i x_i \in MN$, and we have shown that $N \subset MN$ or $N/MN = 0$. Because $N = (0)$, g is an isomorphism, which means that I is free. Because free ideals in a commutative ring are principal, we know that I is a principal ideal. Hence, to show that R is a PID it suffices to show that M is principal.

To prove that M is principal, consider any element $x \neq 0$ in M. If we let $S = \{x^n\}_{n \geq 0}$, then S is a multiplicative subset of R, and $R \subset R_S \subset Q$ where Q is the field of quotients of R and of R_S. But R_S must itself be a field. For otherwise, R_S contains a maximal ideal $\mathfrak{P}' \neq (0)$ and $\mathfrak{P}' \cap R$ is then a nonzero prime ideal of R. In that case, we would have $\mathfrak{P}' \cap R = M$ because M is the only other prime ideal contained in R. But because $x \in M$, $x \in \mathfrak{P}' \cap R$ so that $x \in \mathfrak{P}'$. This is absurd

because x is a unit in R_S and thus could not be contained in $\mathfrak{P}'$. Hence, R_S is a field and must therefore be all of Q.

We have thus shown that if z is any element of Q and x is any nonzero element of M, then $z = r/x^n$ with $r \in R$ and $n \geq 0$. Now let b be a fixed nonzero element of M and x any nonzero element of M. Then $1/b = r/x^n$ for some $r \in R$ and $n \geq 0$ so that we have $x^n = rb$. This shows that for each element x of M there is an integer $n \geq 0$ such that $x^n \in (b)$. From the fact that R is noetherian, we know that M is finitely generated, say by $\{x_1, \ldots, x_t\}$. Then $x_i^{n_i} \in (b)$ and hence, if we take n sufficiently large [say $n \geq (t-1)\max(n_i)$], we have $M^n \subset (b)$. Now let us take the smallest integer n such that $M^n \subset (b)$. Then M^{n-1} is not contained in (b), so we may find an element $c \in M^{n-1}$ with $c \notin (b)$. Because $cM \subset M^n \subset (b)$, we see that $(c/b)M \subset R$ and, because $c \notin (b)$, $c/b \notin R$.

Now $(c/b)M$ is clearly an ideal of R. Therefore, we either have $(c/b)M \subset M$ or $(c/b)M = R$. We claim that $(c/b)M$ cannot be contained in M. For if $z \in Q$ is such that $zM \subset M$, then M is an $R[z]$-module which is faithful as an $R[z]$-module and finitely generated as an R-module. Hence, z is integral over R (by Basic Properties 2.1) and, because we are assuming that R is integrally closed, $z \in R$. But if $z = c/b$, then $z \notin R$. Therefore, we have $(c/b)M \not\subset M$ and hence $(c/b)M = R$. However, R is a free R-module and multiplication by c/b is a monomorphism. Thus, M is isomorphic to R and is therefore a principal ideal. This completes the proof of the lemma and also the proof of Proposition 3.1.

Suppose, now, that R is a Dedekind domain, K the field of quotients of R, L a separable extension of K of finite degree, and S the integral closure of R in L. Then we know that S is a noetherian integrally closed domain. If we show that every nonzero prime ideal $\mathfrak{P}$ of S is maximal, then by Proposition 3.1 we know that S is Dedekind. The fact that every nonzero prime ideal of R is maximal is a special case of the following.

Proposition 3.3
Let $R \subset S$ be commutative rings with S integral over R, and let $\mathfrak{P}_1 \subset \mathfrak{P}_2$ be prime ideals in S such that $\mathfrak{P}_1 \cap R = \mathfrak{P}_2 \cap R$. Then $\mathfrak{P}_1 = \mathfrak{P}_2$.

PROOF: For the sake of convenience, we reduce the problem to the case of integral domains. For, if S is integral over R and I is any ideal of S, then S/I contains $R/I \cap R$ and S/I is integral over $R/I \cap R$. Consequently, we know that $S/\mathfrak{P}_1$ is integral over $R/\mathfrak{P}_1 \cap R$. We now show that if $R \subset S$ are integral domains with S integral over R, and if P is a prime ideal in S such that $P \cap R = (0)$, then $P = (0)$.

If $\mathfrak{P} \neq (0)$, there is an $a \in \mathfrak{P}$ with $a \neq 0$. Because S is integral over R, there is a monic polynomial $f \in R[X]$ such that $f(a) = 0$. If $f = X^n + c_1X^{n-1} + \cdots + c_n$, with $c_i \in R$, we may assume that $c_n \neq 0$. For if $c_n = 0$, we have $f = Xg$ with g monic in $R[X]$ and because $f(a) = ag(a) = 0$ with $a \neq 0$, we have $g(a) = 0$. Continuing in this way we see that we eventually come to a monic polynomial in $R[X]$ with a nonzero constant term of which a is a root. Hence, we may assume that $c_n \neq 0$. Because $f(a) = 0$, we have $a^n + c_1a^{n-1} + \cdots + c_n = 0$ or $c_n = -a(a^{n-1} + c_1a^{n-2} + \cdots + c_{n-1}) \in \mathfrak{P}$. But c_n is also in R, so $c_n \in \mathfrak{P} \cap R = 0$ or $c_n = 0$, which is absurd. This proves Proposition 3.3.

Theorem 3.4
Let R be a Dedekind domain with field of quotients K, and let L be a separable extension of K of finite degree. If S is the integral closure of R in L, then S is a Dedekind domain.

PROOF: We know that S is a noetherian integrally closed integral domain. We want to show that every prime ideal other than (0) is maximal. Suppose $\mathfrak{P}_1$ is a nonzero prime ideal in S which is not maximal. Then $\mathfrak{P}_1$ is contained in some maximal ideal $\mathfrak{P}_2$. Because $1 \notin \mathfrak{P}_2$, we have $\mathfrak{P}_1 \cap R \neq R$ and $\mathfrak{P}_2 \cap R \neq R$. Also, by Proposition 3.3, because $\mathfrak{P}_1 \neq (0)$, $\mathfrak{P}_1 \cap R \neq (0)$. Thus, $\mathfrak{P}_1 \cap R$ is a maximal ideal of R and so is $\mathfrak{P}_2 \cap R$. But $\mathfrak{P}_1 \cap R \subset \mathfrak{P}_2 \cap R$ so that $\mathfrak{P}_1 \cap R$ and $\mathfrak{P}_2 \cap R$ must be equal. By Proposition 3.3, we have $\mathfrak{P}_1 = \mathfrak{P}_2$ which is a contradiction. This tells us that $\mathfrak{P}_1$ must be maximal and our proof is complete.

We use this theorem to show that there are Dedekind domains which are not PID's. In Example 2.2 we found that if $R = \mathbf{Z}$ and $L = Q(\sqrt{d})$ where d is a square free integer, then the integral closure S of $\mathbf{Z}$ in L is the set of all elements $x = a + b\sqrt{d}$ satisfy $2a$ and $a^2 - b^2 d$ are in $\mathbf{Z}$. By Theorem 3.4 we now know that S is a Dedekind domain. We consider the ring S for $d = -5$.

Example 3.5 We first show that when $d = -5$, S consists of all $x = a + b\sqrt{-5}$ with a and b integers. Certainly all such elements are in S. It therefore remains to show that if a and b are rational numbers such that $2a$ and $a^2 + 5b^2$ are integers, then a and b are integers. We see this as follows.

If $2a = n$ and $a^2 + 5b^2 = m$ where n and m are integers, $n^2 + 20b^2 = 4m$. Setting $b = u/v$ with u and v relatively prime integers, we have $20u^2 = v^2(4m - n^2)$. If n is even, we get $5u^2 = v^2(m - n^2/4)$. Thus, $v^2 | 5$ so $v = \pm 1$. In this case a and b are both integers.

If n is odd, then $n^2 = 4t + 1$ for some integer t and we have $20u^2 = v^2(4(m - t) - 1)$. Because $4 | 20u^2$ and 4 does not divide $4(m - t) - 1$, we see that $v = 2w$. Hence, $5u^2 = w^2(4(m - t) - 1)$. This equation in $\mathbf{Z}/4\mathbf{Z}$ becomes $\bar{u}^2 + \bar{w}^2 = 0$. It is easy to check that this implies $\bar{u}^2 = 0 = \bar{w}^2$. Hence, u and w must be even, which contradicts the assumption that u and v are relatively prime. This finishes the proof that the elements x of S are precisely of the form $a + b\sqrt{-5}$ with a and b in $\mathbf{Z}$.

It is easy to see that in S, the element $6 = 3 \cdot 2 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. Using the description of the elements of S, it is not hard to show that 2 is an irreducible element of S but 2 does not divide either $1 + \sqrt{-5}$ or $1 - \sqrt{-5}$. Hence 2 is not a prime element. This shows that S is not a UFD and hence not a PID although it is a Dedekind domain.

4. IDEALS

Before studying arbitrary finitely generated modules over Dedekind domains, we study the multiplicative structure of the ideals in R.

Definition
Let R be an integral domain with field of quotients K. An R-submodule J of K will be called an **ideal of R** if there is a nonzero element $c \in R$ such that $cx \in R$ for

all $x \in J$. If J is an ideal of R contained in R (that is, our usual ideal), then J is called an **integral ideal** of R.

Basic Properties 4.1

(a) If J is an ideal of R and R is noetherian, then J is a finitely generated R-module.

(b) If J_1 and J_2 are ideals of R, then the submodule generated by J_1 and J_2 is an ideal of R, denoted by (J_1, J_2).

(c) If J_1 and J_2 are ideals of R, then the submodule generated by all elements of the form a_1a_2 with $a_1 \in J_1$ and $a_2 \in J_2$ is an ideal of R denoted by J_1J_2. The proofs are left as exercises.

Definition

If R is an integral domain and J is an ideal of R, J is called an **invertible ideal** of R if there is an ideal J' such that $JJ' = R$.

Basic Properties 4.2

(a) An ideal J of R is invertible if and only if J is a projective R-module. In this case J is finitely generated.

(b) The set $\mathscr{C}$ of nonzero ideals of R forms a commutative monoid under multiplication in which the invertible elements are precisely the invertible ideals.

(c) R is a Dedekind domain if and only if every nonzero ideal of R is invertible.

(d) R is a Dedekind domain if and only if the set $\mathscr{C}$ of nonzero ideals of R forms a group under multiplication.

(e) R is a Dedekind domain if and only if every integral ideal of R is the product of a finite number of integral prime ideals of R. This factorization into prime ideals is unique when R is Dedekind.

PROOF: (a) If J is invertible, there is an ideal J' such that $JJ' = R$. Hence, we may find $a_1, \ldots, a_n \in J$ and $b_1, \ldots, b_n \in J'$ with $\Sigma a_ib_i = 1$. Define $f_i : J \to R$ by $f_i(a) = ab_i$ for $a \in J$. f_i is an R-morphism of J into R and, for every $a \in J$ we have $a = a \cdot 1 = a \Sigma a_ib_i = \Sigma a_i(ab_i) = \Sigma f_i(a)a_i$. Thus, J is seen to be projective.

Conversely, if J is projective, we may find $\{a_i\} \subset J$ and $f_i : J \to R$ such that for all $a \in J$ $f_i(a) = 0$ except for a finite number of i and $a = \Sigma f_i(a)a_i$. Let $c \in R$ be a nonzero element such that $cJ \subset R$. Then for any $a_1, a_2 \in J$ and $f_i : J \to R$, we have $f_i(ca_1a_2) = ca_1f_i(a_2) = ca_2f_i(a_1)$ so that, because $c \neq 0$, we obtain $a_1f_i(a_2) = a_2f_i(a_1)$. Thus, $f_i(a)/a = q_i$ in K, the field of quotients of R, is the same for all a in J. Because only a finite number of the $f_i(a) \neq 0$, only a finite number, say $q_1, \ldots, q_n$, are not zero. If $a \in J$ and $a \neq 0$, we have $a = \Sigma_{i=1}^n f_i(a)a_i = \Sigma_{i=1}^n (q_ia)a_i = a \Sigma_{i=1}^n q_ia_i$. Hence, $\Sigma_{i=1}^n q_ia_i = 1$, and we see again that J is generated by $\{a_1, \ldots, a_n\}$. Let J' be the R-submodule of K generated by $\{q_1, \ldots, q_n\}$. Because J' is finitely generated, we find a nonzero $c \in R$ such that $cJ' \subset R$ by taking the common denominator of $q_1, \ldots, q_n$. Hence, J' is an ideal. Clearly, $JJ' = R$ and thus J is invertible.

(b) Left to the reader.

(c), (d) and (e) Here what we shall do is prove that if R is Dedekind, then every ideal of R is invertible. Then we prove that if every ideal in R is invertible, the set $\mathscr{C}$ of nonzero ideals of R is a group under multiplication. Next we will

prove that if $\mathscr{C}$ is a group under multiplication, then every integral ideal is the product of a finite number of integral prime ideals. Finally, we will prove that if every integral ideal of R is a product of finitely many prime ideals, then R is Dedekind. This will complete the cycle and take care of (c), (d) and (e). The last part of (e) will be proven separately.

Suppose that R is Dedekind. Then every integral ideal of R is projective and hence invertible. For if J is any ideal of R, there is a nonzero $c \in R$ such that cJ is an integral ideal. Because cJ is projective, J is projective (being isomorphic to cJ), and so J is invertible.

Now if every nonzero ideal of R is invertible, it follows from (b) that $\mathscr{C}$ is a group under multiplication.

If $\mathscr{C}$ is a group under multiplication, let us first show that R must be noetherian. Because $\mathscr{C}$ is a group, every ideal is invertible (because R is the identity element of the group), and thus every ideal is finitely generated [by (a)].

Now suppose that not every integral ideal is the product of a finite number of prime ideals. Because R is noetherian, there is a maximal such integral ideal, say I. I cannot be a maximal ideal, because then it would be prime. Hence, I is contained in some maximal ideal $\mathfrak{M}$. Because $I \subset \mathfrak{M}$, we have $\mathfrak{M}^{-1}I \subset \mathfrak{M}^{-1}M = R$. Notice that for any invertible ideal J, $J^{-1} = \{x \in K | xJ \subset R\}$. Because since $J^{-1}J = R$, J^{-1} is certainly a subset of this set (which we shall call J', even when J is not assumed to be invertible). Since $R = J^{-1}J \subset J'J \subset R$, we have $J^{-1}J = J'J$ so that, multiplying by J^{-1} on the right we get $J^{-1} = J'$. In particular, if J is an integral ideal, $J' \supset R$ so that in our case $\mathfrak{M}^{-1} = \mathfrak{M}' \supset R$. Hence, $\mathfrak{M}^{-1}I \supset I$. If $\mathfrak{M}^{-1}I = I$, we get $I = \mathfrak{M}I$ and, because R is a noetherian integral domain, $I = 0$. This follows from the fact that localizing at $\mathfrak{M}$, we get $IR_{\mathfrak{M}} = \mathfrak{M}\, IR_{\mathfrak{M}}$ so that $IR_{\mathfrak{M}} = (0)$. Because R is an integral domain, $IR_{\mathfrak{M}} = (0)$ implies that $I = (0)$. However, we are assuming that $I \neq (0)$ so $\mathfrak{M}^{-1}I$ properly contains I and is therefore the product of a finite number of prime ideals; that is, $\mathfrak{M}^{-1}I = \mathfrak{P}_1 \cdots \mathfrak{P}_n$. But then $I = \mathfrak{M}\mathfrak{P}_1 \cdots \mathfrak{P}_s$ is also the product of a finite number of prime ideals, contrary to the nature of I. Thus we have our conclusion.

The last step is to show that if every nonzero integral ideal is the product of a finite number of prime ideals, then R is Dedekind. It is sufficient to prove that the hypothesis implies that every nonzero integral ideal is invertible, from which the fact that R is Dedekind follows immediately due to (a). In order to show that every nonzero integral ideal is invertible, it suffices to prove that every prime ideal is invertible because a product of invertible ideals is clearly invertible and every ideal is assumed to be a product of prime ideals.

Notice first that every nonzero prime ideal contains an invertible prime ideal. For, let $\mathfrak{P} \neq (0)$ be a prime ideal and let $x \in \mathfrak{P}$, $x \neq 0$. Then $(x) \subset \mathfrak{P}$ and $(x) = \mathfrak{P}_1 \cdots \mathfrak{P}_t$ where the $\mathfrak{P}_i$ are prime ideals. Because $(x) \subset \mathfrak{P}$, we have $\mathfrak{P}_1 \cdots \mathfrak{P}_t \subset \mathfrak{P}$ from which it follows that some $\mathfrak{P}_i$ is contained in $\mathfrak{P}$. Now clearly (x) is an invertible ideal (it's free!) and, from general principles in any commutative monoid we know that if a product of elements is invertible, then each of the factors is invertible. Hence, each $\mathfrak{P}_i$ is invertible, and $\mathfrak{P}$ contains an invertible prime ideal.

Next we show that if $\mathfrak{P}$ is an invertible prime ideal, then $\mathfrak{P}$ is a maximal ideal. To prove that $\mathfrak{P}$ is maximal, it suffices to show that if a is any element of R not

contained in P, then $(\mathfrak{P}, a) = R$. However, because $\mathfrak{P}$ is invertible, it will suffice to prove that $\mathfrak{P}(\mathfrak{P}, a) = \mathfrak{P}$, for then, multiplying both sides by $\mathfrak{P}^{-1}$, we get $(\mathfrak{P}, a) = R$. Because $\mathfrak{P}$ obviously contains $\mathfrak{P}(\mathfrak{P}, a)$, all we must do now is show that $\mathfrak{P}$ is contained in $\mathfrak{P}(\mathfrak{P}, a)$.

In R, the ideal $(\mathfrak{P}, a)$ is a product of prime ideals: $(\mathfrak{P}, a) = \mathfrak{Q}_1 \cdots \mathfrak{Q}_t$. In $\bar{R} = R/\mathfrak{P}$, we therefore have $\overline{(\mathfrak{P}, a)} = \bar{\mathfrak{Q}}_1 \cdots \bar{\mathfrak{Q}}_t$ where $\bar{I}$ in $\bar{R}$ means $I/\mathfrak{P}$ for any integral ideal I in R containing $\mathfrak{P}$. Now $\overline{(\mathfrak{P}, a)}$ is a principal ideal in the integral domain $\bar{R}$, so that $\overline{(\mathfrak{P}, a)}$ is invertible. Hence, each of the ideals $\bar{\mathfrak{Q}}_i$ is an invertible prime ideal in $\bar{R}$. The ideal $\overline{(\mathfrak{P}, a^2)}$ is also a principal, hence invertible, ideal of $\bar{R}$ and, because $\overline{(\mathfrak{P}, a^2)} = \overline{(\mathfrak{P}, a)}^2$, we have $\overline{(\mathfrak{P}, a^2)} = \bar{\mathfrak{Q}}_1^2, \ldots, \bar{\mathfrak{Q}}_t^2$.

Writing $(\mathfrak{P}, a^2)$ as a product of prime ideals in R, we have $(\mathfrak{P}, a^2) = \mathfrak{Q}_1' \cdots \mathfrak{Q}_s'$, so that $\overline{(\mathfrak{P}, a^2)} = \bar{\mathfrak{Q}}_1' \cdots \bar{\mathfrak{Q}}_s'$, and each of the $\bar{\mathfrak{Q}}_i'$ is an invertible prime ideal of $\bar{R}$. We have now got the ideal $\overline{(\mathfrak{P}, a^2)}$ in $\bar{R}$ written as the product of invertible prime ideals in two different ways. To see what this implies, let us prove the following.

Lemma 4.3
Let R be an integral domain. If I is an integral ideal of R and $I = \mathfrak{P}_1 \cdots \mathfrak{P}_t = \mathfrak{Q}_1 \cdots \mathfrak{Q}_s$, where the $\mathfrak{P}_i$ and $\mathfrak{Q}_j$ are proper invertible prime ideals of R, then $s = t$ and, after renumbering the Q_i, we have $\mathfrak{P}_i = \mathfrak{Q}_i$ for $i = 1, \ldots, s$.

PROOF: We proceed by induction on t. The case $t = 1$ is left to the reader.

For $t > 1$, consider the set $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_t\}$ of prime ideals and from this set take a minimal element, say $\mathfrak{P}_1$. Because $\mathfrak{P}_1 \supset \mathfrak{P}_1 \cdots \mathfrak{P}_t = \mathfrak{Q}_1 \cdots \mathfrak{Q}_s$, we must have $\mathfrak{P}_1$ containing some $\mathfrak{Q}_i$, say $\mathfrak{Q}_1$. Similarly, $\mathfrak{Q}_1$ must contain some $\mathfrak{P}_i$ so we have $\mathfrak{P}_1 \supset \mathfrak{Q}_1 \supset \mathfrak{P}_i$. But having chosen $\mathfrak{P}_1$ to be a minimal element of the set $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_t\}$, we have $\mathfrak{P}_1 = \mathfrak{P}_i$, and so $\mathfrak{P}_1 = \mathfrak{Q}_1$. Multiplying by $\mathfrak{P}_1^{-1}$, we get $\mathfrak{P}_2 \cdots \mathfrak{P}_t = \mathfrak{Q}_2 \cdots \mathfrak{Q}_s$, and our result follows by induction.

Applying this fact to our situation, we have $\overline{(\mathfrak{P}, a^2)} = \bar{\mathfrak{Q}}_1' \cdots \bar{\mathfrak{Q}}_s' = \bar{\mathfrak{Q}}_1\bar{\mathfrak{Q}}_1\bar{\mathfrak{Q}}_2\bar{\mathfrak{Q}}_2 \cdots \bar{\mathfrak{Q}}_t\bar{\mathfrak{Q}}_t$ so that $s = 2t$ and, by renumbering, we have $\mathfrak{Q}_{2j-1}' = \mathfrak{Q}_{2j}' = \mathfrak{Q}_j$ for $j = 1, \ldots, t$. Because the $\mathfrak{Q}_j'$ and $\mathfrak{Q}_i'$ contain $\mathfrak{P}$, we therefore have $\mathfrak{Q}_{2j-1}' = \mathfrak{Q}_{2j}' = Q_j$ for $j = 1, \ldots, t$. Hence, $(\mathfrak{P}, a^2) = \mathfrak{Q}_1^2 \cdots \mathfrak{Q}_t^2 = (\mathfrak{P}, a)^2$ in R. Now $\mathfrak{P}$ is contained in $(\mathfrak{P}, a^2)$, and so $\mathfrak{P}$ is seen to be contained in $(\mathfrak{P}, a)^2$. This means that if $x \in \mathfrak{P}$, then $x = y + za + ra^2$ where $y \in \mathfrak{P}^2$, $z \in \mathfrak{P}$, and $r \in R$. Then $ra^2 = x - y - za \in \mathfrak{P}$ and, because $a \notin \mathfrak{P}$, $r \in \mathfrak{P}$. Thus, $x = y + (z + ra)a$ where $y \in \mathfrak{P}^2$ and $z + ra \in \mathfrak{P}$, which means that $x \in (\mathfrak{P}^2, \mathfrak{P}a) = \mathfrak{P}(\mathfrak{P}, a)$. Because this is true for every $x \in \mathfrak{P}$, we have $\mathfrak{P} \subset \mathfrak{P}(\mathfrak{P}, a)$ which is what we wanted to prove.

Having shown that every prime ideal in R contains an invertible prime ideal, and that every invertible prime ideal is maximal, we now know that every prime ideal in R is invertible. This proves that every integral ideal is invertible and hence R is Dedekind. That the factorization into a product of prime ideals is unique in a Dedekind domain follows from Lemma 4.3.

These basic properties show that Dedekind domains share many arithmetic properties with PID's. If we think only in terms of ideals rather than elements, we still get unique factorization into prime ideals in Dedekind domains.

This unique factorization into prime ideals is very useful. For instance, sup-

pose we have an integral ideal I of a Dedekind domain R. Then $I = \mathfrak{P}_1, \ldots, \mathfrak{P}_r$ with each $\mathfrak{P}_i$ a prime ideal. The ideals, $\mathfrak{P}_i$, are the only prime ideals of R containing I for if $\mathfrak{P}$ is a prime containing I, $\mathfrak{P}$ contains the product $\mathfrak{P}_1 \cdots \mathfrak{P}_r$ and therefore contains some $\mathfrak{P}_i$. But all the nonzero prime ideals of a Dedekind domain are maximal, so that $\mathfrak{P} = \mathfrak{P}_i$. Now if a is an element of R not contained in any of the ideals $\mathfrak{P}_1, \ldots, \mathfrak{P}_r$, then $(I, a) = R$. For if $\mathfrak{P}$ is a proper prime ideal containing (I, a), it must also contain I and must therefore be one of the ideals $\mathfrak{P}_i$. But because $\mathfrak{P}$ contains (I, a), it contains a and hence $a \in \mathfrak{P}_i$ which is impossible. Thus, $(I, a) = R$.

If we let $S = R - UP_i$, we know that S is a multiplicative set and that R_S is a Dedekind semilocal domain. Therefore, by Corollary 1.4, R_S is a PID. Also, we know that $R_S/IR_S = (R/I)_S$. Because $(I, a) = R$ for every $a \in S$, multiplication by every element of S is an isomorphism on R/I. From this it follows that $(R/I)_S = R/I$. As a result we have the following.

Proposition 4.4
If R is a Dedekind domain and I is a proper integral ideal of R, then every ideal of R/I is principal. In fact, $R/I \approx R_S/IR_S$ where $S = R - \cup P_i$ and $I = \Pi P_i$.

PROOF: We have already seen that $R/I \approx R_S/IR_S$. Because we know that R_S is a PID (being a semilocal Dedekind domain), it follows that every ideal of R_S/IR_S is principal, and so the same is true of R/I.

As an immediate corollary we have the following.

Corollary 4.5
Let R be a Dedekind domain. Then every ideal of R may be generated by two elements. In fact, if I is an ideal of R and $a \in I$ is a nonzero element, there is a $b \in I$ such that $I = (a, b)$.

Knowing that every ideal in a Dedekind domain may be generated by two elements with one of the two elements arbitrarily preassigned enables us to prove a useful lemma.

Lemma 4.6
Let R be a Dedekind domain, and let I_1 and I_2 be ideals of R with I_2 integral. Then there is an integral ideal J_1 isomorphic to I_1^{-1} with $(J_1, I_2) = R$.

PROOF: Let $I_1 = (a, b)$ with a a nonzero element of $I_1 I_2$. Because $I_1 I_2 \subset I_1$ (do not forget that I_2 is assumed to be integral), we have $(I_1 I_2, b) \subset I_1 = (a, b) \subset (I_1 I_2, b)$ so that $I_1 = (I_1 I_2, b)$. Now $(b) = I_1(I_1^{-1} b)$; so we have $I_1 = (I_1 I_2, b) = (I_1 I_2, I_1(I_1^{-1} b) = I_1(I_2, I_1^{-1} b))$. Letting $J_1 = I_1^{-1} b$, we have $J_1 \subset R$ (because $b \in I_1$), and from the fact that $I_1 = I_1(I_2, J_1)$ we get $(I_2, J_1) = R$.

The reader may well wonder why, of all possible lemmas, we have chosen to record Lemma 4.6. To justify our choice, consider two (not necessarily distinct) integral ideals I_1 and I_2 in a Dedekind domain R. Applying Lemma 4.6 to the ideals I_1^{-1} and I_2, we obtain an integral ideal $J_1 = (I_1^{-1})^{-1} b = I_1 b$ such that $(J_1 I_2) = R$. Because J_1 and I_1 are isomorphic, the sum $I_1 \amalg I_2$ is isomorphic to $J_1 \amalg I_2$. The mor-

phism $\sigma: J_1 \amalg I_2 \to R$ defined by $\sigma(a_1, a_2) = a_1 + a_2$ is an epimorphism because $(J_1, I_2) = R$ and Ker $\sigma \approx J_1 \cap I_2 = J_1 I_2$ [the last equality being true because $(J_1, I_2) = R$]. We therefore have the splittable exact sequence $0 \to J_1 I_2 \to J_1 \amalg I_2 \to R \to 0$ so that $J_1 \amalg I_2 \approx R \amalg J_1 I_2$. Finally, because $J_1 = I_1 b$, it follows that $J_1 I_2 \approx I_1 I_2$ and we therefore get $I_1 \amalg I_2 \approx J_1 \amalg I_2 \approx R \amalg J_1 I_2 \approx R \amalg I_1 I_2$. From this discussion we obtain the following.

Theorem 4.7
Let R be a Dedekind domain and M a finitely generated projective R-module. Then M is isomorphic to the sum $F \amalg I$ where F is a free R-module of finite rank, and I is an (integral) ideal of R.

PROOF: By Chapter 10, Theorem 1.1 we know that $M \approx I_1 \amalg \cdots \amalg I_n$ where each I_i is an integral ideal of R. If $n = 1$, our proposition is true because we may then take $F = (0)$. If $n = 2$, our preceding discussion has shown us that $I_1 \amalg I_2 \approx R \amalg I_1 I_2$ and so we may take $F = R$ and $I = I_1 I_2$. Proceeding by induction, suppose we know that $I_1 \amalg \cdots \amalg I_{n-1} \approx F' \amalg I_1 \cdots I_{n-1}$. Then $I_1 \amalg \cdots \amalg I_n \approx F' \amalg I_1 \cdots I_{n-1} \amalg I_n \approx F' \amalg R \amalg I_1 \cdots I_{n-1} I_n \approx F \amalg I$ where $F = F' \amalg R$ and $I = I_1 \cdots I_n$. This proves the theorem.

Theorem 4.7 as it now stands is not completely satisfactory, for we would like to have some uniqueness theorem. Namely, we would like to say that if $M \approx F_1 \amalg I_1 \approx F_2 \amalg I_2$, then $F_1 \approx F_2$ and $I_1 \approx I_2$. That $F_1 \approx F_2$ is clear because they both are free modules having the same rank. The problem, then, is the uniqueness of the ideals I_1, I_2. In the exercises, you will be asked to prove that if $M = F \amalg I$ with F free of rank n, then $I^{-1} \approx \mathrm{Sk}_{n+1}(M)$. This, then, will show that I^{-1}, and hence, also I, is determined uniquely by M.

5. FINITELY GENERATED MODULES OVER DEDEKIND DOMAINS

Having determined the structure of finitely generated projective modules over Dedekind domains, we consider arbitrary finitely generated modules.

If M is a finitely generated R-module we have the exact sequence

$$0 \longrightarrow t(M) \longrightarrow M \xrightarrow{k} \frac{M}{t(M)} \longrightarrow 0$$

Because $M/t(M)$ is a finitely generated torsion free R-module, we know by Chapter 9, Proposition 3.3 that $M/t(M)$ is a submodule of a free R-module. Hence, $M/t(M)$ is a projective R-module, because R is a Dedekind domain. Thus, k is a splittable epimorphism which implies that $M \approx t(M) \amalg M/t(M)$. Because we know what $M/t(M)$ looks like, we turn our attention to determining the structure of finitely generated torsion modules over Dedekind domains.

To this end, let M be a finitely generated torsion module over the Dedekind domain R. Then it is clear that the annihilator of M is not zero, and we let $I = \mathrm{ann}(M)$. If $I = \mathfrak{P}_1 \cdots \mathfrak{P}_s$ and $S = R - U\mathfrak{P}_i$, we know that $R/I = R_S/IR_S \approx R/I \underset{R}{\otimes} R_S$. Also, because $IM = (0)$, we have $M = M/IM \approx R/I \underset{R}{\otimes} M \approx$

$(R/I \underset{R}{\otimes} R_S) \underset{R}{\otimes} M \approx R/I \underset{R}{\otimes} (R_S \underset{R}{\otimes} M)$. Letting $M' = R_S \underset{R}{\otimes} M$ and $R' = R_S$, we know that M' is a finitely generated torsion module over the PID R'. Hence, M' is the sum $R'/I_1' \amalg R'/I_2' \cdots \amalg R'/I_n'$ where I_j' are nonzero ideals of R' and $I_1' \subset \cdots \subset I_n'$. In fact, because $\operatorname{ann}(M') = I_1' = IR_S$, and because $I_j' = (I_j' \cap R)R'$, we have

$$\frac{R}{I} \underset{R}{\otimes} M' \approx \frac{R}{I} \underset{R}{\otimes} \left(\frac{R'}{I_1'} \amalg \cdots \amalg \frac{R'}{I_n'} \right) \approx \frac{R}{I} \underset{R}{\otimes} \left(\frac{R}{I_1 \underset{R}{\otimes} R'} \amalg \cdots \amalg \frac{R}{I_n \underset{R}{\otimes} R'} \right)$$

where $I_j = I_j' \cap R$ and $I = I_1 \subset I_2 \subset \cdots \subset I_n$. Hence, M' is isomorphic to $(R/I_1 \amalg \cdots \amalg R/I_n) \underset{R}{\otimes} R'$. Finally, because multiplication by every element of S is an isomorphism on R/I, the same is true for R/I_j for $j = 1, \ldots, n$. Thus, $R/I_j \underset{R}{\otimes} R' \approx R/I_j$. Because $M' = M$, we have $M \approx R/I_1 \amalg \cdots \amalg R/I_n$ with $I_1 \subset I_2 \subset \cdots \subset I_n$. We now summarize these results in the following.

Theorem 5.1
If M is a finitely generated module over the Dedekind domain R, then M is isomorphic to a sum $R/I_1 \amalg \cdots \amalg R/I_n \amalg F \amalg I$ where $I_1 \subset \cdots \subset I_n$ are integral nonzero ideals of R, F is a free module of finite rank, and I is an invertible ideal. The ideals $I, I_1, \ldots, I_n$ are uniquely determined by M and so is the module F.

PROOF: We have already seen that M is the sum of $t(M)$ and $M/t(M)$ and that these modules are uniquely determined by M. By our preceding discussions, $M/t(M)$ is projective and is therefore isomorphic to $F \amalg I$, while $t(M)$ is isomorphic to $R/I_1 \amalg \cdots \amalg R/I_n$ with $I_1 \subset \cdots \subset I_n$. Our uniqueness theorem, Chapter 19, Theorem 5.5, tells us that the ideals I_j are uniquely determined by $t(M)$, while $F \amalg I$ is determined by $M/t(M)$. We have already observed that the ideal I is uniquely determined by $M/t(M)$.

The fact that F is also uniquely determined by $M/t(M)$ is obvious.

EXERCISES

(1) Let R be a Dedekind domain and M an R-module with $M = F \amalg I$ where F is a finitely generated free R-module of rank n and I is an ideal of R. Prove that $\operatorname{Sk}_{n+1}(M) \approx I^{-1}$. [*Hint*: Using the fact that $\operatorname{Sk}_p(I \amalg R) \approx \operatorname{Sk}_p(I) \amalg \operatorname{Sk}_{p-1}(I)$, show that $\operatorname{Sk}_{n+1}(M) \approx \coprod_{j=1}^{n+1} \binom{n+1}{j} \operatorname{Sk}_j(I)$ where $\binom{n+1}{j}$ is the binomial coefficient and $\binom{n+1}{j} \operatorname{Sk}_j(I)$ means the sum of $\operatorname{Sk}_j(I)$ with itself $\binom{n+1}{j}$ times. Then using the fact that $\operatorname{Sk}_p(I) = 0$ for $p \geq 2$, show that $\operatorname{Sk}_{n+1}(M) \approx \operatorname{Sk}_1(I) \approx \operatorname{Hom}_R(I, R) = I^{-1}$.
(2) Let R be an integrally closed integral domain with field of quotients K, and let L be an algebraic extension of K. If x is an element of L which is integral over R, and $f(X) \in K[X]$ is the minimal polynomial of x over K, prove that $f(X)$ is in $R[X]$.
(3) Let R be an integrally closed domain with field of quotients K. Let L be a field containing K and α an element of L which is algebraic over K. If $f(X)$ in $K[X]$ is

the minimal polynomial of α over K, prove that α is integral over R if and only if $f(X)$ is in $R[X]$.

(4) Let ζ be a primitive mth root of unity.

(a) Prove that if $f(X)$ in $\mathbf{Q}[X]$ is the minimal polynomial for ζ over $\mathbf{Q}$, then $f(X)$ is in $\mathbf{Z}[X]$ where $\mathbf{Z}$ is the ring of integers.

(b) Prove that there is a monic polynomial $g(X)$ in $\mathbf{Z}[X]$ such that $X^m - 1 = f(X)g(X)$.

(c) If p is a prime number and $f(\zeta^p) \neq 0$, prove that there is a monic polynomial $k(X)$ in $\mathbf{Z}[X]$ such that $g(X^p) = f(X)k(X)$, where $g(X)$ is the polynomial whose existence was proven in part (b).

(d) Under the hypotheses of part (c), prove that the polynomial $X^m - 1$ in $\mathbf{Z}/p\mathbf{Z}[X]$ has a zero derivative. [*Hint*: Use the fact that $g(X^p) = f(X)k(X)$ and that the coset of $g(X^p)$ in $\mathbf{Z}/p\mathbf{Z}[X]$ is the same as that of $g(X)^p$ in $\mathbf{Z}/p\mathbf{Z}[X]$. Thus, the reductions modulo p of $f(X)$ and $g(X)$ have a common nonzero root.]

(e) Conclude from part (d) that if p is a prime which does not divide m, then ζ^p must be a root of f.

(f) Conclude from the above that $\Phi_m(X)$ must be an irreducible polynomial in $\mathbf{Z}[X]$, that is, $\Phi_m(X)$ is the minimal polynomial for the primitive mth roots of unity.

(5) Let K be a quadratic extension of the rational numbers $\mathbf{Q}$.

(a) Prove that there is a square-free integer d (that is, an integer d having no square factors) such that K is isomorphic to $\mathbf{Q}(\sqrt{d})$.

(b) Let $\mathbf{Q}(\sqrt{d})$ be the integral closure of $\mathbf{Z}$ in $\mathbf{Q}(\sqrt{d})$. Prove that if $d = 2 \bmod 4$ or $d \equiv 3 \bmod 4$, then the elements $1, \sqrt{d}$ are a free basis of $\mathbf{Q}(\sqrt{d})$ over $\mathbf{Z}$.

(c) Prove that if $d \equiv 1(4)$, then the elements $1, 1/2(1+\sqrt{d})$ are a free basis of $\mathbf{Q}(\sqrt{d})$ over $\mathbf{Z}$.

The next set of exercises is devoted to an exposition of the theorem known as the Hilbert Nullstellensatz.

(6) Let R be a subring of the integral domain S.

(a) Suppose that x is an element of R which has an inverse $1/x$ in S. Show that $1/x$ is in R if $1/x$ is integral over R.

(b) Suppose S is integral over R and S is a field. Show that R is then a field.

(7) Show that the following statements are equivalent for an integral domain R with field of quotients K.

(a) There is a finitely generated ring extension $R[x_1, \ldots, x_n]$ of R which is a field that is algebraic over K (because $R[x_1, \ldots, x_n]$ is a field containing R, it clearly contains K).

(b) There is a nonzero element y in R such that R_V is a field where V is the multiplicative subset $\{y^n\}_{n \in \mathbf{N}}$ of R.

(c) The intersection of all the nonzero prime ideals of R is not zero. [*Hint*: To show that (a) implies (b), first show that there are monic polynomials $f_1, \ldots, f_n$ in $K[X]$ such that $f_i(x_i) = 0$ for $i = 1, \ldots, n$. Next, show that there is a nonzero element y in R such that all the polynomials $f_1, \ldots, f_n$ are in $R_V[X]$. Finally, show that $R[x_1, \ldots, x_n]$ is an integral extension of R_V, and hence R_V is also a field.]

(8) Suppose $K \subset L$ are fields such that L is a finitely generated ring extension of K; that is, there is a finite family of elements $x_1, \ldots, x_n$ in L such that $L = K[x_1, \ldots, x_n]$. Then L is an algebraic extension of K with $[L:K] < \infty$. [*Hint*: It suffices to show that all the x_i are algebraic over K. Suppose this is not the case. Then after suitable relabeling we can assume that $x_1, \ldots, x_t$ is a transcendence basis for L over K with $t \geq 1$. Show that this implies that the ring $K[x_1, \ldots, x_t]$ has a nonzero radical and is isomorphic to the polynomial ring $K[X_1, \ldots, X_t]$. Finally show that this leads to a contradiction by showing that if S is any integral domain, then the radical of $S[X]$ is zero.]

(9) Suppose K is a field, M is a maximal ideal of $K[X_1, \ldots, X_n]$, and L is the field $K[X_1, \ldots, X_n]/M$.

(a) Show that the composition of ring morphisms $K \to K[X_1, \ldots, X_n] \xrightarrow{k} L$ is an injective morphism. This gives an identification of K with a subfield of L which enables us to consider K a subfield of L.

(b) Show that L is an algebraic extension of K with $[L:K] < \infty$. This result is known as the Hilbert Nullstellensatz.

(10) Suppose K is an algebraically closed field. Show that the polynomial ring $K[X_1, \ldots, X_n]$ has the following properties:

(a) An ideal I of $K[X_1, \ldots, X_n]$ is a maximal ideal if and only if there are n elements $c_1, \ldots, c_n$ in K such that I is generated by the elements $X_1 - c_1, \ldots, X_n - c_n$.

(b) Let $c_1, \ldots, c_n$ and $c_1', \ldots, c_n'$ be two families of n elements in K (with possible repeats). Show that the maximal ideals generated by $X_1 - c_1, \ldots, X_n - c_n$ and $X_1 - c_1', \ldots, X_n - c_n'$ are the same if and only if $c_i = c_i'$ for all $i = 1, \ldots, n$.

(11) Show that the following statements are equivalent for an integral domain R.

(a) There is a finitely generated ring extension $R[x_1, \ldots, x_n]$ of R which is a field.

(b) There is a nonzero element y in R such that R_V is a field where V is the multiplicative subset $\{y^n\}_{n \in \mathbb{N}}$ of R.

(c) The intersection of all the nonzero prime ideals of R is not zero. Further, if $S \supset R$ is a finitely generated ring extension of R which is a field, then $S \supset K$ where K is the field of quotients of R and $[S:K] < \infty$.

(12) A commutative ring R is said to be a **Jacobson ring** if each prime ideal $\mathfrak{P}$ of R is the intersection of the maximal ideals of R containing $\mathfrak{P}$. Show that if R is a Jacobson ring, then:

(a) R/I is a Jacobson ring for all ideals I of R.

(b) The radical of R is the ideal of R consisting of all the nilpotent elements of R.

(13) Suppose $S \supset R$ are commutative rings with S a finitely generated ring extension of R. Suppose, also, that R is a noetherian Jacobson ring. Then:

(a) S is a noetherian Jacobson ring.

(b) If M is a maximal ideal of S, then $M' = R \cap M$ is a maximal ideal of R and S/M is a finite algebraic extension of R/M'. [*Hint*: Obviously S is noetherian. Assume that S is not a Jacobson ring. Show that this implies that there is a prime ideal $\mathfrak{P}$ of S which is not the intersection of the maximal ideals of S containing $\mathfrak{P}$ but such that each prime ideal of S properly containing $\mathfrak{P}$ is the intersection of the maximal ideals containing it. Let $R' = R/\mathfrak{P}'$ where $\mathfrak{P}' = R \cap \mathfrak{P}$, and let $S' = S/\mathfrak{P}$. Show that $R' \subset S'$ are integral domains having the

properties:

(i) S' is a finitely generated ring extension of R'.

(ii) Each nonzero prime ideal of S' is the intersection of the maximal ideals containing it.

(iii) The radical of S' is not zero. Show that this leads to a contradiction.]

(14) Let R be a local Dedekind domain with field of quotients K. If $R \neq K$, show that all of K is the only subring of K properly containing R.

(15) Let R be a Dedekind domain which is not equal to its field of quotients K. Let S be a proper subring of K containing R.

(a) Assume that $\mathfrak{P}$ is a nonzero prime ideal of S and $\mathfrak{P}' = R \cap \mathfrak{P}$. Show:

(i) $\mathfrak{P}' \neq 0$, and so $\mathfrak{P}'$ is a maximal ideal of R.

(ii) $R_{\mathfrak{P}'} = S_{\mathfrak{P}'} = S_{\mathfrak{P}}$.

(iii) The natural ring morphism $R/\mathfrak{P}' \to S/\mathfrak{P}'S$ is injective, and hence the ring morphism $S/\mathfrak{P}'S \to S_{\mathfrak{P}'}/(\mathfrak{P}'S)_{\mathfrak{P}'}$ is injective.

(iv) Each of the ring morphisms in the composition $R/\mathfrak{P}' \to S/\mathfrak{P}'S \to S_{\mathfrak{P}'}/(\mathfrak{P}'S)_{\mathfrak{P}'} \to S_{\mathfrak{P}}/\mathfrak{P}S_{\mathfrak{P}}$ is an isomorphism.

(v) $\mathfrak{P}'S$ is a maximal ideal of S.

(vi) $\mathfrak{P} = \mathfrak{P}'S$ and is a finitely generated maximal ideal of S.

(b) S is a noetherian ring with the property that each nonzero prime ideal of S is a maximal ideal of S.

(c) S is a Dedekind ring.

(d) Let X be the set of nonzero prime ideals of R which are the intersection of a prime ideal of S with R. Then $S = \bigcap_{\mathfrak{P} \in X} R_{\mathfrak{P}}$. Further, $S = R$ if and only if X contains each nonzero prime ideal of R.

(16) Let R be a Dedekind domain with field of quotients K and let $C(R)$ be the group of ideals of R.

(a) Show that $C(R)$ is a free abelian group with the integral prime ideals in $C(R)$ a basis for $C(R)$.

(b) Show that if I is an ideal in $C(R)$, then there are integral ideals J_1 and J_2 such that $I = J_1 J_2^{-1}$ and $(J_1, J_2) = R$.

(c) Define the map $f: \cup(K) \to C(R)$ by $f(x) = xR$ for all x in $\cup(K)$ where $\cup(K)$ denotes the group of units of K. Show:

(i) f is a morphism of groups.

(ii) Ker $f = \cup(R)$.

(iii) $\cup(K)/\cup(R)$ is a free abelian group.

(d) Coker f is called the ideal class group of R and is usually denoted by $\mathrm{Cl}(R)$. Show that the following statements are equivalent for any pair of ideals I_1 and I_2 in $C(R)$:

(i) $k(I_1) = k(I_2)$, where $k: C(R) \to \mathrm{Cl}(R)$ is the canonical group surjection.

(ii) There is an x in $\cup(K)$ such that $xI_1 = I_2$.

(iii) I_1 and I_2 are isomorphic R-modules.

(e) Show the image in $Cl(R)$ of an ideal I in $C(R)$ is a torsion element of $Cl(R)$ if and only if $I^n = Rx$ for some positive integer N and some x in K.

(17) Let R be a Dedekind domain with field of quotients K. For each subset A of the nonzero prime ideals of R let R^A be the intersection of all the subrings $R_{\mathfrak{P}}$ of K with $\mathfrak{P}$ a prime ideal not in A. Then:

(a) R^A is a Dedekind domain containing R with field of quotients K.

(b) If $\mathfrak{P}$ is a nonzero prime ideal of R, then $\mathfrak{P}R^{\wedge} = R^{\wedge}$ if and only if $\mathfrak{P}$ is in A.

(c) An element x in K is a unit in $R^{\wedge}$ if and only if the ideal Rx can be written as a product $\Pi\, \mathfrak{P}_i^{\nu_i}$ with the $\mathfrak{P}_i$ in A and the ν_i integers (negative as well as positive integers).

(d) Let A be a finite set. Then the group morphism $g: \cup(R^{\wedge}) \to C(R)$ given by $g(x) = Rx$ for all x in $\cup(R^{\wedge})$ has the following properties:

(i) Im g is contained in the subgroup of $C(R)$ generated by the $\mathfrak{P}$ in A.

(ii) Ker $g = \cup(R)$ and hence $\cup(R^{\wedge})/\cup(R)$ is a free group of rank at most card(A).

(iii) Rank $\cup(R^{\wedge})/\cup(R) =$ card A if and only if the image of $\mathfrak{P}$ in $Cl(R)$ is a torsion element for each $\mathfrak{P}$ in A.

(18) Let R be a Dedekind ring with field of quotients K. Show that the following statements are equivalent.

(a) If S is a subring of K containing R, then there is a multiplicative subset V of R such that $S = R_V$.

(b) $Cl(R)$ is a torsion group. [*Hint*: To show that (a) implies (b) it clearly suffices to show that the image of $\mathfrak{P}$ in $Cl(R)$ is a torsion element for each prime ideal $\mathfrak{P}$ in $C(R)$. Let $\mathfrak{P}$ be a prime ideal in $C(R)$ and let $A = \{\mathfrak{P}\}$. Then because $R^{\wedge}$ is a ring of quotients of R and is different from R, it follows that there is a nonunit in R which is a unit in $R^{\wedge}$. Show that this implies the image of $\mathfrak{P}$ in $Cl(R)$ is a torsion element of R.

To show that (b) implies (a), suppose S is a subring of K containing R. Show that $V = R \cap \cup(S)$ is a multiplicative subset of R and that $R_V = S$. Because $R_V \subset S$, it suffices to show that $S \subset R_V$. To do this, use the fact that if x is in S, then there are integral ideals I_1 and I_2 in R such that $xR = I_1 I_2^{-1}$ and $(I_1, I_2) = R$.]

(19) Let R be a Dedekind domain with field of quotients K.

(a) Show that R is a PID if and only if $Cl(R) = \{1\}$.

(b) Show that if R is a PID and S is a subring of K containing R, then $S = R_V$ where V is the multiplicative subset $\cup(S) \cup R$ of R.

(20) Show that for an integral domain R, the following statements are equivalent:

(a) R is a Dedekind domain.

(b) Given any nonzero element x in an ideal I of R, there is a y in I such that $(x, y) = I$. [*Hint*: To show that (b) implies (a) it suffices to show that $R_{\mathfrak{P}}$ is a PID for each prime ideal $\mathfrak{P}$ of R. Suppose $\mathfrak{P}$ is a nonzero ideal of R and $M = \mathfrak{P}R_{\mathfrak{P}}$ is the maximal ideal of R. Show that (b) implies that the ideal M of $R_{\mathfrak{P}}$ can be generated by a pair of elements x and y with x in M^2. This implies M is actually principal which in turn implies $R_{\mathfrak{P}}$ is a PID.]

INDEX