GRADUATE STUDIES IN MATHEMATICS 180

# Advanced Modern Algebra

## Third Edition, Part 2

**Joseph J. Rotman**


American Mathematical Society

GRADUATE STUDIES
IN MATHEMATICS 180

# Advanced Modern Algebra

## Third Edition, Part 2

Joseph J. Rotman

American Mathematical Society
Providence, Rhode Island

The 2002 edition of this book was previously published by Pearson Education, Inc.

2010 *Mathematics Subject Classification.* Primary 12-01, 13-01, 14-01, 15-01, 16-01, 18-01, 20-01.

---

For additional information and updates on this book, visit
**www.ams.org/bookpages/gsm-180**

---

**Library of Congress Cataloging-in-Publication Data**

Rotman, Joseph J., 1934–
Advanced modern algebra / Joseph J. Rotman. – Third edition.
volumes cm. – (Graduate studies in mathematics ; volume 165)
Includes bibliographical references and index.
ISBN 978-1-4704-1554-9 (alk. paper : pt. 1)
ISBN 978-1-4704-2311-7 (alk. paper : pt. 2)
1. Algebra. I. Title.

QA154.3.R68 2015
512–dc23 2015019659

---



10 9 8 7 6 5 4 3 2 1 22 21 20 19 18 17

# Contents

# Foreword

My friend and UIUC mathematics department colleague Joe Rotman was completely dedicated to his series of books on algebra. He was correcting his draft of this revision of *Advanced Modern Algebra* during what sadly turned out to be his final hospital visit. At that time, Joe and his family asked me to do what I could to see this close-to-finished version to publication.

Two more friends and colleague of Joe's, Jerry Janusz and Paul Weichsel, joined the project. Jerry did a meticulous line-by-line reading of the manuscript, and all three of us answered questions posed by the AMS editorial staff, based on Arlene O'Sean's very careful reading of the manuscript.

It is clear that this book would have been even richer if Joe had been able to continue to work on it. For example, he planned a chapter on algebraic geometry. We include the first paragraph of that chapter, an example of Joe's distinctly personal writing style, as a small memorial to what might have been.

> Mathematical folklore is the "standard" mathematics "everyone" knows. For example, all mathematics graduate students today are familiar with elementary set theory. But folklore changes with time; elementary set theory was not part of nineteenth-century folklore. When we write a proof, we tacitly use folklore, usually not mentioning it explicitly. That folklore depends on the calendar must be one of the major factors complicating the history of mathematics. We can find primary sources and read, say, publications of Ruffini at the beginning of the 1800s, but we can't really follow his proofs unless we are familiar with his contemporary folklore.

I want to express my thanks to Sergei Gelfand and Arlene O'Sean of the AMS and to Jerry Janusz and Paul Weichsel of UIUC for all their help. Our overriding and mutual goal has been to produce a book which embodies Joe Rotman's intentions.

Bruce Reznick

May 31, 2017

# Preface to Third Edition: Part 2

The second half of this third edition of *Advanced Modern Algebra* has Part 1 as prerequisite. This is not to say that everything there must be completely mastered, but the reader should be familiar with what is there and should not be uncomfortable upon seeing the words *category*, *functor*, *module*, or *Zorn*.

The format of Part 2 is standard, but there are interactions between the different chapters. For example, group extensions and factor sets are discussed in the chapter on groups as well as in the chapter on homology. I am reminded of my experience as an aspiring graduate student. In order to qualify for an advanced degree, we were required to take a battery of written exams, one in each of algebra, analysis, geometry, and topology. At the time, I felt that each exam was limited to its own area, but as I was wrestling with an algebra problem, the only way I could see to solve it was to use a compactness argument. I was uncomfortable: compactness arguments belong in the topology exam, not in the algebra exam! Of course, I was naive. The boundaries of areas dividing mathematics are artificial; they really don't describe *what* is being studied but *how* it is being studied. It is a question of attitude and emphasis. Doesn't every area of mathematics study polynomials? But algebraists and analysts view them from different perspectives. After all, mathematics really is one vast subject, and all its parts and emphases are related.

A word about references in the text. If I mention Theorem C-1.2 or Exercise C-1.27 on page 19, then these are names in Part 2 of the third edition. References to names in Part 1 will have the prefix A- or B- and will say, for example, Theorem A-1.2 in Part 1 or Exercise B-1.27 on page 288 in Part 1. In an exercise set, an asterisk before an exercise, say, *C-1.26, means that this exercise is mentioned elsewhere in the text, usually in a proof.

Thanks goes to Ilya Kapovich, Victoria Corkery, Vincenzo Acciaro, and Stephen Ullom.

*Chapter C-1*

# More Groups

We continue investigating the structure of groups in this chapter, beginning by introducing *group actions*, which essentially show that elements of abstract groups can be viewed as permutations of sets. In Part 1, we saw that finitely generated abelian groups are rather uncomplicated: they are direct sums of cyclic groups. The $p$-primary components of a finite abelian group generalize to Sylow $p$-subgroups of finite nonabelian groups $G$: if $p^e$ is the largest power of a prime $p$ dividing $|G|$, then a Sylow $p$-subgroup is a subgroup of $G$ having order $p^e$. Such subgroups always exist; they may not be unique, but the number of them can be computed up to congruence mod $p$; moreover, any two such subgroups are conjugate and, hence, are isomorphic. The notions of normal series and solvability that arose in Galois theory lead to consideration of solvable groups. Here, we will see that solvable groups (and their cousins *nilpotent* groups) are interesting in their own right, outside of Galois theory. The Jordan–Hölder Theorem shows that simple groups are, in a certain sense, building blocks of finite groups. Consequently, we show that the *projective unimodular groups* $\mathrm{PSL}(n, k)$, where $k$ is a field, are simple groups (in addition to the cyclic groups of prime order and the alternating groups $A_n$ for $n \geq 5$ which we have already proved to be simple). We will give two proofs of this: the first involves looking at the general linear groups $\mathrm{GL}(V)$; the second involves showing that these groups act *multiply transitively* on projective space. Free groups and presentations are introduced next, for they are useful in constructing and describing arbitrary groups. The chapter ends with proofs, using topological methods, of the Schreier–Nielsen Theorem, that every subgroup of a free group is itself a free group, and the Kurosh Theorem, that every subgroup of a free product is itself a free product.

## C-1.1. Group Actions

Group theory originated in Galois theory: groups were subgroups of symmetric groups on roots of polynomials. In contrast, an abstract group is a set $G$ equipped with a binary operation (which satisfies some axioms), and its elements need not

be functions, let alone permutations. But it is fruitful to view group elements as permutations, and the next result shows that abstract groups can be so viewed.

**Theorem C-1.1 (Cayley).** *Every group $G$ is isomorphic to a subgroup of the symmetric group $S_G$. In particular, if $|G| = n$, then $G$ is isomorphic to a subgroup of $S_n$.*

**Proof.** For each $a \in G$, define "translation" $\tau_a \colon G \to G$ by $\tau_a(x) = ax$ for every $x \in G$ (if $a \neq 1$, then $\tau_a$ is not a homomorphism). For $a$, $b \in G$, $(\tau_a\tau_b)(x) = \tau_a(\tau_b(x)) = \tau_a(bx) = a(bx) = (ab)x$, by associativity, so that

$$\tau_a\tau_b = \tau_{ab}.$$

It follows that each $\tau_a$ is a bijection, for its inverse is $\tau_{a^{-1}}$:

$$\tau_a\tau_{a^{-1}} = \tau_{aa^{-1}} = \tau_1 = 1_G = \tau_{a^{-1}a},$$

and so $\tau_a \in S_G$.

Define $\varphi : G \to S_G$ by $\varphi(a) = \tau_a$. Rewriting,

$$\varphi(a)\varphi(b) = \tau_a\tau_b = \tau_{ab} = \varphi(ab),$$

so that $\varphi$ is a homomorphism. Finally, $\varphi$ is an injection. Suppose that $\varphi(a) = 1_G$. Now $\varphi(a)(x) = \tau_a(x) = ax$; but $\varphi(a)(x) = 1_G(x) = x$ for all $x \in G$; that is, $a = 1$.

The last statement follows from Exercise A-4.46 on page 157 in Part 1, which says that if $X$ is a set with $|X| = n$, then $S_X \cong S_n$. •

The reader may note, in the proof of Cayley's Theorem, that the permutation $\tau_a \colon x \mapsto ax$ is just the $a$th row of the multiplication table of $G$.

To tell the truth, Cayley's Theorem itself is only mildly interesting, but a generalization having the identical proof is more useful.

**Theorem C-1.2 (Representation on Cosets).** *If $G$ is a finite group and $H$ is a subgroup of index $n$, then there exists a homomorphism $\varphi \colon G \to S_n$ with $\ker\varphi \subseteq H$.*

**Proof.** We denote the family of all the left cosets of $H$ in $G$ by $G/H$, even though $H$ may not be a normal subgroup.

For each $a \in G$, define "translation" $\tau_a \colon G/H \to G/H$ by $\tau_a(xH) = axH$ for every $x \in G$. For $a$, $b \in G$,

$$(\tau_a\tau_b)(xH) = \tau_a(\tau_b(xH)) = \tau_a(bxH) = a(bxH) = (ab)xH,$$

by associativity, so that

$$\tau_a\tau_b = \tau_{ab}.$$

It follows that each $\tau_a$ is a bijection, for its inverse is $\tau_{a^{-1}}$:

$$\tau_a\tau_{a^{-1}} = \tau_{aa^{-1}} = \tau_1 = 1_{G/H} = \tau_{a^{-1}}\tau_a,$$

and so $\tau_a \in S_{G/H}$. Define $\varphi : G \to S_{G/H}$ by $\varphi(a) = \tau_a$. Rewriting,

$$\varphi(a)\varphi(b) = \tau_a\tau_b = \tau_{ab} = \varphi(ab),$$

so that $\varphi$ is a homomorphism. Finally, if $a \in \ker\varphi$, then $\varphi(a) = 1_{G/H}$, so that $\tau_a(xH) = axH = xH$ for all $x \in G$; in particular, when $x = 1$, we have $aH = H$.

Hence, $a \in H$, by Lemma A-4.42 in Part 1, and $\varphi$ is an injection. The result follows from Exercise A-4.46 on page 157 in Part 1, for $|G/H| = n$, and so $S_{G/H} \cong S_n$. •

When $H = \{1\}$, this is the Cayley Theorem, for then $\ker \varphi = \{1\}$ and $\varphi$ is an injection.

We are now going to classify all groups of order up to 7. By Example A-4.55 in Part 1, every group of prime order $p$ is isomorphic to $\mathbb{Z}_p$, and so, up to isomorphism, there is just one group of order $p$. Of the possible orders through 7, four are primes, namely, 2, 3, 5, and 7, and so we need look only at orders 4 and 6.

**Proposition C-1.3.** *Every group $G$ of order 4 is abelian, and either $G \cong \mathbb{Z}_4$ or $G \cong \mathbf{V}$, the four-group. Moreover, $\mathbb{Z}_4$ and $\mathbf{V}$ are not isomorphic.*

**Proof.** By Lagrange's Theorem, every element in $G$ has order 1, 2, or 4. If there is an element of order 4, then $G$ is cyclic. Otherwise, $x^2 = 1$ for all $x \in G$, so that Exercise A-4.31 on page 138 in Part 1 shows that $G$ is abelian.

Assume that $G$ is not cyclic. If distinct elements $x$ and $y$ in $G$ are chosen, neither being 1, then we quickly check that $xy \notin \{1, x, y\}$; hence, $G = \{1, x, y, xy\}$. It is easy to see that the bijection $f\colon G \to \mathbf{V}$, defined by $f(1) = 1$, $f(x) = (1\ 2)(3\ 4)$, $f(y) = (1\ 3)(2\ 4)$, and $f(xy) = (1\ 4)(2\ 3)$, is an isomorphism, for the product of any two nonidentity elements is the third one. We have already seen, in Example A-4.56 in Part 1, that $\mathbb{Z}_4 \ncong \mathbf{V}$. •

Another proof of Proposition C-1.3 uses Cayley's Theorem: $G$ is isomorphic to a subgroup of $S_4$, and it is not too difficult to show, using Table 1 on page 121 in Part 1, that every subgroup of $S_4$ of order 4 is either cyclic or isomorphic to the four-group.

**Proposition C-1.4.** *If $G$ is a group of order 6, then $G$ is isomorphic to either $\mathbb{Z}_6$ or $S_3$.*[1] *Moreover, $\mathbb{Z}_6$ and $S_3$ are not isomorphic.*

**Proof.**[2] By Lagrange's Theorem, the only possible orders of nonidentity elements are 2, 3, and 6. Of course, $G \cong \mathbb{Z}_6$ if $G$ has an element of order 6. Now Exercise A-4.33 on page 138 in Part 1 shows that $G$ must contain an element of order 2, say, $t$. We distinguish two cases.

**Case 1**. $G$ is abelian.

If there is a second element of order 2, say, $a$, then it is easy to see, using $at = ta$, that $H = \{1, a, t, at\}$ is a subgroup of $G$. This contradicts Lagrange's

[1] Cayley states this proposition in an article he wrote in 1854. However, in 1878, in the *American Journal of Mathematics*, he wrote, "The general problem is to find all groups of a given order $n$; ... if $n = 6$, there are three groups; a group

$$1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5 \quad (\alpha^6 = 1),$$

and two more groups

$$1, \beta, \beta^2, \alpha, \alpha\beta, \alpha\beta^2 \qquad (\alpha^2 = 1, \beta^3 = 1),$$

viz., in the first of these $\alpha\beta = \beta\alpha$ while in the other of them, we have $\alpha\beta = \beta^2\alpha, \alpha\beta^2 = \beta\alpha$." Cayley's list is $\mathbb{Z}_6$, $\mathbb{Z}_2 \times \mathbb{Z}_3$, and $S_3$; of course, $\mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_6$. Even Homer nods.

[2] We give another proof in Proposition C-1.131.

Theorem, because 4 is not a divisor of 6. It follows that $G$ must contain an element $b$ of order 3.[3] But $tb$ has order 6, by Proposition A-4.86 in Part 1. Therefore, $G$ is cyclic if it is abelian.

**Case 2**. $G$ is not abelian.

If $G$ has no elements of order 3, then $x^2 = 1$ for all $x \in G$, and $G$ is abelian, by Exercise A-4.31 on page 138 in Part 1. Therefore, $G$ contains an element $s$ of order 3 as well as the element $t$ of order 2.

Now $|\langle s\rangle| = 3$, so that $[G : \langle s\rangle] = |G|/|\langle s\rangle| = 6/3 = 2$, and so $\langle s\rangle$ is a normal subgroup of $G$, by Proposition A-4.65 in Part 1. Since $t = t^{-1}$, we have $tst \in \langle s\rangle$; hence, $tst = s^i$ for $i = 0$, 1, or 2. Now $i \neq 0$, for $tst = s^0 = 1$ implies $s = 1$. If $i = 1$, then $s$ and $t$ commute, and this gives $st$ of order 6, as in Case 1 (which forces $G$ to be cyclic, hence abelian, contrary to our present hypothesis). Therefore, $tst = s^2 = s^{-1}$.

We construct an isomorphism $G \to S_3$. Let $H = \langle t\rangle$, and consider the homomorphism $\varphi : G \to S_{G/H}$ given by

$$\varphi(g) : xH \mapsto gxH.$$

By Theorem C-1.2, $\ker\varphi \subseteq H$, so that either $\ker\varphi = \{1\}$ (and $\varphi$ is injective) or $\ker\varphi = H = \langle t\rangle$. Now $G/H = \{H, sH, s^2H\}$ and, in two-rowed permutation notation,

$$\varphi(t) = \begin{pmatrix} H & sH & s^2H \\ tH & tsH & ts^2H \end{pmatrix}.$$

If $\varphi(t)$ is the identity permutation, then $tsH = sH$, so that $s^{-1}ts \in H = \langle t\rangle = \{1, t\}$, by Lemma A-4.42 in Part 1. But now $s^{-1}ts = t$ (it cannot be 1); hence, $ts = st$, contradicting $t$ and $s$ not commuting. Thus, $t \notin \ker\varphi$, and $\varphi\colon G \to S_{G/H} \cong S_3$ is an injective homomorphism. Since both $G$ and $S_3$ have order 6, $\varphi$ must be a bijection, and so $G \cong S_3$.

It is clear that $\mathbb{Z}_6 \ncong S_3$, for $\mathbb{Z}_6$ is abelian and $S_3$ is not. •

One consequence of this result is that $\mathbb{Z}_6 \cong \mathbb{Z}_2 \times \mathbb{Z}_3$.

Classifying groups of order 8 is more difficult, for we have not yet developed enough theory. Theorem C-1.132 below says there are only five nonisomorphic groups of order 8; three are abelian: $\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, and $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$; two are nonabelian: the dihedral group $D_8$ and the quaternions $\mathbf{Q}$.

We can continue this discussion for larger orders, but things soon get out of hand, as Table 1 shows (the number of groups of order 1024 can be found in [**19**]). Making a telephone directory of groups is not the way to study them.

Groups arose by abstracting the fundamental properties enjoyed by permutations. But there is an important feature of permutations that the axioms do not mention: permutations are functions. We shall see that there are interesting consequences when this feature is restored.

[3] We will soon prove Cauchy's Theorem which says that a finite group whose order is divisible by a prime $p$ must contain an element of order $p$. In particular, groups of order 6 must contain an element of order 3.

| Order of Group | Number of Groups |
|---|---|
| 2 | 1 |
| 4 | 2 |
| 8 | 5 |
| 16 | 14 |
| 32 | 51 |
| 64 | 267 |
| 128 | 2,328 |
| 256 | 56,092 |
| 512 | 10,494,213 |
| 1024 | 49,487,365,422 |

**Table 1.** Too many 2-groups.

**Definition.** A group $G$ ***acts*** on a set $X$ if there is a function $G \times X \to X$, denoted by $(g, x) \mapsto gx$, such that

(i) $(gh)x = g(hx)$ for all $g$, $h \in G$ and $x \in X$,

(ii) $1x = x$ for all $x \in X$, where 1 is the identity in $G$.

If $G$ acts on $X$, we call $X$ a ***$G$-set*** and we call $|X|$ the ***degree*** of $X$.

**Remark.** In the definition of action just given, the elements of $G$ act on the left. As for modules, there is also a "right" version of $G$-set which is sometimes convenient. Define a ***right action*** to be a function $X \times G \to X$, denoted by $(x, g) \mapsto xg$, satisfying

(i) $x(gh) = (xg)h$ for all $g$, $h \in G$ and $x \in X$,

(ii) $x1 = x$ for all $x \in X$, where 1 is the identity in $G$.

Given a right $G$-set, we can make it into a left $G$-set by defining $gx = xg^{-1}$ for $x \in X$ and $g \in G$.

It is easy to see that every right $G$-set gives rise to a (left) $G$-set if we define $G\colon G \times X \to X$ by $g^{-1}x = xg$. ◀

We now show that an action of a group $G$ on a set $X$ is merely another way of viewing a homomorphism $G \to S_X$, where $S_X$ is the symmetric group on $X$.

**Theorem C-1.5.** *Every action $\alpha\colon G \times X \to X$ of a group $G$ on a set $X$ determines a homomorphism $G \to S_X$. Conversely, every homomorphism $G \to S_X$ makes $X$ into a $G$-set.*

**Proof.** Given an action $\alpha$, fixing the first variable, say, $g$, gives a function $\alpha_g\colon X \to X$, namely, $\alpha_g\colon x \mapsto gx$. This function is a permutation of $X$, for its inverse is $\alpha_{g^{-1}}$: $\alpha_g\alpha_{g^{-1}} = \alpha_1 = 1_X = \alpha_{g^{-1}}\alpha_g$. It is easy to see that $\alpha\colon G \to S_X$, defined by $\alpha\colon g \mapsto \alpha_g$, is a homomorphism.

Conversely, given any homomorphism $\varphi\colon G \to S_X$, define $\alpha\colon G \times X \to X$ by $\alpha(g, x) = \varphi(g)(x)$. It is easy to see that $\alpha$ is an action. •

We say that a $G$-set $X$ is ***faithful*** if the homomorphism $\alpha\colon G \to S_X$ is an injection; that is, $G$ can be viewed (via $\alpha$) as a subgroup of $S_X$.

The following definitions are fundamental.

**Definition.** If $G$ acts on $X$ and $x \in X$, then the ***orbit*** of $x$, denoted by $\mathcal{O}(x)$, is the subset

$$\mathcal{O}(x) = \{gx : g \in G\} \subseteq X.$$

We say that $G$ acts ***transitively*** on a set $X$ if there is only one orbit.

The ***stabilizer*** of $x \in X$, denoted by $G_x$, is the subgroup

$$G_x = \{g \in G : gx = x\} \subseteq G.$$

The ***orbit space***, denoted by $X/G$, is the set of all the orbits.

If $G$ acts on a set $X$, define a relation on $X$ by $x \equiv y$ in case there exists $g \in G$ with $y = gx$. It is easy to see that this is an equivalence relation whose equivalence classes are the orbits. The orbit space is the family of equivalence classes.

Cayley's Theorem says that a group $G$ acts on itself by (left) translations, and its generalization, Theorem C-1.2, shows that $G$ also acts on the family of (left) cosets of a subgroup $H$ by (left) translations.

**Example C-1.6.** We show that $G$ acts on itself by conjugation. For each $g \in G$, define $\alpha_g\colon G \to G$ to be conjugation

$$\alpha_g(x) = gxg^{-1}.$$

To verify axiom (i), note that for each $x \in G$,

$$\begin{aligned}(\alpha_g\alpha_h)(x) &= \alpha_g(\alpha_h(x)) = \alpha_g(hxh^{-1})\\ &= g(hxh^{-1})g^{-1} = (gh)x(gh)^{-1} = \alpha_{gh}(x).\end{aligned}$$

Therefore, $\alpha_g\alpha_h = \alpha_{gh}$. To prove axiom (ii), note that $\alpha_1(x) = 1x1^{-1} = x$ for each $x \in G$, and so $\alpha_1 = 1_G$. ◀

Let us find some orbits and stabilizers.

**Example C-1.7.** Cayley's Theorem says that $G$ acts on itself by translations: $\tau_g\colon a \mapsto ga$. If $a \in G$, then the orbit $\mathcal{O}(a) = G$, for if $b \in G$, then $b = (ba^{-1})a = \tau_{ba^{-1}}(a)$. The stabilizer $G_a$ of $a \in G$ is $\{1\}$, for if $a = \tau_g(a) = ga$, then $g = 1$.

More generally, $G$ acts transitively on $G/H$ (the family of (left) cosets of a (not necessarily normal) subgroup $H$) by translations $\tau_g : aH \mapsto gaH$. The orbit $\mathcal{O}(aH) = G/H$, for if $bH \in G/H$, then $\tau_{ba^{-1}}\colon aH \mapsto bH$. The stabilizer $G_{aH}$ of the coset $aH$ is $aHa^{-1}$, for $gaH = aH$ if and only if $a^{-1}ga \in H$ if and only if $g \in aHa^{-1}$. ◀

**Example C-1.8.** Let $X = \{1, 2, \ldots, n\}$, let $\alpha \in S_n$, and define the obvious action of the cyclic group $G = \langle\alpha\rangle$ on $X$ by $\alpha^k \cdot i = \alpha^k(i)$. If $i \in X$, then

$$\mathcal{O}(i) = \{\alpha^k(i) : 0 \le k < |G|\}.$$

Suppose the complete factorization of $\alpha$ is $\alpha = \beta_1 \cdots \beta_{t(\alpha)}$ and $i = i_1$ is moved by $\alpha$. If the cycle involving $i_1$ is $\beta_j = (i_1\ i_2\ \ldots\ i_r)$, then the proof of Theorem A-4.4 in

Part 1 shows that $i_{k+1} = \alpha^k(i_1)$ for all $k < r$, while $\alpha^r(i_1) = \alpha(\alpha^{r-1}(i_1)) = \alpha(i_r) = i_1$. Therefore,

$$\mathcal{O}(i) = \{i_1, i_2, \ldots, i_r\},$$

where $i = i_1$. It follows that $|\mathcal{O}(i)| = r$, the length of the cycle $\beta_j$.

The stabilizer $G_i$ of a number $i$ is $G$ if $\alpha$ fixes $i$; however, if $\alpha$ moves $i$, then $G_i$ depends on the size of the orbit $\mathcal{O}(i)$. For example, if $\alpha = (1\ 2\ 3)(4\ 5)(6)$, then $G_6 = G$, $G_1 = \langle \alpha^3 \rangle$, and $G_4 = \langle \alpha^2 \rangle$. ◀

**Example C-1.9.** Let $k$ be a field and let $f(x) \in k[x]$ have distinct roots. If $E/k$ is a splitting field of $f$, then $\mathrm{Gal}(E/k)$ acts faithfully on the set $X$ of the roots of $f$. Moreover, $f$ is irreducible if and only if the action of $\mathrm{Gal}(E/k)$ on $X$ is transitive (Rotman [**188**], p. 95). ◀

**Example C-1.10.** The general linear group $\mathrm{GL}(\mathbb{R}^n)$ acts on $\mathbb{R}^n - \{0\}$: if $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is a nonsingular linear transformation, then the action is given by $v \mapsto Tv$. This action is transitive, and if $v \in \mathbb{R}^n$ is nonzero, then the stabilizer of $v$ is the line containing $v$ and the origin.

There is a matrix version of this action. If $\mathrm{GL}(n, \mathbb{R})$ is the multiplicative group of all nonsingular $n \times n$ real matrices, then $\mathrm{GL}(n, \mathbb{R})$ acts once a basis of $\mathbb{R}^n$ is chosen. In particular, if $e_1, \ldots, e_n$ is the standard basis, then each nonzero $v \in \mathbb{R}^n$ has coordinates $v = (x_1, \ldots, x_n)$, and if $A \in \mathrm{GL}(n, \mathbb{R})$, then $Av$ is the matrix product $Av^\top$. ◀

**Example C-1.11.** Let $X = \{v_0,\ v_1,\ v_2,\ v_3\}$ be the vertices of a square, and let $G$ be the dihedral group $D_8$ acting on $X$, as in Figure C-1.1 (for clarity, the vertices in the figure are labeled 0, 1, 2, 3 instead of $v_0$, $v_1$, $v_2$, $v_3$). The lines in the bottom four squares are axes of symmetry.

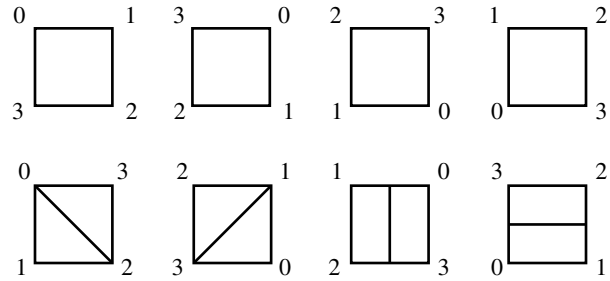



**Figure C-1.1.** Dihedral group $D_8$.

Thus,

$$\begin{aligned} G &= \{\text{rotations}\} \cup \{\text{reflections}\} \\ &= \{(1), (0\ 1\ 2\ 3), (0\ 2)(1\ 3), (0\ 3\ 2\ 1)\} \cup \{(1\ 3), (0\ 2), (0\ 1)(2\ 3), (0\ 3)(1\ 2)\}. \end{aligned}$$

For each vertex $v_i \in X$, there is some $g \in G$ with $gv_0 = v_i$; therefore, $\mathcal{O}(v_0) = X$ and $D_8$ acts transitively.

What is the stabilizer $G_{v_0}$ of $v_0$? Aside from the identity, only one $g \in D_8$ fixes $v_0$, namely, $g = (1\ 3)$; hence, $G_{v_0}$ is a subgroup of order 2. (This example can be generalized to the dihedral group $D_{2n}$ acting on a regular $n$-gon.) ◀

**Example C-1.12.** When a group $G$ acts on itself by conjugation, then the orbit $\mathcal{O}(x)$ of $x \in G$ is commonly denoted by

$$x^G = \{y \in G : y = axa^{-1} \text{ for some } a \in G\}.$$

The orbit $\mathcal{O}(x)$ is called the ***conjugacy class*** of $x$ (we have already mentioned conjugacy classes in Exercise A-4.49 on page 157 in Part 1). Theorem A-4.7 in Part 1 shows that if $\alpha \in S_n$, then the conjugacy class of $\alpha$ consists of all the permutations in $S_n$ having the same cycle structure as $\alpha$. Also, an element $z$ lies in the center $Z(G)$ if and only if $z^G = \{z\}$; that is, no other elements in $G$ are conjugate to $z$.

If $x \in G$, then the stabilizer $G_x$ of $x$ is

$$C_G(x) = \{g \in G : gxg^{-1} = x\}.$$

This subgroup of $G$, consisting of all $g \in G$ that commute with $x$, is called the ***centralizer*** of $x$ in $G$. ◀

**Example C-1.13.** Every group $G$ acts by conjugation on the set $X$ of all its subgroups: if $a \in G$, then $a$ acts by $H \mapsto aHa^{-1}$, where $H \subseteq G$.

If $H$ is a subgroup of a group $G$, then a ***conjugate*** of $H$ is a subgroup of $G$ of the form

$$aHa^{-1} = \{aha^{-1} : h \in H\},$$

where $a \in G$. Conjugation $h \mapsto aha^{-1}$ is an injection $H \to G$ with image $aHa^{-1}$. It follows that conjugate subgroups of $G$ are isomorphic; the inverse is given by $g \mapsto a^{-1}ga$. For example, in $S_3$, all cyclic subgroups of order 2 are conjugate (for their generators are conjugate).

The orbit of a subgroup $H$ consists of all its conjugates; notice that $H$ is the only element in its orbit if and only if $H \lhd G$; that is, $aHa^{-1} = H$ for all $a \in G$. The stabilizer of $H$ is

$$N_G(H) = \{g \in G : gHg^{-1} = H\}.$$

This subgroup of $G$ is called the ***normalizer*** of $H$ in $G$. Of course, $H \lhd N_G(H)$; indeed, the normalizer is the largest subgroup of $G$ in which $H$ is normal. ◀

We have already defined the centralizer of an element; we now define the centralizer of a subgroup.

**Definition.** If $H$ is a subgroup of a group $G$, then the ***centralizer*** of $H$ in $G$ is

$$C_G(H) = \{g \in G : gh = hg \text{ for all } h \in H\}.$$

It is easy to see that $C_G(H)$ is a subgroup of $G$, and $C_G(G) = Z(G)$. Note that $C_G(H) \subseteq N_G(H)$.

Recall that an automorphism $\alpha$ of a group $G$ is *inner* if it is conjugation by some $a \in G$; that is, $\alpha = \alpha_a \colon x \mapsto axa^{-1}$.

**Proposition C-1.14 ($N/C$ Lemma).**

(i) *If $H \subseteq G$, then $C_G(H) \lhd N_G(H)$ and there is an imbedding*

$$N_G(H)/C_G(H) \hookrightarrow \mathrm{Aut}(H).$$

(ii) *$G/Z(G) \cong \mathrm{Inn}(G)$, where $\mathrm{Inn}(G)$ is the subgroup of $\mathrm{Aut}(G)$ consisting of all the inner automorphisms.*

**Proof.**

(i) If $a \in G$, denote conjugation $g \mapsto aga^{-1}$ by $\gamma_a$. Define $\varphi\colon N_G(H) \to \mathrm{Aut}(H)$ by $\varphi\colon a \mapsto \gamma_a|H$. Note that $\varphi$ is well-defined, for $\gamma_a|H \in \mathrm{Aut}(H)$ because $a \in N_G(H)$. It is routine to check that $\varphi$ is a homomorphism. Now the following statements are equivalent: $a \in \ker\varphi$; $\gamma_a|H = 1_H$; $aha^{-1} = h$ for all $h \in H$; $a \in C_G(H)$. The First Isomorphism Theorem gives $C_G(H) \lhd N_G(H)$ and $N_G(H)/C_G(H) \cong \mathrm{im}\,\varphi \subseteq \mathrm{Aut}(H)$.

(ii) In the special case $H = G$, we have $N_G(H) = G$, $C_G(H) = Z(G)$, and $\mathrm{im}\,\varphi = \mathrm{Inn}(G)$. •

**Remark.** We claim that $\mathrm{Inn}(G) \lhd \mathrm{Aut}(G)$. If $\varphi \in \mathrm{Aut}(G)$ and $g \in G$, then

$$\varphi\gamma_a\varphi^{-1}\colon g \mapsto \varphi^{-1}g \mapsto a\varphi^{-1}ga^{-1} \mapsto \varphi(a)g\varphi(a^{-1}).$$

Thus, $\varphi\gamma_a\varphi^{-1} = \gamma_{\varphi(a)} \in \mathrm{Inn}(G)$. Recall that an automorphism is called ***outer*** if it is not inner; the ***outer automorphism group*** is defined by

$$\mathrm{Out}(G) = \mathrm{Aut}(G)/\mathrm{Inn}(G). \quad \blacktriangleleft$$

**Proposition C-1.15.** *If $G$ acts on a set $X$, then $X$ is the disjoint union of the orbits. If $X$ is finite, then*

$$|X| = \sum_i |\mathcal{O}(x_i)|,$$

*where one $x_i$ is chosen from each orbit.*

**Proof.** As we have mentioned earlier, the relation on $X$, given by $x \equiv y$ if there exists $g \in G$ with $y = gx$, is an equivalence relation whose equivalence classes are the orbits. Therefore, the orbits partition $X$.

The count given in the second statement is correct: since the orbits are disjoint, no element in $X$ is counted twice. •

Here is the connection between orbits and stabilizers.

**Theorem C-1.16 (Orbit-Stabilizer Theorem).** *If $G$ acts on a set $X$ and $x \in X$, then*

$$|\mathcal{O}(x)| = [G : G_x],$$

*the index of the stabilizer $G_x$ in $G$.*

**Proof.** Let $G/G_x$ denote the family of all the left cosets of $G_x$ in $G$. We will exhibit a bijection $\varphi\colon G/G_x \to \mathcal{O}(x)$, and this will give the result, since $|G/G_x| = [G : G_x]$. Define $\varphi\colon gG_x \mapsto gx$. Now $\varphi$ is well-defined: if $gG_x = hG_x$, then $h = gf$ for some $f \in G_x$; that is, $fx = x$; hence, $hx = gfx = gx$. Now $\varphi$ is an injection: if

$gx = \varphi(gG_x) = \varphi(hG_x) = hx$, then $h^{-1}gx = x$; hence, $h^{-1}g \in G_x$, and $gG_x = hG_x$. Lastly, $\varphi$ is a surjection: if $y \in \mathcal{O}(x)$, then $y = gx$ for some $g \in G$, and so $y = \varphi(gG_x)$. •

In Example C-1.11, $D_8$ acting on the four corners of a square, we saw that $|\mathcal{O}(v_0)| = 4$, $|G_{v_0}| = 2$, and $[G : G_{v_0}] = 8/2 = 4$. In Example C-1.8, $G = \langle\alpha\rangle \subseteq S_n$ acting on $X = \{1, 2, \ldots, n\}$, we saw that if $\alpha = \beta_1 \cdots \beta_t$ is the complete factorization into disjoint cycles and $\ell$ occurs in the $r_j$-cycle $\beta_j$, then $r_j = |\mathcal{O}(\ell)|$. Theorem C-1.16 says that $r_j$ is a divisor of the order $k$ of $\alpha$ (but Theorem A-4.24 in Part 1 tells us more: $k$ is the lcm of the lengths of the cycles occurring in the factorization).

**Corollary C-1.17.** *If a finite group $G$ acts on a finite set $X$, then the number of elements in any orbit is a divisor of $|G|$.*

**Proof.** This follows at once from Lagrange's Theorem and Theorem C-1.16. •

Table 1 on page 121 in Part 1 displays the number of permutations in $S_4$ of each cycle structure; these numbers are 1, 6, 8, 6, 3. Note that each of these numbers is a divisor of $|S_4| = 24$. Table 2 on page 121 in Part 1 shows that the corresponding numbers for $S_5$ are 1, 10, 20, 30, 24, 20, and 15, and these are all divisors of $|S_5| = 120$. We now recognize these subsets as being conjugacy classes, and the next corollary explains why these numbers divide the group order.

**Corollary C-1.18.** *If $x$ lies in a finite group $G$, then the number of conjugates of $x$ is the index of its centralizer*:

$$|x^G| = [G : C_G(x)],$$

*and hence it is a divisor of $|G|$.*

**Proof.** As in Example C-1.12, the orbit of $x$ is its conjugacy class $x^G$, and the stabilizer $G_x$ is the centralizer $C_G(x)$. •

**Corollary C-1.19.** *If $H$ is a subgroup of a finite group $G$, then the number of conjugates of $H$ in $G$ is $[G : N_G(H)]$.*

**Proof.** As in Example C-1.13, the orbit of $H$ is the family of all its conjugates, and the stabilizer is its normalizer $N_G(H)$. •

When we began classifying groups of order 6, it would have been helpful to be able to assert that any such group has an element of order 3 (we were able to use an earlier exercise to assert the existence of an element of order 2). We now prove that if $p$ is a prime divisor of $|G|$, where $G$ is a finite group, then $G$ contains an element of order $p$ (Proposition A-4.81 in Part 1 proved the special case of this when $G$ is abelian).

**Theorem C-1.20 (Cauchy).** *If $G$ is a finite group whose order is divisible by a prime $p$, then $G$ contains an element of order $p$.*

**Proof.** We prove the theorem by induction on $m \geq 1$, where $|G| = pm$. The base step $m = 1$ is true, for Lagrange's Theorem shows that every nonidentity element in a group of order $p$ has order $p$.

Let us now prove the inductive step. If $x \in G$, then the number of conjugates of $x$ is $|x^G| = [G : C_G(x)]$, where $C_G(x)$ is the centralizer of $x$ in $G$. If $x \notin Z(G)$, then $x^G$ has more than one element, and so $|C_G(x)| < |G|$. We are done if $p \mid |C_G(x)|$, for the inductive hypothesis gives an element of order $p$ in $C_G(x) \subseteq G$. Therefore, we may assume that $p \nmid |C_G(x)|$ for all noncentral $x \in G$. Better, since $p$ is prime and $|G| = [G : C_G(x)]|C_G(x)|$, Euclid's Lemma gives

$$p \mid [G : C_G(x)].$$

After recalling that $Z(G)$ consists of all those elements $x \in G$ with $|x^G| = 1$, we may use Proposition C-1.15 to see that

$$|G| = |Z(G)| + \sum_i [G : C_G(x_i)], \tag{1}$$

where one $x_i$ is selected from each conjugacy class having more than one element. Since $|G|$ and all $[G : C_G(x_i)]$ are divisible by $p$, it follows that $|Z(G)|$ is divisible by $p$. But $Z(G)$ is abelian, and so Proposition A-4.81 in Part 1 says that $Z(G)$, and hence $G$, contains an element of order $p$. •

The next definition, which specializes Proposition C-1.15, gives a name to Eq. (1) in the proof of Cauchy's Theorem.

**Definition.** The ***class equation*** of a finite group $G$ is

$$|G| = |Z(G)| + \sum_i [G : C_G(x_i)],$$

where one $x_i$ is selected from each conjugacy class having more than one element.

**Definition.** If $p$ is prime, then a group $G$ is called a ***p-group*** if every element has order a power of $p$.

**Proposition C-1.21.** *A finite group $G$ is a $p$-group if and only if $|G| = p^n$ for some $n \geq 0$.*

**Proof.** Let $G$ be a finite $p$-group. If $|G| \neq p^n$, then there is some prime $q \neq p$ with $q \mid |G|$. By Cauchy's Theorem, $G$ contains an element of order $q$, a contradiction. The converse follows from Lagrange's Theorem. •

We have seen examples of groups whose center is trivial; for example, $Z(S_3) = \{(1)\}$. Finite $p$-groups, however, are never centerless.

**Theorem C-1.22.** *If $p$ is prime and $G \neq \{1\}$ is a finite $p$-group, then the center of $G$ is nontrivial; $Z(G) \neq \{1\}$.*

**Proof.** Consider the class equation

$$|G| = |Z(G)| + \sum_i [G : C_G(x_i)].$$

Each $C_G(x_i)$ is a proper subgroup of $G$, for $x_i \notin Z(G)$. Since $G$ is a $p$-group, $[G : C_G(x_i)]$ is a divisor of $|G|$, hence is itself a power of $p$. Thus, $p$ divides each of the terms in the class equation other than $|Z(G)|$, and so $p \mid |Z(G)|$ as well. Therefore, $Z(G) \neq \{1\}$. •

McLain gave an example of an infinite $p$-group $G$ with $Z(G) = \{1\}$ (see Robinson [**181**], p. 362).

**Corollary C-1.23.** *If $p$ is prime, then every group $G$ of order $p^2$ is abelian.*

**Proof.** If $G$ is not abelian, then its center $Z(G)$ is a proper subgroup, so that $|Z(G)| = 1$ or $p$, by Lagrange's Theorem. But Theorem C-1.22 says that $Z(G) \neq \{1\}$, and so $|Z(G)| = p$. The center is always a normal subgroup, so that the quotient $G/Z(G)$ is defined; it has order $p$, and hence $G/Z(G)$ is cyclic. This contradicts Exercise A-4.79 on page 172 in Part 1. •

We note that Corollary C-1.23 is false for higher powers of $p$; for example, the subgroup of $\mathrm{GL}(3, \mathbb{F}_p)$ consisting of all upper triangular matrices

$$\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix},$$

for $a, b, c \in \mathbb{F}_p$, is a nonabelian group of order $p^3$.

**Example C-1.24.** Who would have guessed that Cauchy's Theorem (if $G$ is a group whose order is a multiple of a prime $p$, then $G$ has an element of order $p$) and Fermat's Theorem (if $p$ is prime, then $a^p \equiv a \bmod p$) can be proved simultaneously? The elementary yet ingenious proof of Cauchy's Theorem is due to McKay in 1959 (Montgomery–Ralston [**158**], p. 41); A. Mann showed me that McKay's argument also proves Fermat's Theorem.

If $G$ is a finite group and $p$ is prime, denote the cartesian product of $p$ copies of $G$ by $G^p$, and define

$$X = \{(a_0, a_1, \ldots, a_{p-1}) \in G^p : a_0 a_1 \cdots a_{p-1} = 1\}.$$

Note that $|X| = |G|^{p-1}$, for having chosen the last $p-1$ entries arbitrarily, the 0th entry must equal $(a_1 a_2 \cdots a_{p-1})^{-1}$. Introduce an action of $\mathbb{Z}_p$ on $X$ by defining, for $0 \leq i \leq p-1$,

$$[i](a_0, a_1, \ldots, a_{p-1}) = (a_i, a_{i+1}, \ldots, a_{p-1}, a_0, a_1, \ldots, a_{i-1}).$$

The product of the entries in the new $p$-tuple is a conjugate of $a_0 a_1 \cdots a_{p-1}$:

$$a_i a_{i+1} \cdots a_{p-1} a_0 a_1 \cdots a_{i-1} = (a_0 a_1 \cdots a_{i-1})^{-1} (a_0 a_1 \cdots a_{p-1})(a_0 a_1 \cdots a_{i-1}).$$

This conjugate is 1 (for $a_0 a_1 \cdots a_{p-1} = 1$), and so $[i](a_0, a_1, \ldots, a_{p-1}) \in X$. By Corollary C-1.17, the size of every orbit of $X$ is a divisor of $|\mathbb{Z}_p| = p$; since $p$ is prime, these sizes are either 1 or $p$. Now orbits with just one element consist of a

$p$-tuple all of whose entries $a_i$ are equal, for all cyclic permutations of the $p$-tuple are the same. In other words, such an orbit corresponds to an element $a \in G$ with $a^p = 1$. Clearly, $(1, 1, \ldots, 1)$ is such an orbit; if it were the only such orbit, then we would have

$$|G|^{p-1} = |X| = 1 + kp$$

for some $k \geq 0$; that is, $|G|^{p-1} \equiv 1 \bmod p$. If $p$ is a divisor of $|G|$, then we have a contradiction, for $|G|^{p-1} \equiv 0 \bmod p$. We have thus proved Cauchy's Theorem: if a prime $p$ is a divisor of $|G|$, then $G$ has an element of order $p$.

Choose a group $G$ of order $n$, say, $G = \mathbb{Z}_n$, where $n$ is not a multiple of $p$. By Lagrange's Theorem, $G$ has no elements of order $p$, so that if $a^p = 1$, then $a = 1$. Therefore, the only orbit in $G^p$ of size 1 is $(1, 1, \ldots, 1)$, and so

$$n^{p-1} = |G|^{p-1} = |X| = 1 + kp;$$

that is, if $p$ is not a divisor of $n$, then $n^{p-1} \equiv 1 \bmod p$. Multiplying both sides by $n$, we have $n^p \equiv n \bmod p$, a congruence also holding when $p$ is a divisor of $n$; this is Fermat's Theorem. ◀

We have seen, in Proposition A-4.67 in Part 1, that $A_4$ is a group of order 12 having no subgroup of order 6. Thus, the assertion that if $d$ is a divisor of $|G|$, then $G$ must have a subgroup of order $d$, is false. However, this assertion is true when $G$ is a finite $p$-group.

**Proposition C-1.25.** *If $G$ is a $p$-group of order $p^\ell$, then $G$ has a normal subgroup of order $p^k$ for every $k \leq \ell$.*

**Proof.** We prove the result by induction on $\ell \geq 0$. The base step is obviously true, and so we proceed to the inductive step. By Theorem C-1.22, the center of $G$ is a nontrivial normal subgroup: $Z(G) \neq \{1\}$. Let $Z \subseteq Z(G)$ be a subgroup of order $p$; as any subgroup of $Z(G)$, the subgroup $Z$ is a normal subgroup of $G$. If $k \leq \ell$, then $p^{k-1} \leq p^{\ell-1} = |G/Z|$. By induction, $G/Z$ has a normal subgroup $H^*$ of order $p^{k-1}$. The Correspondence Theorem says there is a subgroup $H$ of $G$ containing $Z$ with $H^* = H/Z$; moreover, $H^* \lhd G/Z$ implies $H \lhd G$. But $|H/Z| = p^{k-1}$ implies $|H| = p^k$, as desired. •

Abelian groups (and the quaternions) have the property that every subgroup is normal. At the opposite pole are simple groups $G$ which have no normal subgroups other than the two obvious ones: $\{1\}$ and $G$. We proved, in Part 1, that an abelian group is simple if and only if it is finite and of prime order; Proposition C-1.25 shows that a finite $p$-group of order at least $p^2$ is not simple. We also saw, in Part 1, that the alternating groups $A_n$ are simple for all $n \geq 5$. In fact, $A_5$ is the smallest nonabelian simple group; there are no simple nonabelian groups of order less than 60. We will consider other simple groups later in this chapter.

## Exercises

**C-1.1.** (i) If $V$ is an $n$-dimensional vector space over a field $k$, prove that $\mathrm{GL}(n,k)$ acts on $\mathrm{Mat}_n(k)$ by conjugation: if $P \in \mathrm{GL}(n,k)$ and $A \in \mathrm{Mat}_n(k)$, then the action takes $A$ to $PAP^{-1}$.

(ii) Prove that there is a bijection from the orbit space $\mathrm{Mat}_n(k)/\mathrm{GL}(n,k)$ (the family of all orbits) to $\mathrm{Hom}_k(V,V)$.

**C-1.2.** If $a$ and $b$ are elements in a group $G$, prove that $ab$ and $ba$ have the same order.

**Hint.** Use a conjugation.

**C-1.3.** Prove that if $G$ is a finite group of odd order, then only the identity is conjugate to its inverse.

**Hint.** If $x$ is conjugate to $x^{-1}$, how many elements are in $x^G$?

**C-1.4.** Prove that no two of the following groups of order 8 are isomorphic:

$$\mathbb{Z}_8; \quad \mathbb{Z}_4 \times \mathbb{Z}_2; \quad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2; \quad D_8; \quad \mathbf{Q}.$$

∗ **C-1.5.** Show that $S_4$ has a subgroup isomorphic to $D_8$.

∗ **C-1.6.** Prove that $S_4/\mathbf{V} \cong S_3$, where $\mathbf{V}$ is the four-group.

**Hint.** Use Proposition C-1.4.

∗ **C-1.7.** (i) Prove that $A_4 \ncong D_{12}$.

**Hint.** Recall that $A_4$ has no element of order 6.

(ii) Prove that $D_{12} \cong S_3 \times \mathbb{Z}_2$.
**Hint.** Each element $x \in D_{12}$ has a unique factorization of the form $x = b^i a^j$, where $b^6 = 1$ and $a^2 = 1$.

∗ **C-1.8.** (i) If $G$ is a group, then a normal subgroup $H \lhd G$ is called a ***maximal normal subgroup*** if there is no normal subgroup $K$ of $G$ with $H \subsetneq K \subsetneq G$. Prove that a normal subgroup $H$ is a maximal normal subgroup of $G$ if and only if $G/H$ is a simple group.

(ii) Prove that every finite abelian group $G$ has a subgroup of prime index.
**Hint.** Use Proposition A-4.92 in Part 1.

(iii) Prove that $A_6$ has no subgroup of prime index.

**C-1.9.** (i) (**Landau**) Given a positive integer $n$ and a positive rational $q$, prove that there are only finitely many $n$-tuples $(i_1, \ldots, i_n)$ of positive integers with $q = \sum_{j=1}^{n} 1/i_j$.

(ii) Prove, for every positive integer $n$, that there are only finitely many finite groups having exactly $n$ conjugacy classes.
**Hint.** Use part (i) and the class equation.

**C-1.10.** Find $N_G(H)$ if $G = S_4$ and $H = \langle (1\ 2\ 3) \rangle$.

∗ **C-1.11.** If $H$ is a proper subgroup of a finite group $G$, prove that $G$ is not the union of all the conjugates of $H$: that is, $G \neq \bigcup_{x \in G} xHx^{-1}$.

∗ **C-1.12.** (i) If $H$ is a subgroup of $G$ and $x \in H$, prove that

$$C_H(x) = H \cap C_G(x).$$

(ii) If $H$ is a subgroup of index 2 in a finite group $G$ and $x \in H$, prove that either $|x^H| = |x^G|$ or $|x^H| = \frac{1}{2}|x^G|$, where $x^H$ is the conjugacy class of $x$ in $H$.
**Hint.** Use the Second Isomorphism Theorem.

(iii) Prove that there are two conjugacy classes of 5-cycles in $A_5$, each of which has 12 elements.
**Hint.** If $\alpha = (1\ 2\ 3\ 4\ 5)$, then $|C_{S_5}(\alpha)| = 5$ because $24 = 120/|C_{S_5}(\alpha)|$; hence, $C_{S_5}(\alpha) = \langle \alpha \rangle$. What is $C_{A_5}(\alpha)$?

(iv) Prove that the conjugacy classes in $A_5$ have sizes 1, 12, 12, 15, and 20.

**C-1.13.** (i) Prove that every normal subgroup $H$ of a group $G$ is a union of conjugacy classes of $G$, one of which is $\{1\}$.

(ii) Use part (i) and Exercise C-1.12 to give a second proof of the simplicity of $A_5$.

∗ **C-1.14.** (i) For all $n \geq 5$, prove that all 3-cycles are conjugate in $A_n$.

**Hint.** Show that $(1\ 2\ 3)$ and $(i\ j\ k)$ are conjugate by considering two cases: they are not disjoint (so they move at most 5 letters); they are disjoint.

(ii) Prove that if a normal subgroup $H$ of $A_n$ contains a 3-cycle, where $n \geq 5$, then $H = A_n$. (*Remark.* We have proved this in Lemma A-4.93 in Part 1 when $n = 5$.)

**C-1.15.** Prove that the only normal subgroups of $S_4$ are $\{(1)\}$, $\mathbf{V}$, $A_4$, and $S_4$.

**Hint.** Use Theorem A-4.7 in Part 1, checking the various cycle structures one at a time.

**C-1.16.** Prove that $A_5$ is a group of order 60 that has no subgroup of order 30.

∗ **C-1.17.** (i) Prove, for all $n \geq 5$, that the only normal subgroups of $S_n$ are $\{(1)\}$, $A_n$, and $S_n$.

**Hint.** If $H \lhd S_n$ is a proper subgroup and $H \neq A_n$, then $H \cap A_n = \{(1)\}$.

(ii) Prove that if $n \geq 3$, then $A_n$ is the only subgroup of $S_n$ of order $\frac{1}{2}n!$.
**Hint.** If $H$ is a second such subgroup, then $H \lhd S_n$ and $(H \cap A_n) \lhd A_n$.

(iii) Prove that $S_5$ is a group of order 120 having no subgroup of order 30.
**Hint.** Use the representation on the cosets of a supposed subgroup of order 30, as well as the simplicity of $A_5$.

(iv) Prove that $S_5$ contains no subgroup of order 40.

∗ **C-1.18.** (i) Let $\sigma, \tau \in S_5$, where $\sigma$ is a 5-cycle and $\tau$ is a transposition. Prove that $S_5 = \langle \sigma, \tau \rangle$.

(ii) Give an example showing that $S_n$, for some $n$, contains an $n$-cycle $\sigma$ and a transposition $\tau$ such that $\langle \sigma, \tau \rangle \neq S_n$.

∗ **C-1.19.** Let $G$ be a subgroup of $S_n$.

(i) If $G \cap A_n = \{1\}$, prove that $|G| \leq 2$.

(ii) If $G$ is a simple group with more than two elements, prove that $G \subseteq A_n$.

∗ **C-1.20.** (i) If $n \geq 5$, prove that $S_n$ has no subgroup of index $r$, where $2 < r < n$.

(ii) Prove that if $n \geq 5$, then $A_n$ has no subgroup of index $r$, where $2 \leq r < n$.

**C-1.21.** (i) Prove that if a simple group $G$ has a subgroup of index $n > 1$, then $G$ is isomorphic to a subgroup of $S_n$.

**Hint.** Kernels are normal subgroups.

(ii) Prove that an infinite simple group (such do exist) has no subgroups of finite index $n > 1$.
**Hint.** Use part (i).

∗ **C-1.22.** If $G$ is a group of order $n$, prove that $G$ is isomorphic to a subgroup of $\mathrm{GL}(n, k)$, where $k$ is a field.

**Hint.** If $\sigma \in S_n$, then the $n \times n$ ***permutation matrix*** $P_\sigma$ is the matrix obtained from the $n \times n$ identity matrix by permuting its columns via $\sigma$. Show that $\sigma \mapsto P_\sigma$ is an injective homomorphism $S_n \to \mathrm{GL}(n, k)$.

∗ **C-1.23.** Let $G$ be a group with $|G| = mp$, where $p$ is prime and $1 < m < p$. Prove that $G$ is not simple.

**Hint.** Show that $G$ has a subgroup $H$ of order $p$, and use the representation of $G$ on the cosets of $H$.

**Remark.** We can now show that all but 11 of the numbers smaller than 60 are not orders of nonabelian simple groups (namely, 12, 18, 24, 30, 36, 40, 45, 48, 50, 54, 56). Theorem C-1.22 eliminates all prime powers (for the center is always a normal subgroup), and this exercise eliminates all numbers of the form $mp$, where $p$ is prime and $m < p$. ◀

∗ **C-1.24.** (i) Let a group $G$ act on a set $X$, and suppose that $x, y \in X$ lie in the same orbit: $y = gx$ for some $g \in G$. Prove that $G_y = gG_xg^{-1}$.

(ii) Let $G$ be a finite group acting on a set $X$; prove that if $x, y \in X$ lie in the same orbit, then $|G_x| = |G_y|$.

## ▪ Graphs

The *Cayley graph* of a group $G$ is a nice set on which $G$ acts. Here is a short account of graphs and directed graphs.

**Definition.** A ***graph*** $\Gamma$ is an ordered pair $\Gamma = (V, E)$, where $V$ is a nonempty set, called ***vertices***, and $E$ is a symmetric relation on $V$, called ***adjacency***. If $(u, v) \in E$, we write $u \sim v$, and we call $\{u, v\}$ the ***edge*** connecting $u$ and $v$ (since adjacency is a symmetric relation, $\{v, u\} = \{u, v\}$ is another name of the edge).[4] Given a graph $\Gamma = (V, E)$, we may write

$$\mathrm{Vert}(\Gamma) = V \qquad \text{and} \qquad \mathrm{Edge}(\Gamma) = E.$$

If $\Gamma = (V, E)$ and $\Gamma' = (V', E')$ are graphs, an ***isomorphism*** $\varphi\colon \Gamma \to \Gamma'$ is a bijection $\varphi\colon V \to V'$ that preserves adjacency; that is, if $u \sim v$ in $\Gamma$, then $\varphi(u) \sim \varphi(v)$ in $\Gamma'$. When $\Gamma = \Gamma'$, an isomorphism is called an ***automorphism***, and all automorphisms of $\Gamma$ form a group under composition, denoted by

$$\mathrm{Aut}(\Gamma).$$

[4] A related notion is that of ***multigraph***, which may have more than one edge between a pair of vertices.

An edge of the form $\{v, v\}$ (that is, $v \sim v$) is called ***trivial***. A graph is ***discrete*** if it has no edges; it is ***complete*** if it has no trivial edges and every distinct pair of vertices are adjacent.

Given a graph $\Gamma = (V, E)$ and vertices $u, v \in V$, a ***path*** connecting $u$ and $v$ is a sequence of edges $\{u_0, u_1\}, \{u_1, u_2\}, \ldots, \{u_{n-1}, u_n\}$ with $u_1 = u$ and $u_n = v$. A graph is ***connected*** if, for each pair of distinct vertices, there is a path connecting them. A path $\alpha = e_1 \cdots e_n$ is ***reduced*** if either $\alpha$ is trivial, i.e., $\alpha = (v, v)$, or if no $e_i$ is trivial and no edge $e_j = (u, v)$ is adjacent to its inverse $(v, u)$. A ***circuit*** is a reduced closed path. A ***tree*** is a connected graph having no circuits.

If $\Gamma$ is a finite graph, that is, if $\Gamma$ has only finitely many vertices, say $\mathrm{Vert}(\Gamma) = \{v_1, \ldots, v_n\}$, then its ***adjacency matrix*** $[a_{ij}]$ is the $n \times n$ symmetric matrix, where

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \sim v_j, \\ 0 & \text{if } v_i \not\sim v_j. \end{cases}$$

If $\Gamma$ has no trivial edges, then its adjacency matrix has only 0's on its diagonal. The adjacency matrix of a complete graph with $n$ vertices has 0's on the diagonal and 1's everywhere else.

Graphs can be pictured. For example, if $\Gamma$ is finite with $\mathrm{Vert}(\Gamma) = \{v_1, \ldots, v_n\}$, draw $n$ points in the plane, label each with a vertex, and draw a line segment (an edge) connecting each pair of adjacent vertices. A trivial edge is a point $v$ having a circular edge from $v$ to $v$.

Here are some pictures; the first graph is not connected; the others are connected. Graphs (ii) and (v) are trees.

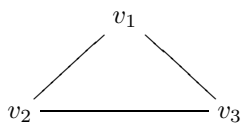
(i) A discrete graph with 2 vertices.

$v_1 \qquad v_2$

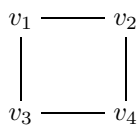
(ii) An edge.

$v_1 \text{ —— } v_2$

(iii) A triangle (which is a complete graph on 3 vertices)




(iv) A square




(v) The natural numbers $\mathbb{N}$, where $n$ and $n+1$ are adjacent for all $n \geq 0$

$0 \text{ —— } 1 \text{ —— } 2 \text{ —— } \cdots \text{ —— } v_n \text{ —— } v_{n+1} \text{ —— } \cdots$

Here are the adjacency matrices of graphs (i) through (iv) (the adjacency matrix of (v) is infinite).

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

**Remark.** If $\Gamma$ is a finite graph with $n$ vertices and no trivial edges, let $V(\Gamma)$ be the vector space over $\mathbb{F}_2$ having basis $\Gamma$. Now the adjacency matrix $A = [a_{ij}]$ of $\Gamma$ is a symmetric matrix of 0's and 1's having all diagonal elements zero. If we regard $A$ as a matrix with entries in $\mathbb{F}_2$, then we may view $V(\Gamma)$ as an inner product space, where $(v_i, v_j) = a_{ij}$. Thus, $(v_i, v_j) = 1$ if and only if $v_i$ and $v_j$ are adjacent in $\Gamma$. If we assume that $A$ is nonsingular, that is, $\det(A)$ is odd, then $V(\Gamma)$ is an alternating space. If $T$ is an $n \times n$ symplectic matrix, then $(Tv_i, Tv_j) = (v_i, v_j)$; that is, $T$ preserves adjacency. In other words, $\mathrm{Aut}(\Gamma)$ is a subgroup of $\mathrm{Sp}(n, \mathbb{F}_2)$.

Note that if its adjacency matrix $A$ is nonsingular, then $\Gamma$ must have an even number of vertices (for nondegenerate alternating spaces are even-dimensional (Corollary B-3.99 in Part 1)). It can be shown that the complete graph $K_{2n}$ is nonsingular (see Theorem 3.1 and §5 in Hughes–Singhi [**100**]). ◀

**Definition.** A ***directed graph*** $\Gamma$ (or ***digraph***) is an ordered pair $\Gamma = (V, E)$, where $V$ is a nonempty set called ***vertices*** and $E \subseteq V \times V$ is a (not necessarily symmetric) relation on $V$. If $(u, v) \in E$, then we call $e = (u, v)$ a (directed) ***edge from*** $u$ ***to*** $v$. If $e = (u, v)$ is an edge, we write $o(e) = u$ (the ***origin*** of $e$) and $t(e) = v$ (the ***terminus*** of $e$). If $\Gamma = (V, E)$ is a directed graph, we may write

$$\mathrm{Vert}(\Gamma) = V \qquad \text{and} \qquad \mathrm{Edge}(\Gamma) = E.$$

Given a directed graph $\Gamma = (V, E)$ and vertices $u, v \in V$, a ***path*** $p$ ***from*** $u$ ***to*** $v$ is a sequence of edges $e_1 = (u_0, u_1), e_2 = (u_1, u_2), \ldots, e_n = (u_{n-1}, u_n)$ with $o(e_1) = u_1 = u$, $t(e_i) = o(e_{i+1})$ for all $1 \leq i < n$, and $t(e_n) = u_n = v$. A directed graph is ***connected*** if, for each pair of distinct vertices, there is a path $p$ connecting them. Define the ***length*** $\ell$ of a path $p = e_1, e_2, \ldots, e_n$ to be $n$; that is, $\ell(p) = n$.

Directed graphs are usually ***labeled***: both vertices and edges have names.

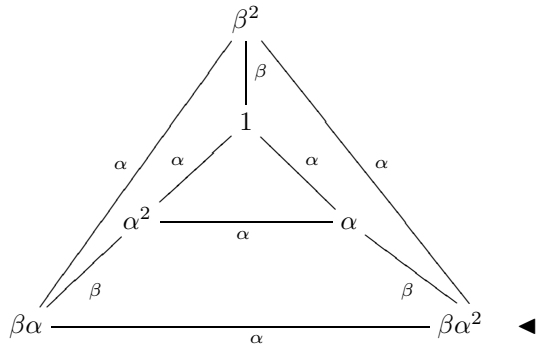
The most important directed graph for us is the *Cayley graph*.

**Definition.** Let $G$ be a finitely generated group and let $S$ be a finite generating set: $G = \langle S \rangle$. The ***Cayley graph*** $\Gamma = \Gamma(G, S)$ is the directed graph with $\mathrm{Vert}(\Gamma) = G$ and edges $e = (g, gs)$, where $g \in G$ and $s \in S$.

There are two variations. Some (e.g., Serre [**202**]) assume, for every edge $e = (u, v)$, there is an ***inverse edge*** $e^{-1} = (v, u)$; moreover, some assume the generating set $S$ is ***symmetric***; that is, if $s \in S$, then $s^{-1} \in S$ (remember, $S$ is

a subset of a group). Note that if $e = (g, gs)$, then $e^{-1} = (gs, g)$. (We note that the Cayley graph of a group $G$ does depend on the choice of generating set $S$.) Exercise C-1.26 on page 19 says that Cayley graphs are always connected.

**Example C-1.26.** Here is the Cayley graph of the symmetric group $S_3$ with generating set $S = \{\alpha, \beta\}$, where $\alpha = (1\ 2\ 3)$ and $\beta = (1\ 2)$.



◀

**Proposition C-1.27.** *If $G = \langle S \rangle$ is a group generated by a subset $S$, then $G$ acts on the Cayley graph $\Gamma = \Gamma(G, S)$.*

**Proof.** If $a \in G$, define $ag$ (where $g \in G = \mathrm{Vert}(\Gamma)$) and $a(g, gs) = (ag, ags)$, where $(g, gs) \in \mathrm{Edge}(\Gamma)$. •

## Exercises

**C-1.25.** If $\varphi\colon \Gamma \to \Gamma'$ is an isomorphism of graphs, prove that $\Gamma$ is connected if and only if $\Gamma'$ is connected, and that $\Gamma$ is a tree if and only if $\Gamma'$ is a tree.

∗ **C-1.26.** Let $G$ be a group generated by a subset $S$, and let $\Gamma(G, S)$ be the corresponding Cayley graph. Use the fact that $S$ generates $G$ to prove that the Cayley graph $\Gamma(G, S)$ is connected.

∗ **C-1.27.** Prove that every Cayley graph $\Gamma(G, S)$ is a metric space if we define

$$d(u, v) = \ell(p),$$

where $p$ is the shortest path from $u$ to $v$ and $\ell(p)$ is its length.

**C-1.28.** Let $G$ be a free abelian group of rank 2 with basis $S = x, y$. Prove that the vertices of the Cayley graph $\Gamma(G, S)$ are all the lattice points in the plane.

## ▪ Counting

We now use groups to solve some difficult counting problems.

**Theorem C-1.28 (Burnside's Lemma).**[5] *Let $G$ be a finite group acting on a finite set $X$. If $N$ is the number of orbits, then*

$$N = \frac{1}{|G|} \sum_{\tau \in G} \mathrm{Fix}(\tau),$$

*where* $\mathrm{Fix}(\tau)$ *is the number of $x \in X$ fixed by $\tau$.*

**Proof.** List the elements of $X$ as follows: choose $x_1 \in X$, and then list all the elements $x_1, x_2, \ldots, x_r$ in the orbit $\mathcal{O}(x_1)$; then choose $x_{r+1} \notin \mathcal{O}(x_1)$, and list the elements $x_{r+1}, x_{r+2}, \ldots$ in $\mathcal{O}(x_{r+1})$; continue this procedure until all the elements of $X$ are listed. Now list the elements $\tau_1, \tau_2, \ldots, \tau_n$ of $G$, and form Figure C-1.2, where

$$f_{i,j} = \begin{cases} 1 & \text{if } \tau_i \text{ fixes } x_j, \\ 0 & \text{if } \tau_i \text{ moves } x_j. \end{cases}$$

| | $x_1$ | $x_2$ | $\cdots$ | $x_{r+1}$ | $x_{r+2}$ | $\cdots$ |
|---|---|---|---|---|---|---|
| $\tau_1$ | $f_{1,1}$ | $f_{1,2}$ | $\cdots$ | $f_{1,r+1}$ | $f_{1,r+2}$ | $\cdots$ |
| $\tau_2$ | $f_{2,1}$ | $f_{2,2}$ | $\cdots$ | $f_{2,r+1}$ | $f_{2,r+2}$ | $\cdots$ |
| $\tau_i$ | $f_{i,1}$ | $f_{i,2}$ | $\cdots$ | $f_{i,r+1}$ | $f_{i,r+2}$ | $\cdots$ |
| $\tau_n$ | $f_{n,1}$ | $f_{n,2}$ | $\cdots$ | $f_{n,r+1}$ | $f_{n,r+2}$ | $\cdots$ |

**Figure C-1.2.** Burnside's Lemma.

Now $\mathrm{Fix}(\tau_i)$, the number of $x$ fixed by $\tau_i$, is the number of 1's in the $i$th row of the array; therefore, $\sum_{\tau \in G} \mathrm{Fix}(\tau)$ is the total number of 1's in the array. Let us now look at the columns. The number of 1's in the first column is the number of $\tau_i$ that fix $x_1$; by definition, these $\tau_i$ comprise $G_{x_1}$. Thus, the number of 1's in column 1 is $|G_{x_1}|$. Similarly, the number of 1's in column 2 is $|G_{x_2}|$. By Exercise C-1.24 on page 16, $|G_{x_1}| = |G_{x_2}|$. By Theorem C-1.16, the number of 1's in the $r$ columns labeled by the $x_i \in \mathcal{O}(x_1)$ is thus

$$r|G_{x_1}| = |\mathcal{O}(x_1)| \cdot |G_{x_1}| = (|G|/|G_{x_1}|)\,|G_{x_1}| = |G|.$$

The same is true for any other orbit: its columns collectively contain exactly $|G|$ 1's. Therefore, if there are $N$ orbits, there are $N|G|$ 1's in the array. We conclude that

$$\sum_{\tau \in G} \mathrm{Fix}(\tau) = N|G|. \quad \bullet$$

[5] Burnside himself attributed this lemma to Frobenius. To avoid the confusion that would be caused by changing a popular name, P. M. Neumann has suggested that it be called "not-Burnside's Lemma". Burnside was a fine mathematician, and there do exist theorems properly attributed to him. For example, Burnside proved that there are no simple groups of order $p^m q^n$, where $p$ and $q$ are primes.

We are going to use Burnside's Lemma to solve problems of the following sort. How many striped flags are there having six stripes (of equal width) each of which can be colored red, white, or blue? Clearly, the two flags in Figure C-1.3 are the same: the bottom flag is just the top one rotated about its center.

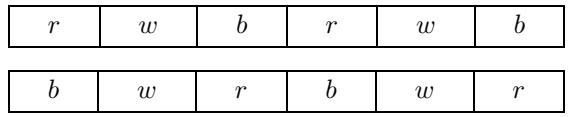



**Figure C-1.3.** Striped flags.

Let $X$ be the set of all 6-tuples of colors; if $x \in X$, then

$$x = (c_1, c_2, c_3, c_4, c_5, c_6),$$

where each $c_i$ denotes either red, white, or blue. Let $\tau$ be the permutation that reverses all the indices:

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 4 & 3 & 2 & 1 \end{pmatrix} = (1\ 6)(2\ 5)(3\ 4)$$

(thus, $\tau$ "rotates" each 6-tuple $x$ of colored stripes). The cyclic group $G = \langle \tau \rangle$ acts on $X$; since $|G| = 2$, the orbit of any 6-tuple $x$ consists of either one or two elements: either $\tau$ fixes $x$ or it does not. Since a flag is unchanged by rotation, it is reasonable to identify a flag with an orbit of a 6-tuple. For example, the orbit consisting of the 6-tuples

$$(r, w, b, r, w, b) \quad \text{and} \quad (b, w, r, b, w, r)$$

describes the flag in Figure C-1.3. The number of flags is thus the number $N$ of orbits; by Burnside's Lemma, $N = \frac{1}{2}[\mathrm{Fix}((1)) + \mathrm{Fix}(\tau)]$. The identity permutation (1) fixes every $x \in X$, and so $\mathrm{Fix}((1)) = 3^6$ (there are three colors). Now $\tau$ fixes a 6-tuple $x$ if and only if it is a "palindrome", that is, if the colors in $x$ read the same forward as backward. For example,

$$x = (r, r, w, w, r, r)$$

is fixed by $\tau$. Conversely, if

$$x = (c_1, c_2, c_3, c_4, c_5, c_6)$$

is fixed by $\tau = (1\ 6)(2\ 5)(3\ 4)$, then $c_1 = c_6$, $c_2 = c_5$, and $c_3 = c_4$; that is, $x$ is a palindrome. It follows that $\mathrm{Fix}(\tau) = 3^3$, for there are three choices for each of $c_1$, $c_2$, and $c_3$. The number of flags is thus

$$N = \tfrac{1}{2}(3^6 + 3^3) = 378.$$

Let us make the notion of coloring more precise.

**Definition.** If a group $G$ acts (on the right) on $X = \{1, \ldots, n\}$ and $\mathcal{C}$ is a set of $q$ *colors*, then $G$ acts on the set $\mathcal{C}^n$ of all $n$-tuples of colors by

$$\sigma(c_1, \ldots, c_n) = (c_{1\sigma}, \ldots, c_{n\sigma}) \quad \text{for all } \sigma \in G.$$

An orbit of $(c_1, \ldots, c_n) \in \mathcal{C}^n$ is called a ***(q, G)-coloring*** of $X$.

Color each square in a $4 \times 4$ grid red or black (adjacent squares may have the same color; indeed, one possibility is that all the squares have the same color).

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 |

| | | | |
|---|---|---|---|
| 13 | 9 | 5 | 1 |
| 14 | 10 | 6 | 2 |
| 15 | 11 | 7 | 3 |
| 16 | 12 | 8 | 4 |

**Figure C-1.4.** Chessboard and a rotation.

If $X$ consists of the 16 squares in the grid and $\mathcal{C}$ consists of the two colors red and black, then the cyclic group $G = \langle R \rangle$ of order 4 acts on $X$, where $R$ is clockwise rotation by 90°; Figure C-1.4 shows how $R$ acts: the right square is $R$'s action on the left square. In cycle notation,

$$\begin{aligned} R &= (1,\ 4,\ 16,\ 13)(2,\ 8,\ 15,\ 9)(3,\ 12,\ 14,\ 5)(6,\ 7,\ 11,\ 10), \\ R^2 &= (1,\ 16)(4,\ 13)(2,\ 15)(8,\ 9)(3,\ 14)(12,\ 5)(6,\ 11)(7,\ 10), \\ R^3 &= (1,\ 13,\ 16,\ 4)(2,\ 9,\ 15,\ 8)(3,\ 5,\ 14,\ 12)(6,\ 10,\ 11,\ 7). \end{aligned}$$

A red-and-black chessboard does not change when it is rotated; it is merely viewed from a different position. Thus, we may regard a chessboard as a 2-coloring of $X$; the orbit of a 16-tuple corresponds to the four ways of viewing the board.

By Burnside's Lemma, the number of chessboards is

$$\tfrac{1}{4}\Big[\mathrm{Fix}((1)) + \mathrm{Fix}(R) + \mathrm{Fix}(R^2) + \mathrm{Fix}(R^3)\Big].$$

Now $\mathrm{Fix}((1)) = 2^{16}$, for every 16-tuple is fixed by the identity. To compute $\mathrm{Fix}(R)$, note that squares 1, 4, 16, 13 must all have the same color in a 16-tuple fixed by $R$. Similarly, squares 2, 8, 15, 9 must have the same color, squares 3, 12, 14, 5 must have the same color, and squares 6, 7, 11, 10 must have the same color. We conclude that $\mathrm{Fix}(R) = 2^4$; note that the exponent 4 is the number of cycles in the complete factorization of $R$. A similar analysis shows that $\mathrm{Fix}(R^2) = 2^8$, for the complete factorization of $R^2$ has 8 cycles, and $\mathrm{Fix}(R^3) = 2^4$, because the cycle structure of $R^3$ is the same as that of $R$. Therefore, the number $N$ of chessboards is

$$N = \tfrac{1}{4}\Big[2^{16} + 2^4 + 2^8 + 2^4\Big] = 16{,}456.$$

We now show, as in the discussion of the $4 \times 4$ chessboard, that the cycle structure of a permutation $\tau$ allows one to calculate $\mathrm{Fix}(\tau)$.

**Lemma C-1.29.** *Let $\mathcal{C}$ be a set of $q$ colors, and let $G$ be a subgroup of $S_n$. If $\tau \in G$, then*

$$\mathrm{Fix}(\tau) = q^{t(\tau)},$$

*where $t(\tau)$ is the number of cycles in the complete factorization of $\tau$.*

**Proof.** Since $\tau(c_1, \ldots, c_n) = (c_{\tau 1}, \ldots, c_{\tau n}) = (c_1, \ldots, c_n)$, we see that $c_{\tau i} = c_i$ for all $i$, and so $\tau i$ has the same color as $i$; that is, under the action of $\langle \tau \rangle$, all the points in the orbit of $i$ have the same color. But if the complete factorization of $\tau$ is $\tau = \beta_1 \cdots \beta_{t(\tau)}$ and $i$ occurs in $\beta_j$, then Example C-1.8 shows that the orbit containing $i$ is the disjoint cycle $\beta_j$. Thus, for an $n$-tuple to be fixed by $\tau$, all the symbols involved in each of the $t(\tau)$ cycles must have the same color; as there are $q$ colors, there are thus $q^{t(\tau)}$ $n$-tuples fixed by $\tau$. •

**Theorem C-1.30.** *Let $G$ act on a finite set $X$. If $N$ is the number of $(q, G)$-colorings of $X$, then*

$$N = \frac{1}{|G|} \sum_{\tau \in G} q^{t(\tau)},$$

*where $t(\tau)$ is the number of cycles in the complete factorization of $\tau$.*

**Proof.** Rewrite Burnside's Lemma using Lemma C-1.29. •

There is a generalization of this technique, due to Pólya (Biggs [**20**], p. 403), giving formulas of the sort that count the number of red, white, blue, and green flags having 20 stripes exactly 7 of which are red and 5 of which are blue.

## Exercises

**C-1.29.** How many flags are there with $n$ stripes of equal width, each of which can be colored any one of $q$ given colors?

**Hint.** The parity of $n$ is relevant.

**C-1.30.** Let $X$ be the squares in an $n \times n$ grid, and let $\rho$ be a rotation by $90^\circ$. Define a ***chessboard*** to be a $(q, G)$-coloring, where the cyclic group $G = \langle \rho \rangle$ of order 4 is acting on $X$. Show that the number of chessboards is

$$\frac{1}{4}\left(q^{n^2} + q^{\lfloor (n^2+1)/2 \rfloor} + 2q^{\lfloor (n^2+3)/4 \rfloor}\right),$$

where $\lfloor x \rfloor$ is the greatest integer in the number $x$.

**C-1.31.** Let $X$ be a disk divided into $n$ congruent circular sectors, and let $\rho$ be a rotation by $(360/n)^\circ$. Define a ***roulette wheel*** to be a $(q, G)$-coloring, where the cyclic group $G = \langle \rho \rangle$ of order $n$ is acting on $X$. Prove that if $n = 6$, then there are $\frac{1}{6}(2q + 2q^2 + q^3 + q^6)$ roulette wheels having 6 sectors. (The formula for the number of roulette wheels with $n$ sectors is

$$\frac{1}{n} \sum_{d|n} \phi(n/d) q^d,$$

where $\phi$ is the Euler $\phi$-function.)

**C-1.32.** Let $X$ be the vertices of a regular $n$-gon, and let the dihedral group $G = D_{2n}$ act on $X$ [as the usual group of symmetries (Examples A-4.27 and A-4.28 in Part 1)]. Define a ***bracelet*** to be a $(q, G)$-coloring of a regular $n$-gon, and call each of its vertices a ***bead***. (Not only can we rotate a bracelet, we can also *flip* it: that is, turn it upside down by rotating it in space about a line joining two beads.)

(i) How many bracelets are there having 5 beads, each of which can be colored any one of $q$ available colors?
**Hint.** The group $G = D_{10}$ is acting. Use Example A-4.28 in Part 1 to assign to each symmetry a permutation of the vertices, and then show that the number of bracelets is
$$\frac{1}{10}\big(q^5 + 4q + 5q^3\big).$$

(ii) How many bracelets are there having 6 beads, each of which can be colored any one of $q$ available colors?
**Hint.** The group $G = D_{12}$ is acting. Assign a permutation of the vertices to each symmetry, and then show that the number of bracelets is
$$\frac{1}{12}\big(q^6 + 2q^4 + 4q^3 + 3q^2 + 2q\big).$$

## C-1.2. Sylow Theorems

Recall that a group $G$ is called *simple* if $G \neq \{1\}$ and it has no normal subgroups other than $\{1\}$ and $G$ itself. We saw, in Proposition A-4.92 in Part 1, that the abelian simple groups are precisely the (finite) cyclic groups $\mathbb{Z}_p$ of prime order $p$, and we saw, in Theorem A-4.97 in Part 1, that $A_n$ is a nonabelian simple group for all $n \geq 5$. In fact, $A_5$ is the nonabelian simple group of smallest order. How can we prove that a nonabelian group $G$ of order less than $60 = |A_5|$ is not simple? Exercise C-1.23 on page 16 states that if $G$ is a group of order $mp$, where $p$ is prime and $1 < m < p$, then $G$ is not simple. This exercise shows that many of the numbers less than 60 are not orders of simple groups. After throwing out all prime powers (finite $p$-groups of order at least $p^2$ are never simple, by Proposition C-1.25), the only remaining possibilities are

$$12,\ 18,\ 24,\ 30,\ 36,\ 40,\ 45,\ 48,\ 50,\ 54,\ 56.$$

The solution to Exercise C-1.23 uses Cauchy's Theorem, which says that $G$ has a subgroup of order $p$. We shall see that if $G$ has a subgroup of prime power order $p^e$ instead of $p$, then the exercise can be generalized and the list of candidates can be shortened. Proposition C-1.40 below uses this result to show that $A_5$ is, indeed, the smallest nonabelian simple group.

The first book on group theory, Jordan [**117**], was published in 1870; more than half of it is devoted to Galois theory, then called the theory of equations. At about the same time, but too late for publication in Jordan's book, three fundamental theorems were discovered. In 1868, Schering proved the Basis Theorem: every finite abelian group is a direct product of primary cyclic groups. In 1870, Kronecker, unaware of Schering's proof, also proved this result. In 1878, Frobenius and Stickelberger proved the Fundamental Theorem of Finite Abelian Groups. In 1872, Sylow showed, for every finite group $G$ and every prime $p$, that if $p^e$ is the largest power

of $p$ dividing $|G|$, then $G$ has a subgroup of order $p^e$ (nowadays called a *Sylow p-subgroup*). Indeed, Sylow subgroups are analogs for finite nonabelian groups of the primary components of finite abelian groups. We will use Sylow subgroups to generalize Exercise C-1.23.

Our strategy for proving the Sylow Theorems works best if we adopt the following definition.

**Definition.** Let $p$ be prime. A ***Sylow p-subgroup*** of a group $G$ is a maximal $p$-subgroup $P$.

Maximality means that if $Q$ is a $p$-subgroup of $G$ and $P \subseteq Q$, then $P = Q$.

It follows from Lagrange's Theorem that if $p^e$ is the largest power of $p$ dividing the order of a finite group $G$, then a subgroup of $G$ of order $p^e$, should it exist, is a maximal $p$-subgroup. While it is not clear at the outset that $G$ has any subgroups of order $p^e$, it is clear, using Zorn's Lemma, that maximal $p$-subgroups always exist. We shall prove, in Theorem C-1.35, that Sylow $p$-subgroups do have order $p^e$.

Let us show that if $S$ is any $p$-subgroup of a finite group $G$ (perhaps $S = \{1\}$), then there exists a Sylow $p$-subgroup $P$ containing $S$. If there is no $p$-subgroup strictly containing $S$, then $S$ itself is a Sylow $p$-subgroup. Otherwise, there is a $p$-subgroup $P_1$ with $S \subsetneq P_1$. If $P_1$ is maximal, it is Sylow, and we are done. Otherwise, there is some $p$-subgroup $P_2$ with $P_1 \subsetneq P_2$. This procedure of producing larger and larger $p$-subgroups $P_i$ must end after a finite number of steps because $|P_i| \leq |G|$ for all $i$; the largest $P_i$ must, therefore, be a Sylow $p$-subgroup.

Recall that a *conjugate* of a subgroup $H \subseteq G$ is a subgroup of $G$ of the form

$$aHa^{-1} = \{aha^{-1} : h \in H\},$$

where $a \in G$. The *normalizer* of $H$ in $G$ is the subgroup

$$N_G(H) = \{a \in G : aHa^{-1} = H\},$$

and Corollary C-1.19 states that if $H$ is a subgroup of a finite group $G$, then the number of conjugates of $H$ in $G$ is $[G : N_G(H)]$.

It is obvious that $H \lhd N_G(H)$, and so the quotient group $N_G(H)/H$ is defined.

**Lemma C-1.31.** *Let $P$ be a Sylow $p$-subgroup of a finite group $G$.*

(i) *Every conjugate of $P$ is also a Sylow $p$-subgroup of $G$.*

(ii) *$|N_G(P)/P|$ is prime to $p$.*

(iii) *If $a \in G$ has order some power of $p$ and $aPa^{-1} = P$, then $a \in P$.*

**Proof.**

(i) If $a \in G$ and $aPa^{-1}$ is not a Sylow $p$-subgroup of $G$, then there is a $p$-subgroup $Q$ with $aPa^{-1} \subsetneq Q$. Now $a^{-1}Qa$ is a $p$-subgroup strictly containing $P$: $P \subsetneq a^{-1}Qa$, which contradicts the maximality of $P$.

(ii) If $p$ divides $|N_G(P)/P|$, then Cauchy's Theorem shows that $N_G(P)/P$ contains an element $aP$ of order $p$, and hence $N_G(P)/P$ contains a subgroup $S^* = \langle aP \rangle$ of order $p$. By the Correspondence Theorem, there is a subgroup $S$ with $P \subseteq S \subseteq N_G(P)$ such that $S/P \cong S^*$. But $S$ is a $p$-subgroup of

$N_G(P) \subseteq G$ (by Exercise A-4.85 on page 172 in Part 1) strictly larger than $P$, contradicting the maximality of $P$. We conclude that $p$ does not divide $|N_G(P)/P|$.

(iii) By the definition of normalizer, the element $a$ lies in $N_G(P)$. If $a \notin P$, then the coset $aP$ is a nontrivial element of $N_G(P)/P$ having order some power of $p$; in light of part (ii), this contradicts Lagrange's Theorem. •

Since every conjugate of a Sylow $p$-subgroup is a Sylow $p$-subgroup, it is natural to let $G$ act on the Sylow $p$-subgroups by conjugation.

**Theorem C-1.32 (Sylow).** *Let $G$ be a finite group of order $p_1^{e_1} \cdots p_t^{e_t}$, where the $p_i$ are distinct primes, and let $P$ be a Sylow $p$-subgroup of $G$ for some prime $p = p_j$.*

(i) *Every Sylow $p$-subgroup is conjugate to $P$.*[6]

(ii) *If there are $r_j$ Sylow $p_j$-subgroups, then $r_j$ is a divisor of $|G|/p_j^{e_j}$ and*

$$r_j \equiv 1 \bmod p_j.$$

**Proof.** Let $X = \{P_1, \ldots, P_{r_j}\}$ be the set of all the conjugates of $P$, where we have denoted $P$ by $P_1$. If $Q$ is any Sylow $p$-subgroup of $G$, then $Q$ acts on $X$ by conjugation: if $a \in Q$, then it sends

$$P_i = g_i P g_i^{-1} \mapsto a\big(g_i P g_i^{-1}\big)a^{-1} = (ag_i)P(ag_i)^{-1} \in X.$$

By Corollary C-1.17, the number of elements in any orbit is a divisor of $|Q|$; that is, every orbit has size some power of $p$ (because $Q$ is a $p$-group). If there is an orbit of size 1, then there is some $P_i$ with $aP_ia^{-1} = P_i$ for all $a \in Q$. By Lemma C-1.31, we have $a \in P_i$ for all $a \in Q$; that is, $Q \subseteq P_i$. But $Q$, being a Sylow $p$-subgroup, is a maximal $p$-subgroup of $G$, and so $Q = P_i$. In particular, if $Q = P_1$, then there is only one orbit of size 1, namely, $\{P_1\}$, and all the other orbits have sizes that are honest powers of $p$. We conclude that $|X| = r_j \equiv 1 \bmod p_j$.

Suppose now that there is some Sylow $p$-subgroup $Q$ that is not a conjugate of $P$; thus, $Q \neq P_i$ for any $i$. Again, we let $Q$ act on $X$, and again we ask if there is an orbit of size 1, say, $\{P_k\}$. As in the previous paragraph, this implies $Q = P_k$, contrary to our present assumption that $Q \notin X$. Hence, there are no orbits of size 1, which says that each orbit has size an honest power of $p$. It follows that $|X| = r_j$ is a multiple of $p$; that is, $r_j \equiv 0 \bmod p_j$, which contradicts the congruence $r_j \equiv 1 \bmod p_j$. Therefore, no such $Q$ can exist, and so all Sylow $p$-subgroups are conjugate to $P$.

Finally, since all Sylow $p$-subgroups are conjugate, we have $r_j = [G : N_G(P)]$, and so $r_j$ is a divisor of $|G|$. But $r_j \equiv 1 \bmod p_j$ implies $(r_j, p_j^{e_j}) = 1$, so that Euclid's Lemma gives $r_j$ a divisor of $|G|/p_j^{e_j}$. •

**Corollary C-1.33.** *A finite group $G$ has a unique Sylow $p$-subgroup $P$ for some prime $p$ if and only if $P \lhd G$.*

[6] It follows that all Sylow $p$-subgroups of a finite group $G$ are isomorphic. This may not be true for infinite groups. However, when $G$ is a periodic linear group over a field (a group is *linear* if it is isomorphic to a subgroup of a matrix group; for nonabelian groups, the term *periodic* is used instead of *torsion*), then Wehrfritz [**229**] shows that its Sylow $p$-subgroups are conjugate.

**Proof.** Assume that $P$, a Sylow $p$-subgroup of $G$, is unique. For each $a \in G$, the conjugate $aPa^{-1}$ is also a Sylow $p$-subgroup; by uniqueness, $aPa^{-1} = P$ for all $a \in G$, and so $P \lhd G$.

Conversely, assume that $P \lhd G$. If $Q$ is any Sylow $p$-subgroup, then $Q = aPa^{-1}$ for some $a \in G$; but $aPa^{-1} = P$, by normality, and so $Q = P$. •

**Corollary C-1.34.** *If $P$ is a Sylow $p$-subgroup of a finite group $G$, then its normalizer $N = N_G(P)$ is **self-normalizing**; that is, $N_G(N) = N$.*

**Proof.** Let $x \in N = N_G(P)$. Since $P \lhd N$, we have $xPx^{-1} \subseteq N$, so that both $P$ and $xPx^{-1}$ are Sylow $p$-subgroups of $N$. By Theorem C-1.32(i), there is $n \in N$ with $nPn^{-1} = xPx^{-1}$. Hence, $nx^{-1} \in N = N_G(P)$, and $x \in N$. Therefore, $N_G(N) \subseteq N = N_G(P)$. The reverse inclusion always holds, and so $N_G(N) = N$. •

The next result shows that the order of a Sylow $p$-subgroup of a group $G$ is the largest power of $p$ dividing $|G|$.

**Theorem C-1.35 (Sylow).** *If $G$ is a finite group of order $p^e m$, where $p$ is prime and $p \nmid m$, then every Sylow $p$-subgroup $P$ of $G$ has order $p^e$.*

**Proof.** We first show that $p \nmid [G : P]$. Now $[G : P] = [G : N_G(P)][N_G(P) : P]$. The first factor, $[G : N_G(P)] = r$, is the number of conjugates of $P$ in $G$, and so $p$ does not divide $[G : N_G(P)]$ because $r \equiv 1 \bmod p$. The second factor, $[N_G(P) : P] = |N_G(P)/P|$, is also not divisible by $p$, by Lemma C-1.31. Therefore, $p$ does not divide $[G : P]$, by Euclid's Lemma.

Now $|P| = p^k$ for some $k \leq e$, and so $[G : P] = |G|/|P| = p^e m/p^k = p^{e-k}m$. Since $p$ does not divide $[G : P]$, we must have $k = e$; that is, $|P| = p^e$. •

**Example C-1.36.**

(i) Let $G = S_4$. Since $|S_4| = 24 = 2^3 3$, a Sylow 2-subgroup $P$ of $S_4$ has order 8. We have seen, in Exercise C-1.5, that $S_4$ contains a copy of the dihedral group $D_8$ (the symmetries of a square). The Sylow Theorem says that all subgroups of $S_4$ of order 8 are conjugate, hence isomorphic; thus, $P \cong D_8$. Moreover, the number $r$ of Sylow 2-subgroups is a divisor of 24 congruent to 1 mod 2; that is, $r$ is an odd divisor of 24. Since $r \neq 1$ (Exercise C-1.34), there are exactly three Sylow 2-subgroups.

(ii) If $G$ is a finite abelian group, then a Sylow $p$-subgroup is just its $p$-primary component (since $G$ is abelian, every subgroup is normal, and so there is a unique Sylow $p$-subgroup for every prime $p$). ◀

Here is a second proof of the last Sylow Theorem.

**Theorem C-1.37.** *If $G$ is a finite group of order $p^e m$, where $p$ is a prime and $p \nmid m$, then $G$ has a subgroup of order $p^e$.*

**Proof (Wielandt).** If $X$ is the family of all those *subsets* of $G$ having exactly $p^e$ elements, then $|X| = \binom{p^e m}{p^e}$, and $p \nmid |X|$.[7] Now $G$ acts on $X$: define $gB$, for $g \in G$ and $B \in X$, by

$$gB = \{gb : b \in B\}.$$

If $p$ divides $|\mathcal{O}(B)|$ for every $B \in X$, where $\mathcal{O}(B)$ is the orbit of $B$, then $p$ is a divisor of $|X|$, for $X$ is the disjoint union of orbits (Proposition C-1.15). As $p \nmid |X|$, there exists a subset $B$ with $|B| = p^e$ and with $|\mathcal{O}(B)|$ not divisible by $p$. If $G_B$ is the stabilizer of this subset, Theorem C-1.16 gives $[G : G_B] = |\mathcal{O}(B)|$, and so $|G| = |G_B| \cdot |\mathcal{O}(B)|$. Since $p^e \mid |G|$ and $(p^e, |\mathcal{O}(B)|) = 1$, Euclid's Lemma gives $p^e \mid |G_B|$. Therefore, $p^e \leq |G_B|$.

For the reverse inequality, choose an element $b \in B$. Now $G_B b$ is a right coset of $G_B$, and so $|G_B| = |G_B b|$. But $G_B b \subseteq B$, because $G_B b = \{gb : g \in G_B\}$ and $gb \in gB \subseteq B$ (for $g \in G_B$). Therefore, $|G_B| \leq |B| = p^e$. We conclude that $G_B$ is a subgroup of $G$ of order $p^e$. •

**Proposition C-1.38.** *A finite group $G$ all of whose Sylow subgroups are normal is the direct product of its Sylow subgroups.*[8]

**Proof.** Let $|G| = p_1^{e_1} \cdots p_t^{e_t}$ and let $G_{p_i}$ be the Sylow $p_i$-subgroup of $G$. We use Exercise C-1.33 on page 32. The subgroup $S$ generated by all the Sylow subgroups is $G$, for $p_i^{e_i} \mid |S|$ for all $i$; hence, $|S| = |G|$. Finally, if $x \in G_{p_i} \cap \left\langle \bigcup_{j \neq i} G_{p_j} \right\rangle$, then $x = s_i \in G_{p_i}$ and $x = \prod_{j \neq i} s_j$, where $s_j \in G_{p_j}$. Now $x^{p_i^n} = 1$ for some $n \geq 1$. On the other hand, there is some power of $p_j$, say, $q_j$, with $s_j^{q_j} = 1$ for all $j$. Since the $s_j$ commute with each other, by Exercise C-1.33(i), we have $1 = x^q = (\prod_{j \neq i} s_j)^q$, where $q = \prod_{j \neq i} q_j$. Since $(p_i^n, q) = 1$, there are integers $u$ and $v$ with $1 = up_i^n + vq$, and so $x = x^1 = x^{up_i^n} x^{vq} = 1$. Thus, $G$ is the direct product of its Sylow subgroups. •

We can now generalize Exercise C-1.23 on page 16 and its solution.

**Lemma C-1.39.** *There is no nonabelian simple group $G$ of order $|G| = p^e m$, where $p$ is prime, $p \nmid m$, and $p^e \nmid (m-1)!$.*

**Proof.** Suppose that such a simple group $G$ exists. By Sylow's Theorem, $G$ contains a subgroup $P$ of order $p^e$, hence of index $m$. We may assume that $m > 1$, for nonabelian $p$-groups are never simple. By Theorem C-1.2, there exists a homomorphism $\varphi\colon G \to S_m$ with $\ker \varphi \subseteq P$. Since $G$ is simple, however, it has no proper normal subgroups; hence, $\ker \varphi = \{1\}$ and $\varphi$ is an injection; that is, $G \cong \varphi(G) \subseteq S_m$. By Lagrange's Theorem, $p^e m \mid m!$, and so $p^e \mid (m-1)!$, contrary to the hypothesis. •

**Proposition C-1.40.** *There are no nonabelian simple groups of order less than* 60.

[7] If $n = p^e m$, where $p$ is a prime not dividing $m$, then $p \nmid \binom{n}{p^e}$; otherwise, cross multiply and use Euclid's Lemma.

[8] Such finite groups $G$ are called ***nilpotent***.

**Proof.** The reader may now check that the only integers $n$ between 2 and 59, neither a prime power nor having a factorization of the form $n = p^e m$ as in the statement of Lemma C-1.39, are $n = 30$, 40, and 56, and so these three numbers are the only candidates for orders of nonabelian simple groups of order $< 60$.

Suppose there is a simple group $G$ of order 30. Let $P$ be a Sylow 5-subgroup of $G$, so that $|P| = 5$. The number $r_5$ of conjugates of $P$ is a divisor of 30 and $r_5 \equiv 1 \bmod 5$. Now $r_5 \neq 1$ lest $P \lhd G$, so that $r_5 = 6$. By Lagrange's Theorem, the intersection of any two of these is trivial (intersections of Sylow subgroups can be more complicated; see Exercise C-1.35 on page 32). There are four nonidentity elements in each of these subgroups, and so there are $6 \times 4 = 24$ nonidentity elements in their union. Similarly, the number $r_3$ of Sylow 3-subgroups of $G$ is 10 (for $r_3 \neq 1$, $r_3$ is a divisor of 30, and $r_3 \equiv 1 \bmod 3$). There are two nonidentity elements in each of these subgroups, and so the union of these subgroups has 20 nonidentity elements. We have exceeded 30, the number of elements in $G$, and so $G$ cannot be simple.

Let $G$ be a group of order 40, and let $P$ be a Sylow 5-subgroup of $G$. If $r$ is the number of conjugates of $P$, then $r \mid 40$ and $r \equiv 1 \bmod 5$. These conditions force $r = 1$, so that $P \lhd G$; therefore, no simple group of order 40 can exist.

Finally, suppose there is a simple group $G$ of order 56. If $P$ is a Sylow 7-subgroup of $G$, then $P$ must have $r_7 = 8$ conjugates (for $r_7 \mid 56$ and $r_7 \equiv 1 \bmod 7$). Since these groups are cyclic of prime order, the intersection of any pair of them is $\{1\}$, and so there are 48 nonidentity elements in their union. Thus, adding the identity, we have accounted for 49 elements of $G$. Now a Sylow 2-subgroup $Q$ has order 8, and so it contributes 7 more nonidentity elements, giving 56 elements. But there is a second Sylow 2-subgroup, lest $Q \lhd G$, and we have exceeded our quota. Therefore, there is no simple group of order 56. •

The order of the next nonabelian simple group is 168.

The "converse" of Lagrange's Theorem is false: if $G$ is a finite group of order $n$ and $d \mid n$, then $G$ may not have a subgroup of order $d$. For example, we proved, in Proposition A-4.67 in Part 1, that the alternating group $A_4$ is a group of order 12 having no subgroup of order 6.

**Proposition C-1.41.** *Let $G$ be a finite group. If $p$ is prime and $p^k$ divides $|G|$, then $G$ has a subgroup of order $p^k$.*

**Proof.** If $|G| = p^e m$, where $p \nmid m$, then a Sylow $p$-subgroup $P$ of $G$ has order $p^e$. Hence, if $p^k$ divides $|G|$, then $p^k$ divides $|P|$. By Proposition C-1.25, $P$ has a subgroup of order $p^k$; a fortiori, $G$ has a subgroup of order $p^k$. •

What examples of $p$-groups have we seen? Of course, cyclic groups of order $p^n$ are $p$-groups, as is any direct product of copies of these. By the Fundamental Theorem, this describes all (finite) abelian $p$-groups. The only nonabelian examples we have seen so far are the dihedral groups $D_{2n}$ (which are 2-groups when $n$ is a power of 2) and the quaternions $\mathbf{Q}$ of order 8 (of course, for every 2-group $A$, the direct products $D_8 \times A$ and $\mathbf{Q} \times A$ are also nonabelian 2-groups). Here are some new examples.

**Definition.** A ***unitriangular*** matrix over a field $k$ is an upper triangular matrix each of whose diagonal terms is 1. Define $\mathrm{UT}(n,k)$ to be the set of all $n \times n$ unitriangular matrices over $k$.

For example, $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$ is upper triangular, and $\begin{bmatrix} 1 & a_{12} & a_{13} \\ 0 & 1 & a_{23} \\ 0 & 0 & 1 \end{bmatrix}$ is unitriangular.

**Remark.** We can generalize this definition by allowing $k$ to be any commutative ring. For example, the group $\mathrm{UT}(n,\mathbb{Z})$ is an interesting group; it is a finitely generated torsion-free nilpotent group (we will define (possibly infinite) *nilpotent groups* in the next chapter). ◀

**Proposition C-1.42.** *If $k$ is a field, then $\mathrm{UT}(n,k)$ is a subgroup of $\mathrm{GL}(n,k)$.*

**Proof.** Of course, the identity matrix $I$ is unitriangular, so that $I \in \mathrm{UT}(n,k)$. If $A \in \mathrm{UT}(n,k)$, then $A = I + N$, where $N$ is *strictly* upper triangular; that is, $N$ is an upper triangular matrix having only 0's on its diagonal. Note that $N^n = 0$, by Exercise A-7.17 on page 269 in Part 1.

The same Exercise A-7.17 also says that if $N, M$ are strictly upper triangular, then so are $N + M$ and $NM$. Hence, $(I + N)(I + M) = I + (N + M + NM)$ is unitriangular, and $\mathrm{UT}(n,k)$ is closed. Now unitriangular matrices are nonsingular because $\det(I + N) = 1$, but a proof of this without determinants will also show that $A^{-1} = (I + N)^{-1}$ is unitriangular. Recalling the power series expansion $1/(1 + x) = 1 - x + x^2 - x^3 + \cdots$, we define $B = I - N + N^2 - N^3 + \cdots$ (this series stops after $n - 1$ terms because $N^n = 0$). The reader may now check that $BA = B(I + N) = I$, so that $B = A^{-1}$. Moreover, $N$ strictly upper triangular implies that $-N + N^2 - N^3 + \cdots \pm N^{n-1}$ is also strictly upper triangular, and so $A^{-1} = B$ is unitriangular. Therefore, $\mathrm{UT}(n,k)$ is a subgroup of $\mathrm{GL}(n,k)$. •

**Proposition C-1.43.** *Let $q = p^e$, where $p$ is prime. For each $n \geq 2$, $\mathrm{UT}(n,\mathbb{F}_q)$ is a $p$-group of order $q^{\binom{n}{2}} = q^{n(n-1)/2}$.*

**Proof.** The number of entries in an $n \times n$ unitriangular matrix lying strictly above the diagonal is $\binom{n}{2} = \frac{1}{2}n(n-1)$ (throw away $n$ diagonal entries from the total of $n^2$ entries; half of the remaining $n^2 - n$ entries are above the diagonal). Since each of these entries can be any element of $\mathbb{F}_q$, there are exactly $q^{\binom{n}{2}}$ $n \times n$ unitriangular matrices over $\mathbb{F}_q$, and so this is the order of $\mathrm{UT}(n,\mathbb{F}_q)$. •

Recall Exercise A-4.31 on page 138 in Part 1: if $G$ is a group and $x^2 = 1$ for all $x \in G$, then $G$ is abelian. We now ask whether a group $G$ satisfying $x^p = 1$ for all $x \in G$, where $p$ is an odd prime, must also be abelian.

**Proposition C-1.44.** *If $p$ is an odd prime, then there exists a nonabelian group $G$ of order $p^3$ with $x^p = 1$ for all $x \in G$.*

**Proof.** If $G = \mathrm{UT}(3,\mathbb{F}_p)$, then $|G| = p^3$. Now $G$ is not abelian; for example, the matrices $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ do not commute. If $A \in G$, then $A = I + N$, where $N$

is strictly upper triangular; since $p$ is an odd prime, $p \geq 3$, and $N^p = 0$. Finally, Corollary A-3.6 in Part 1 says that

$$A^p = (I + N)^p = I^p + N^p = I. \quad \bullet$$

**Theorem C-1.45.** *Let $\mathbb{F}_q$ denote the finite field with $q$ elements. Then*

$$|\mathrm{GL}(n, \mathbb{F}_q)| = (q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1}).$$

**Proof.** Let $V$ be an $n$-dimensional vector space over $\mathbb{F}_q$. We show first that there is a bijection $\Phi\colon \mathrm{GL}(n, \mathbb{F}_q) \to \mathcal{B}$, where $\mathcal{B}$ is the set of all bases of $V$.[9] Choose, once for all, a basis $e_1, \ldots, e_n$ of $V$. If $T \in \mathrm{GL}(n, \mathbb{F}_q)$, define

$$\Phi(T) = Te_1, \ldots, Te_n.$$

By Lemma A-7.26 in Part 1, $\Phi(T) \in \mathcal{B}$ because $T$, being nonsingular, carries a basis into a basis. But $\Phi$ is a bijection, for given a basis $v_1, \ldots, v_n$, there is a unique linear transformation $S$, necessarily nonsingular (by Lemma A-7.26 in Part 1), with $Se_i = v_i$ for all $i$ (by Theorem A-7.28 in Part 1).

Our problem now is to count the number of bases $v_1, \ldots, v_n$ of $V$. There are $q^n$ vectors in $V$, and so there are $q^n - 1$ candidates for $v_1$ (the zero vector is not a candidate). Having chosen $v_1$, we see that the candidates for $v_2$ are those vectors not in $\langle v_1 \rangle$, the subspace spanned by $v_1$; there are thus $q^n - q$ candidates for $v_2$. More generally, having chosen a linearly independent list $v_1, \ldots, v_i$, we see that $v_{i+1}$ can be any vector not in $\langle v_1, \ldots, v_i \rangle$. Thus, there are $q^n - q^i$ candidates for $v_{i+1}$. The result follows by induction on $n$. $\bullet$

**Theorem C-1.46.** *If $p$ is prime and $q = p^m$, then the unitriangular group $\mathrm{UT}(n, \mathbb{F}_q)$ is a Sylow $p$-subgroup of $\mathrm{GL}(n, \mathbb{F}_q)$.*

**Proof.** Since $q^n - q^i = q^i(q^{n-i} - 1)$, the highest power of $p$ dividing $|\mathrm{GL}(n, \mathbb{F}_q)|$ is

$$qq^2q^3 \cdots q^{n-1} = q^{\binom{n}{2}}.$$

But $|\mathrm{UT}(n, \mathbb{F}_q)| = q^{\binom{n}{2}}$, and so $\mathrm{UT}(n, \mathbb{F}_q)$ must be a Sylow $p$-subgroup. $\bullet$

**Corollary C-1.47.** *If $p$ is prime and $G$ is a finite $p$-group, then $G$ is isomorphic to a subgroup of the unitriangular group $\mathrm{UT}(|G|, \mathbb{F}_p)$.*

**Proof.** A modest generalization of Exercise C-1.22 on page 16 shows, for any field $k$, that every group of order $m$ can be imbedded in $\mathrm{GL}(m, k)$. In particular, $G$ can be imbedded in $\mathrm{GL}(m, \mathbb{F}_p)$. Now $G$ is a $p$-group, and so it is contained in a Sylow $p$-subgroup $P$ of $\mathrm{GL}(m, \mathbb{F}_p)$, for every $p$-subgroup lies in some Sylow $p$-subgroup. Since all Sylow $p$-subgroups are conjugate, there is $a \in \mathrm{GL}(m, \mathbb{F}_p)$ with $P = a\,(\mathrm{UT}(m, \mathbb{F}_p))\,a^{-1}$. Therefore,

$$G \cong a^{-1}Ga \subseteq a^{-1}Pa \subseteq \mathrm{UT}(m, \mathbb{F}_p). \quad \bullet$$

[9] Recall that a basis of a finite-dimensional vector space is an (ordered) *list* of vectors, not merely a *set* of vectors. This distinction is critical in this proof.

A natural question is to find the Sylow subgroups of symmetric groups. This can be done, and the answer is in terms of a construction called *wreath product* (see Rotman [**188**], p. 176).

## Exercises

∗ **C-1.33.** (i) Let $G$ be an arbitrary (possibly nonabelian) group, and let $S$ and $T$ be normal subgroups of $G$. Prove that if $S \cap T = \{1\}$, then $st = ts$ for all $s \in S$ and $t \in T$.

**Hint.** Show that $sts^{-1}t^{-1} \in S \cap T$.

(ii) If $S_1, \ldots, S_n$ are normal subgroups of a group $G$ with $G = \langle S_1, \ldots, S_n \rangle$, prove that $G$ is the direct product $S_1 \times \cdots \times S_n$.

∗ **C-1.34.** Show that $S_4$ has more than one Sylow 2-subgroup.

∗ **C-1.35.** Give an example of a finite group $G$ having Sylow $p$-subgroups (for some prime $p$) $P$, $Q$, and $R$ such that $P \cap Q = \{1\}$ and $P \cap R \neq \{1\}$.

**Hint.** Consider $S_3 \times S_3$.

∗ **C-1.36.** A subgroup $H$ of a group $G$ is called ***characteristic*** if $\varphi(H) \subseteq H$ for every isomorphism $\varphi\colon G \to G$. A subgroup $S$ of a group $G$ is called ***fully invariant*** if $\varphi(S) \subseteq S$ for every homomorphism $\varphi\colon G \to G$.

(i) Prove that every fully invariant subgroup is a characteristic subgroup, and that every characteristic subgroup is a normal subgroup.

(ii) Prove that the commutator subgroup, $G'$, is a normal subgroup of $G$ by showing that it is a fully invariant subgroup.

(iii) Give an example of a group $G$ having a normal subgroup $H$ that is not a characteristic subgroup.

(iv) Prove that $Z(G)$, the center of a group $G$, is a characteristic subgroup (and so $Z(G) \lhd G$), but that it need not be a fully invariant subgroup.
**Hint.** Let $G = S_3 \times \mathbb{Z}_2$.

(v) For any group $G$, prove that if $H \lhd G$, then $Z(H) \lhd G$.

**C-1.37.** If $G$ is an abelian group, prove, for all positive integers $m$, that $mG$ and $G[m]$ are fully invariant subgroups.

**C-1.38.** Prove that $\mathrm{UT}(3, \mathbb{F}_2) \cong D_8$; conclude that $D_8$ is a Sylow 2-subgroup of $\mathrm{GL}(3, \mathbb{F}_2)$.

**Hint.** You may use the fact that the only nonabelian groups of order 8 are $D_8$ and $\mathbf{Q}$.

∗ **C-1.39.** Prove that the group of quaternions $\mathbf{Q}$ is isomorphic to a Sylow 2-subgroup of the special linear group $\mathrm{SL}(2, 5)$.

**Hint.** Use Exercise A-4.67 on page 159 in Part 1.

∗ **C-1.40.** (i) Exhibit all the subgroups of $S_4$; aside from $S_4$ and $\{(1)\}$, there are 26 of them.

(ii) Prove that if $d$ is a positive divisor of 24, then $S_4$ has a subgroup of order $d$.

(iii) If $d \neq 4$, prove that any two subgroups of $S_4$ having order $d$ are isomorphic.

**C-1.41.** Prove that a Sylow 2-subgroup of $A_5$ has exactly five conjugates.

**C-1.42.** (i) Find a Sylow 3-subgroup of $S_6$.
**Hint.** $\{1,2,3,4,5,6\} = \{1,2,3\} \cup \{4,5,6\}$.

(ii) Show that a Sylow 2-subgroup of $S_6$ is isomorphic to $D_8 \times \mathbb{Z}_2$.
**Hint.** $\{1,2,3,4,5,6\} = \{1,2,3,4\} \cup \{5,6\}$.

∗ **C-1.43.** (i) Prove that a Sylow 2-subgroup of $A_6$ is isomorphic to $D_8$.

(ii) Prove that a Sylow 3-subgroup of $A_6$ is isomorphic to $\mathbb{Z}_3 \times \mathbb{Z}_3$.

(iii) Prove that the normalizer of a Sylow 5-subgroup of $A_6$ is isomorphic to $D_{10}$.

∗ **C-1.44.** Let $Q$ be a normal $p$-subgroup of a finite group $G$. Prove that $Q \subseteq P$ for every Sylow $p$-subgroup $P$ of $G$.
**Hint.** Use the fact that any other Sylow $p$-subgroup of $G$ is conjugate to $P$.

**C-1.45.** (i) Let $G$ be a finite group and let $P$ be a Sylow $p$-subgroup of $G$. If $H \lhd G$, prove that $HP/H$ is a Sylow $p$-subgroup of $G/H$ and $H \cap P$ is a Sylow $p$-subgroup of $H$.
**Hint.** Show that $[G/H : HP/H]$ and $[H : H \cap P]$ are prime to $p$.

(ii) Let $P$ be a Sylow $p$-subgroup of a finite group $G$. Give an example of a subgroup $H$ of $G$ with $H \cap P$ not a Sylow $p$-subgroup of $H$.
**Hint.** Choose a subgroup $H$ of $S_4$ with $H \cong S_3$, and find a Sylow 3-subgroup $P$ of $S_4$ with $H \cap P = \{1\}$.

∗ **C-1.46.** Let $G$ be a group of order 90.

(i) If a Sylow 5-subgroup $P$ of $G$ is not normal, prove that it has six conjugates.
**Hint.** If $P$ has 18 conjugates, there are 72 elements in $G$ of order 5. Show that $G$ has more than 18 other elements.

(ii) Prove that $G$ is not simple.
**Hint.** Use Exercises C-1.19 and C-1.20(ii) on page 15.

**C-1.47.** Prove that there is no simple group of order 96, 120, 150, 300, 312, or 1000.

## C-1.3. Solvable and Nilpotent Groups

Galois introduced groups to investigate polynomials in $k[x]$, where $k$ is a field of characteristic 0, and he proved that such a polynomial is solvable by radicals if and only if its Galois group is a solvable group. Solvable groups are an interesting family of groups in their own right, and we now examine them a bit more. We first review some results from Part 1.

Recall that a ***normal series*** of a group $G$ is a finite sequence of subgroups, $G = G_0,\ G_1,\ \ldots,\ G_n$, with

$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n = \{1\}$$

and $G_{i+1} \lhd G_i$ for all $i$. The ***factor groups*** of the series are the groups

$$G_0/G_1, \quad G_1/G_2, \quad \ldots, \quad G_{n-1}/G_n,$$

and the ***length*** of the series is the number of strict inclusions (equivalently, the length is the number of nontrivial factor groups). A ***composition series*** of a group is a normal series of maximal length; its factor groups are called ***composition factors***.

**Definition.** Two normal series of a group $G$ are ***equivalent*** if there is a bijection between the lists of nontrivial factor groups of each so that corresponding factor groups are isomorphic.

The Jordan–Hölder Theorem says that any two composition series of a group are equivalent; it follows from a more general theorem, due to Schreier.

**Definition.** A ***refinement*** of a normal series of a group $G$ is a normal series $G = N_0, \ldots, N_k = \{1\}$ having the original series as a subsequence.

In other words, a refinement of a normal series is a normal series obtained from the original one by inserting more subgroups.

**Theorem C-1.48 (Schreier Refinement Theorem).** *Any two normal series*

$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n = \{1\}$$

*and*

$$G = N_0 \supseteq N_1 \supseteq \cdots \supseteq N_k = \{1\}$$

*of a group $G$ have equivalent refinements.*

**Proof.** Theorem A-5.29 in Part 1. •

## ▪ Solvable Groups

We now pass from general groups to solvable groups. We recall the definition.

**Definition.** A group $G$ is ***solvable*** if it has a normal series all of whose factor groups are abelian (British mathematicians call these groups ***soluble***).

In Part 1, solvable groups arose in determining those polynomials that are solvable by radicals, and so we focussed on finite solvable groups. But there are purely group-theoretic theorems about solvable groups making no direct reference to Galois theory and polynomials. For example, a theorem of Burnside says that if the order of a finite group $G$ is divisible by only two primes, that is, $|G| = p^m q^n$, where $p$ and $q$ are primes, then $G$ is solvable. The deep *Feit–Thompson Theorem* states that every group of odd order is solvable (Exercise C-1.59 on page 42 says that this is equivalent to every nonabelian finite simple group having even order).

Of course, every abelian group is solvable.

**Example C-1.49.** The ***infinite dihedral group***

$$D_\infty = \langle a, x : xax^{-1} = a^{-1}, x^2 = 1 \rangle$$

is solvable, for it has a normal series with factor groups $\mathbb{Z}$ and $\mathbb{Z}_2$. Note that $(xa)^2 = 1$ and $D_\infty = \langle xa, x \rangle$.

A group $G$ is ***polycyclic*** if it has a normal series each of whose factor groups is a (possibly infinite) cyclic group. Polycyclic groups are solvable. Every finitely

generated abelian group is polycyclic, as is $D_\infty$; thus, the finite dihedral groups $D_{2n}$ are also solvable.

Every polycyclic group $G$ has a normal subgroup $H$ of finite index where $H$ has a normal series each of whose factor groups is infinite cyclic (it follows that $H$ is torsion-free). Moreover, an infinite polycyclic group has a nontrivial torsion-free abelian normal subgroup (Robinson [**181**], p. 153).

A group $G$ satisfies the ***maximal condition*** if every nonempty family $\mathcal{S}$ of subgroups of $G$ has a maximal element; that is, there is a subgroup $M \in \mathcal{S}$ with $S \subseteq M$ for all $S \in \mathcal{S}$. A group is polycyclic if and only if it is solvable and satisfies the maximal condition (Robinson [**181**], p. 152).

Mal′cev proved that every solvable subgroup of $\mathrm{GL}(n, \mathbb{Z})$ is polycyclic, while L. Auslander[10] and Swan, independently, proved the converse: every polycyclic group can be imbedded in $\mathrm{GL}(n, \mathbb{Z})$ for some $n$. See Wehrfritz [**230**] for a more complete account of polycyclic groups. ◀

Solvability of a group is preserved by standard group-theoretic constructions. For example, every quotient of a solvable group is itself a solvable group (Proposition A-5.22 in Part 1), and every subgroup of a solvable group is itself solvable (Proposition A-5.23 in Part 1). An extension of one solvable group by another is itself solvable (Proposition A-5.25 in Part 1): if $K \lhd G$ and both $K$ and $G/K$ are solvable, then $G$ is solvable. It follows that a direct product of solvable groups is itself solvable (Corollary A-5.26 in Part 1).

**Proposition C-1.50.** *Every finite p-group $G$ is solvable.*

**Proof.** If $G$ is abelian, then $G$ is solvable. Otherwise, its center, $Z(G)$, is a proper nontrivial normal abelian subgroup, by Theorem C-1.22. Now $Z(G)$ is solvable because it is abelian, and $G/Z(G)$ is solvable, by induction on $|G|$, and so $G$ is solvable, by Proposition A-5.25 in Part 1. •

It follows that a direct product of finite $p$-groups, for various primes $p$, is solvable.

**Definition.** If $G$ is a group and $x, y \in G$, then the ***commutator*** $[x, y]$ is the element

$$[x, y] = xyx^{-1}y^{-1}.$$

The ***commutator subgroup*** $G'$ of a group $G$ is the subgroup generated by all the commutators.

It is clear that two elements $x$ and $y$ in a group $G$ commute if and only if their commutator $[x, y]$ is 1.

The subset consisting of all the commutators, while closed under inverses, need not be closed under products, and so the *set* of all commutators may not be a subgroup. The smallest group in which a product of two commutators is not a commutator has order 96 (there are two such groups). See Carmichael's Exercise C-1.57 on page 42.

[10] Louis Auslander was a geometer; his brother Maurice Auslander was an algebraist.

**Proposition C-1.51.** *Let $G$ be a group.*

(i) *The commutator subgroup $G'$ is a normal subgroup of $G$.*

(ii) *$G/G'$ is abelian.*

(iii) *If $H \lhd G$ and $G/H$ is abelian, then $G' \subseteq H$.*

**Proof.**

(i) The inverse of a commutator $xyx^{-1}y^{-1}$ is itself a commutator: $[x,y]^{-1} = yxy^{-1}x^{-1} = [y,x]$. Therefore, each element of $G'$ is a product of commutators. But any conjugate of a commutator (and, hence, a product of commutators) is another commutator:

$$\begin{aligned} a[x,y]a^{-1} &= a(xyx^{-1}y^{-1})a^{-1} \\ &= axa^{-1}aya^{-1}ax^{-1}a^{-1}ay^{-1}a^{-1} \\ &= [axa^{-1}, aya^{-1}]. \end{aligned}$$

Therefore, $G' \lhd G$. (Alternatively, $G' \lhd G$ because it is fully invariant: if $\varphi\colon G \to G$ is a homomorphism, then $\varphi([x,y]) = [\varphi(x), \varphi(y)] \in G'$.)

(ii) If $aG', bG' \in G/G'$, then

$$aG'bG'(aG')^{-1}(bG')^{-1} = aba^{-1}b^{-1}G' = [a,b]G' = G',$$

and so $G/G'$ is abelian.

(iii) Suppose that $H \lhd G$ and $G/H$ is abelian. If $a, b \in G$, then $aHbH = bHaH$; that is, $abH = baH$, and so $b^{-1}a^{-1}ba \in H$. As every commutator has the form $b^{-1}a^{-1}ba$, we have $G' \subseteq H$. •

**Example C-1.52.**

(i) A group $G$ is abelian if and only if $G' = \{1\}$.

(ii) If $G$ is a simple group, then $G' = \{1\}$ or $G' = G$, for $G'$ is a normal subgroup. The first case occurs when $G$ has prime order; the second case occurs otherwise. In particular, $(A_n)' = A_n$ for all $n \geq 5$. A group $G$ for which $G' = G$ is called ***perfect***; thus, every nonabelian simple group is perfect. There are perfect groups which are not simple; see Exercise C-1.53 on page 41.

(iii) What is $(S_n)'$? Since $S_n/A_n \cong \mathbb{Z}_2$ is abelian, Proposition C-1.51 shows that $(S_n)' \subseteq A_n$. For the reverse inclusion, first note that $(S_n)' \cap A_n \lhd A_n$; hence, if $n \geq 5$, simplicity of $A_n$ implies that this intersection is trivial or $A_n$. But $(S_n)' \cap A_n \neq \{(1)\}$, so $(S_n)' \cap A_n = A_n$ and $A_n \subseteq (S_n)'$. Therefore, $(S_n)' = A_n$ for all $n \geq 5$. Exercise C-1.53 on page 41 shows that the equality $(S_n)' = A_n$ also holds for $n = 2, 3, 4$. ◀

Let us iterate the formation of the commutator subgroup.

**Definition.** The ***derived series***[11] of $G$ is

$$G = G^{(0)} \supseteq G^{(1)} \supseteq G^{(2)} \supseteq \cdots \supseteq G^{(i)} \supseteq G^{(i+1)} \supseteq \cdots ,$$

where $G^{(0)} = G$, $G^{(1)} = G'$, and, more generally, $G^{(i+1)} = (G^{(i)})'$ for all $i \geq 0$.

It is easy to prove, by induction on $i \geq 0$, that $G^{(i)}$ is fully invariant (see Exercise C-1.36 on page 32), which implies that $G^{(i)} \lhd G$. It follows that $G^{(i)} \lhd G^{(i-1)}$, and so the derived series is a normal series. We now prove that $G$ is solvable if and only if its derived series reaches $\{1\}$.

**Proposition C-1.53.** *A group $G$ is solvable if and only if there is some $n$ with $G^{(n)} = \{1\}$.*

**Proof.** If $G$ is solvable, there is a normal series

$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n = \{1\}$$

whose factor groups $G_i/G_{i+1}$ are abelian. We show, by induction on $i \geq 0$, that $G^{(i)} \subseteq G_i$. Since $G^{(0)} = G = G_0$, the base step is obviously true. For the inductive step, since $G_i/G_{i+1}$ is abelian, Proposition C-1.51 gives $(G_i)' \subseteq G_{i+1}$. On the other hand, the inductive hypothesis gives $G^{(i)} \subseteq G_i$, which implies that

$$G^{(i+1)} = (G^{(i)})' \subseteq (G_i)' \subseteq G_{i+1}.$$

In particular, $G^{(n)} \subseteq G_n = \{1\}$, which is what we wished to show.

Conversely, if $G^{(n)} = \{1\}$, then the derived series is a normal series (a normal series must end with $\{1\}$) with abelian factor groups, and so $G$ is solvable. •

For example, the derived series of $G = S_4$ is easily seen to be

$$S_4 \supsetneq A_4 \supsetneq \mathbf{V} \supsetneq \{(1)\},$$

where $\mathbf{V}$ is the four-group.

Nowadays, most authors define a solvable group as one whose derived series reaches $\{1\}$ after a finite number of steps. In Exercise C-1.56 on page 41, the reader is asked to prove, using the criterion in Proposition C-1.53, that subgroups, quotient groups, and extensions of solvable groups are also solvable.

**Definition.** A normal subgroup $H$ of a group $G$ is a ***minimal normal subgroup*** if $H \neq \{1\}$ and there is no normal subgroup $K$ of $G$ with $\{1\} \subsetneq K \subsetneq H$.

Minimal normal subgroups may not exist (the abelian group $\mathbb{Q}$ has none), but they always exist in nontrivial finite groups.

**Theorem C-1.54.** *If $G$ is a finite solvable group, then every minimal normal subgroup $H$ is elementary abelian; that is, $H$ is a vector space over $\mathbb{Z}_p$ for some prime $p$.*

[11] I wonder whether the derived series is so-called because the notation $G'$ looks like the notation for the derivative $f'(x)$.

**Proof.** Exercise C-1.36 on page 32 says that if $K$ char $H$, then $K \lhd G$. Since $H$ is minimal, either $K = \{1\}$ or $K = H$. In particular, $H'$ char $H$, so that $H' = \{1\}$ or $H' = H$. As $G$ is solvable, so is its subgroup $H$, so that $H' \neq H$; hence, $H$ is abelian. Since $H$ is abelian, a Sylow $p$-subgroup $P$ of $H$, for some prime $p$, is characteristic in $H$, so that we may assume that $H$ is an abelian $p$-group. Finally, $\{x \in H : x^p = 1\}$ char $H$, so that minimality gives $H$ an elementary $p$-group. •

**Corollary C-1.55.** *Every finite solvable group $G$ acts on a finite-dimensional vector space over $\mathbb{F}_p$ for some prime $p$.*

**Proof.** If $H \lhd G$ is a normal subgroup of $G$, then $G$ acts on $H$ by conjugation. Now let $H$ be a minimal normal subgroup of $G$. •

The next theorem was proved by Philip Hall in 1928 (there is a contemporary group theorist named Marshall Hall, Jr.). We begin with a lemma (proved in 1885) which turns out to be quite useful.

**Lemma C-1.56 (Frattini Argument).** *Let $K$ be a normal subgroup of a finite group $G$. If $P$ is a Sylow $p$-subgroup of $K$ (for some prime $p$), then*

$$G = KN_G(P).$$

**Proof.** If $g \in G$, then $gPg^{-1} \subseteq gKg^{-1} = K$, because $K \lhd G$. It follows that $gPg^{-1}$ is a Sylow $p$-subgroup of $K$, and so there exists $k \in K$ with $kPk^{-1} = gPg^{-1}$. Hence, $P = (k^{-1}g)P(k^{-1}g)^{-1}$, so that $k^{-1}g \in N_G(P)$. The required factorization is thus $g = k(k^{-1}g)$. •

**Theorem C-1.57 (P. Hall).** *If $G$ is a finite solvable group of order $ab$, where $\gcd(a, b) = 1$, then $G$ contains a subgroup of order $a$.*

**Remark.** Hall also proved that any two subgroups of order $a$ are conjugate (see Rotman [**188**], p. 108). ◄

**Proof.** The proof is by induction on $|G| \geq 1$, the base step being trivially true.

*Case* 1: $G$ contains a normal subgroup $H$ of order $a'b'$, where $a' \mid a$, $b' \mid b$, and $b' < b$.

In this case, $G/H$ is a solvable group of order $(a/a')(b/b')$, which is strictly less than $ab = |G|$. By induction, $G/H$ has a subgroup $A/H$ of order $a/a'$, where $A \subseteq G$. Now $A$ has order $(a/a')|H| = ab' < ab$. As $A$ is solvable, induction shows that it, and hence $G$, has a subgroup of order $a$.

If there is some proper normal subgroup of $G$ whose order is not divisible by $b$, then the theorem has been proved. We may, therefore, assume that $b \mid |N|$ for every proper normal subgroup $N$. However, if $N$ is a minimal normal subgroup (which exists because $G$ is finite), then Theorem C-1.54 says that $N$ is an (elementary) abelian $p$-group for some prime $p$. Thus, we may assume that $b = p^m$, so that $N$ is a Sylow $p$-subgroup of $G$. Moreover, $N$ is the unique such subgroup, for normal Sylow subgroups are unique (Corollary C-1.33). We have reduced the problem to the following case.

*Case* 2: $|G| = ap^m$, where $p \nmid a$, $G$ has a normal abelian Sylow $p$-subgroup $H$, and $H$ is the unique minimal normal subgroup in $G$.

**Remark.** One of the first results of *cohomology of groups* is the Schur–Zassenhaus Lemma, from which this case follows at once. ◀

The group $G/H$ is a solvable group of order $a$. If $S$ is a minimal normal subgroup of $G/H$, then $|S/H| = q^n$ for some prime $q \neq p$. Hence, $|S| = p^m q^n$; if $Q$ is a Sylow $q$-subgroup of $S$, then $S = HQ$. Let $N^* = N_G(Q)$, and let $N = N^* \cap S = N_S(Q)$. We claim that $|N^*| = a$.

The Frattini Argument, Lemma C-1.56, gives $G = SN^*$. Since

$$G/S = SN^*/S \cong N^*/(N^* \cap S) = N^*/N,$$

we have $|N^*| = |G||N|/|S|$. But $S = HQ$ and $Q \subseteq N \subseteq S$ gives $S = HN$. Hence, $|S| = |HN| = |H||N|/|H \cap N|$, so that

$$\begin{aligned} |N^*| = |G||N|/|S| &= |G||N||H \cap N|/|H||N| \\ &= (|G|/|H|)|H \cap N| = a|H \cap N|. \end{aligned}$$

It follows that $|S| = a$ if $S \cap N = \{1\}$. We show that $H \cap N = \{1\}$ in two stages: (i) $H \cap N \subseteq Z(S)$; (ii) $Z(S) = \{1\}$.

(i) Let $x \in H \cap N$. Every $s \in S = HQ$ has the form $s = hy$ for $h \in H$ and $y \in Q$. Now $x$ commutes with $H$, for $H$ is abelian, and so it suffices to show that $x$ commutes with $y$. But $xyx^{-1}y^{-1} \in Q$, because $x$ normalizes $Q$, and $x(yx^{-1}y^{-1}) \in H$, because $H$ is normal. Therefore, $xyx^{-1}y^{-1} \in Q \cap H = \{1\}$.

(ii) Now $Z(S) \lhd G$, by Exercise C-1.36 on page 32. If $Z(S) \neq \{1\}$, then it contains a minimal subgroup which must be a minimal normal subgroup of $G$. Hence, $H \subseteq Z(S)$, for $H$ is the unique minimal normal subgroup of $G$. But since $S = HQ$, it follows that $Q$ char $S$. Thus, $Q \lhd G$, by Exercise C-1.36 on page 32, and so $H \subseteq Q$, a contradiction. Therefore, $Z(S) = \{1\}$, $H \cap N = \{1\}$, and $|N^*| = a$. •

Because of this theorem, a subgroup of a finite group is called a ***Hall subgroup*** if its order and index are relatively prime. For example, Sylow subgroups are Hall subgroups. In more detail, we may call $H$ a ***Hall $p$-subgroup*** if it is a Hall subgroup and a $p$-group, while we may call a subgroup $K$ a ***Hall $p'$-subgroup*** if it is a Hall subgroup and $p \nmid |K|$.

**Definition.** If $G$ be a group of order $ap^k$, where $p$ is prime and $p \nmid a$, then a ***$p$-complement*** is a subgroup of order $a$.

A ***$p'$-complement*** of a finite group $G$ is a subgroup $H$ such that $p \nmid |H|$ and $[G : H]$ is a power of $p$.

If a $p$-complement exists, then it is a Hall subgroup. Let $G$ be a finite group of order $ap^k$, where $p$ is a prime not dividing $a$. If $P$ is a Sylow $p$-subgroup of $G$ and $G$ has a $p$-complement $H$, then $|G| = |HP|$ and $G = HP$.

In general, $p$-complements do not exist. Up to isomorphism, there is only one group of order 15 (Rotman [**188**], p. 83), namely, $\mathbb{Z}_{15}$. Hence, $A_5$, a group of order $60 = 2^2 \cdot 15$, has no 2-complement, for it has no element of order 15.

Hall's Theorem implies that a finite solvable group has a $p$-complement for every prime $p$. Now if a finite group $G$ has order $p^m q^n$, where $p$ and $q$ are primes, then $G$ has a $p$-complement and a $q$-complement (namely, a Sylow $q$-subgroup and a Sylow $p$-subgroup). In the coming proof of the converse of Theorem C-1.57, we will use a special case: ***Burnside's Theorem***: every group of order $p^m q^n$ is solvable (Burnside's Theorem is proved in Chapter C-2, our chapter on representation theory).

Here is a second theorem of Philip Hall.

**Theorem C-1.58 (P. Hall).** *If a finite group $G$ has a $p$-complement for every prime divisor $p$ of $|G|$, then $G$ is solvable.*

**Proof.** The proof is by induction on $|G|$. Assume, on the contrary, that there are nonsolvable groups $G$ satisfying the hypothesis; choose such a $G$ of smallest possible order.

If $G$ has a nontrivial normal subgroup $N$ and $H$ is a Hall $p'$-subgroup of $G$, then checking orders shows that $H \cap N$ is a Hall $p'$-subgroup of $N$ and $HN/N \cong H/(H \cap N)$ is a $p'$-subgroup of $G/N$. Then both $N$ and $G/N$ are solvable, for their orders are strictly smaller than $|G|$. But $G$ is an extension of $N$ by $G/N$, so that $G$ is solvable, a contradiction.

Therefore, we may assume that $G$ is simple. Let $|G| = p_1^{e_1} \cdots p_n^{e_n}$, where the $p_i$ are distinct primes and $e_i > 0$ for all $i$. For each $i$, let $H_i$ be a Hall $p_i$-complement of $G$, so that $[G : H_i] = p_i^{e_i}$ and $|H_i| = \prod_{j \neq i} p_j^{e_j}$, by Exercise C-1.65 on page 42. If $D = H_3 \cap \cdots \cap H_n$, then $[G : D] = \prod_{i=3}^{n} p_i^{e_i}$ and $|D| = p_1^{e_1} p_2^{e_2}$. Now $D$ is solvable, by Burnside's Theorem. If $N$ is a minimal normal subgroup of $D$, then Theorem C-1.54 says that $N$ is elementary abelian; for notation, assume that $N$ is a $p_1$-group. Exercise C-1.65 shows that $[G : D \cap H_2] = \prod_{i=2}^{n} p_i^{e_i}$; thus, $D$ is a Sylow $p_1$-subgroup of $D$. By Exercise C-1.44 on page 33, $N \subseteq D \cap H_2$, and so $N \subseteq H_2$. But, as above, $|D \cap H_1| = p_2^{e_2}$, and comparison of orders gives $G = H_2(D \cap H_1)$. If $g \in G$, then $g = hd$, where $h \in H_2$ and $d \in D \cap H_1$. If $x \in N$, then $gxg^{-1} = hdxd^{-1}h^{-1} = hth^{-1}$, where $y = dxd^{-1} \in N$ (because $N \lhd D$) and $hyh^{-1} \in H_2$ (because $N \subseteq H_2$). Therefore, $N^G \subseteq H_2$, where $N^G$ is the normal subgroup of $G$ generated by $N$. Since $H_2 \subsetneq G$, we have $N^G \neq \{1\}$ a proper normal subgroup of $G$, contradicting the hypothesis that $G$ is simple. •

This proof exhibits two recurring themes in finite group theory. Many theorems having the form "If a finite group $G$ has property $P$, then it also has property $Q$" are proved by induction on $|G|$ in the following style: assume that $G$ is a ***least criminal***; that is, $G$ is a group of smallest order having property $P$ but not property $Q$, and reach a contradiction. The second theme is the reduction of a problem to the special case of simple groups. This is one reason why the Classification Theorem of Finite Simple Groups is so important.

See Doerk–Hawkes [**55**] for more about solvable groups.

## Exercises

**C-1.48.** Prove that every finitely generated solvable torsion group is finite.

**C-1.49.** Let $G$ be a polycyclic group. Prove that the number $h$, called the ***Hirsch length*** or ***Hirsch number***, of infinite cyclic factor groups in a normal series of $G$ is independent of the series.

**C-1.50.** If $f\colon G \to H$ is a group homomorphism, prove that $f(G') \subseteq H'$ and the induced map $f^{\#}$, given by $f^{\#}\colon gG' \mapsto f(g)H'$, is a homomorphism $G/G' \to H/H'$. Conclude that there is a functor $F\colon \mathbf{Groups} \to \mathbf{Ab}$ with $F(G) = G/G'$.

**C-1.51.** Let $p$ be prime and let $G$ be a nonabelian group of order $p^3$. Prove that $Z(G) = G'$.

**Hint.** Show first that both subgroups have order $p$.

**C-1.52.** Prove that if $H$ is a subgroup of a group $G$ and $G' \subseteq H$, then $H \lhd G$.

**Hint.** Use the Correspondence Theorem.

∗ **C-1.53.** (i) Prove that $(S_n)' = A_n$ for $n = 2, 3, 4$ (see Example C-1.52(ii) for $n \geq 5$).

(ii) Prove that $(\mathrm{GL}(n,k))' \subseteq \mathrm{SL}(n,k)$. (The reverse inclusion is also true; see Exercise C-1.81 on page 58 for the case $n = 2$.)

(iii) Prove that $\mathrm{SL}(2,5)$ is a perfect group which is not simple.

∗ **C-1.54.** If $G$ is a finite group and $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n = \{1\}$ is a normal series, prove that the order of $G$ is the product of the orders of the factor groups:

$$|G| = \prod_{i=0}^{n-1} |G_i/G_{i+1}|.$$

**C-1.55.** Prove that any two finite solvable groups of the same order have the same composition factors.

∗ **C-1.56.** Let $G$ be an arbitrary, possibly infinite, group.

(i) Prove that if $H \subseteq G$, then $H^{(i)} \subseteq G^{(i)}$ for all $i$. Conclude, using Proposition C-1.53, that every subgroup of a solvable group is solvable.

(ii) Prove that if $f\colon G \to K$ is a surjective homomorphism, then

$$f(G^{(i)}) = K^{(i)}$$

for all $i$. Conclude, using Proposition C-1.53, that every quotient of a solvable group is also solvable.

(iii) For every group $G$, prove, by double induction, that

$$G^{(m+n)} = (G^{(m)})^{(n)}.$$

(iv) Prove, using Proposition C-1.53, that if $H \lhd G$ and both $H$ and $G/H$ are solvable, then $G$ is solvable.

∗ **C-1.57.** (**Carmichael**) Let $G$ be the subgroup of $S_{16}$ generated by the following permutations:

$$\begin{aligned} (a\ c)(b\ d); &\qquad (e\ g)(f\ h);\\ (i\ k)(j\ \ell); &\qquad (m\ o)(n\ p);\\ (a\ c)(e\ g)(i\ k); &\qquad (a\ b)(c\ d)(m\ o);\\ (e\ f)(g\ h)(m\ n)(o\ p); &\qquad (i\ j)(k\ \ell). \end{aligned}$$

Prove that $|G| = 256$, $|G'| = 16$, and

$$(i\ k)(j\ \ell)(m\ o)(n\ p) \in G'$$

is not a commutator.

**C-1.58.** Let $p$ and $q$ be primes.

(i) Prove that every group of order $pq$ is solvable.
**Hint.** If $p = q$, then $G$ is abelian. If $p < q$, then a divisor $r$ of $pq$ for which $r \equiv 1 \bmod q$ must equal 1.

(ii) Prove that every group $G$ of order $p^2q$ is solvable.
**Hint.** If $G$ is not simple, use Proposition A-5.25 in Part 1. If $p > q$, then $r \equiv 1 \bmod p$ forces $r = 1$. If $p < q$, then $r = p^2$ and there are more than $p^2q$ elements in $G$.

∗ **C-1.59.** Show that the Feit–Thompson Theorem, "Every finite group of odd order is solvable," is equivalent to "Every nonabelian finite simple group has even order."

**Hint.** For sufficiency, choose a "least criminal": a nonsolvable group $G$ of smallest odd order. By hypothesis, $G$ is not simple, and so it has a proper nontrivial normal subgroup.

**C-1.60.** (i) Prove that the infinite cyclic group $\mathbb{Z}$ does not have a composition series.

(ii) Prove that an *abelian* group $G$ has a composition series if and only if $G$ is finite.

**C-1.61.** Prove that if $G$ is a finite group and $H \lhd G$, then there is a composition series of $G$ one of whose terms is $H$.

**Hint.** Use Schreier's Theorem.

**C-1.62.** (i) Prove that if $S$ and $T$ are solvable subgroups of a group $G$ and $S \lhd G$, then $ST$ is a solvable subgroup of $G$.

**Hint.** Use the Second Isomorphism Theorem.

(ii) If $G$ is a finite group, define $\mathcal{NS}(G)$ to be the subgroup of $G$ generated by all normal solvable subgroups of $G$. Prove that $\mathcal{NS}(G)$ is the unique maximal normal solvable subgroup of $G$ and that $G/\mathcal{NS}(G)$ has no nontrivial normal solvable subgroups.

**C-1.63.** (i) Prove that the dihedral groups $D_{2n}$ are solvable.

(ii) Give a composition series for $D_{2n}$.

**C-1.64.** (**Rosset**) Let $G$ be a group containing elements $x$ and $y$ such that the orders of $x$, $y$, and $xy$ are pairwise relatively prime; prove that $G$ is not solvable.

∗ **C-1.65.** (i) If $H$ and $K$ are subgroups of finite index in a group $G$, prove that $H \cap K$ also has finite index in $G$.

**Hint.** Show that $[G : H \cap K] \leq [G : H][G : K]$.

(ii) If $H$ has finite index in $G$, prove that the intersection of all the conjugates of $H$ is a normal subgroup of $G$ having finite index in $G$. Conclude that an infinite simple group has no proper subgroup of finite index.

(iii) If $\gcd([G:H],[G:K])=1$, prove that $[G:H\cap K]=[G:H][G:K]$.

## ▪ Nilpotent Groups

It turns out that the same methods giving properties of $p$-groups extend to a larger class, *nilpotent groups*, which may be regarded as generalized $p$-groups. We first introduce a convenient notation.

**Notation.** If $X$ and $Y$ are nonempty subsets of a group $G$, then $[X,Y]$ is defined by

$$[X,Y]=\big\langle [x,y] : x\in X \text{ and } y\in Y\big\rangle,$$

the subgroup of $G$ generated by all commutators $[x,y]=xyx^{-1}y^{-1}$, where $x\in X$ and $y\in Y$.

The commutator subgroup $G'$ can be written as $[G,G]$, and the higher commutator subgroups can be denoted by $G^{(i+1)}=[G^{(i)},G^{(i)}]$. Since $[x,y]=[y,x]^{-1}$, we see that $[X,Y]=[Y,X]$.

The following lemma relates $[H,K]$ and centers.

**Lemma C-1.59.** *Let $H$ and $K$ be subgroups of a group $G$.*

(i) *If $K\lhd G$ and $K\subseteq H\subseteq G$, then*

$$[H,G]\subseteq K \text{ if and only if } H/K\subseteq Z(G/K).$$

(ii) *If $f\colon G\to L$, for some group $L$, then $f([H,K])=[f(H),f(K)]$.*

**Proof.**

(i) If $h\in H$ and $g\in G$, then $hKgK=gKhK$ if and only if $[h,g]K=K$ if and only if $[h,g]\in K$.

(ii) Both sides are generated by all $f([h,k])=[f(h),f(k)]$. •

**Definition.** Define subgroups $\gamma_i(G)$ of a group $G$ by induction:

$$\gamma_1(G)=G; \qquad \gamma_{i+1}(G)=[\gamma_i(G),G].$$

It is easy to see that the subgroups $\gamma_i(G)$ are characteristic, hence normal, subgroups of $G$.

**Definition.** The ***lower central series*** of a group $G$ (also called the ***descending central series***) is

$$G=\gamma_1(G)\supseteq\gamma_2(G)\supseteq\cdots,$$

where $\gamma_{i+1}(G)=[\gamma_i(G),G]$.

The lower central series is a normal series only if it reaches $\{1\}$.

**Definition.** A group $G$ is called ***nilpotent*** if the lower central series reaches $\{1\}$, that is, if $\gamma_{c+1}(G) = \{1\}$ for some $c$. The smallest integer $c$ for which $\gamma_{c+1}(G) = \{1\}$ is called the (nilpotency) ***class*** of $G$.

Note that $\gamma_2(G) = [G, G] = G' = G^{(1)}$, the commutator subgroup, but the derived series and the lower central series may differ afterward; for example, $\gamma_3(G) = [G', G] \supseteq G^{(2)}$, with strict inequality possible. Moreover, Lemma C-1.59(i) shows that $\gamma_{i+1}(G) = [\gamma_i(G), G]$ gives $\gamma_i(G)/\gamma_{i+1}(G) \subseteq Z(G/\gamma_{i+1}(G))$.

The class of nilpotent groups is closed under subgroups, quotients, and finite direct products, but it is not closed under extensions. Here are the details.

**Proposition C-1.60.**

(i) *Every nilpotent group $G$ is solvable.*

(ii) *If $G$ is nilpotent of class $c$ and $S$ is a subgroup, then $S$ is nilpotent of class $\leq c$.*

(iii) *If $G$ is nilpotent of class $c$ and $H \lhd G$, then $G/H$ is nilpotent of class $\leq c$.*

(iv) *If $H$ and $K$ are nilpotent, then the direct product $H \times K$ is nilpotent.* ◀

**Proof.**

(i) An induction on $i$ shows that $G^{(i)} \subseteq \gamma_i(G)$ for all $i$. It follows that if $\gamma_{c+1}(G) = \{1\}$, then $G^{(c+1)} = \{1\}$; that is, if $G$ is nilpotent, then $G$ is solvable.

(ii) An induction on $i$ shows that $S \subseteq G$ implies $\gamma_i(S) \subseteq \gamma_i(G)$ for all $i$. Hence, $\gamma_{c+1}(G) = \{1\}$ implies $\gamma_{c+1}(S) = \{1\}$.

(iii) If $f\colon G \to G/H$ is the natural map, then Lemma C-1.59(ii) shows that $\gamma_{c+1}(G/H) \subseteq f(\gamma_{c+1}(G))$; hence, $\gamma_{c+1}(G) = \{1\}$ forces $\gamma_{c+1}(G/H) = \{1\}$.

(iv) An induction on $i$ shows that $\gamma_i(H \times K) \subseteq \gamma_i(H) \times \gamma_i(K)$. Suppose that $H$ has class $c$ and $K$ has class $d$; that is, $\gamma_{c+1}(H) = \{1\}$ and $\gamma_{d+1}(K) = \{1\}$. If $m = \max\{c, d\}$, then $\gamma_{m+1}(H \times K) = \{1\}$. •

There is another series of interest.

**Definition.** The ***higher centers*** of a group $G$ are defined by induction:

$$\zeta^0(G) = \{1\}; \qquad \zeta^{i+1}(G)/\zeta^i(G) = Z\left(G/\zeta^i(G)\right);$$

that is, $\zeta^{i+1}(G)$ is the inverse image of the center $Z\left(G/\zeta^i(G)\right)$ under the natural map $G \to G/\zeta^i(G)$.

The first higher center is $\zeta^1(G) = Z(G)$.

**Definition.** The ***upper central series*** of a group $G$ (also called the ***ascending central series***) is

$$\{1\} = \zeta^0(G) \subseteq \zeta^1(G) = Z(G) \subseteq \zeta^2(G) \subseteq \cdots.$$

We abbreviate $\gamma_i(G)$ as $\gamma_i$ and $\zeta^i(G)$ as $\zeta^i$ in the next proof.

**Proposition C-1.61.** *If $G$ is a group, then there is an integer $c$ with $\gamma_{c+1}(G) = \{1\}$ (that is, $G$ is nilpotent of class $\leq c$) if and only if $\zeta^c(G) = G$. Moreover,*

$$\gamma_{i+1}(G) \subseteq \zeta^{c-i}(G)$$

*for all $i$.*

**Proof.** If $\zeta^c = G$, we prove that the inclusion holds by induction on $i$. If $i = 0$, then $\gamma_1 = G = \zeta^c$. For the inductive step, assume that $\gamma_{i+1} \subseteq \zeta^{c-i}$. Therefore,

$$\gamma_{i+2} = [\gamma_{i+1}, G] \subseteq [\zeta^{c-i}, G] \subseteq \zeta^{c-i+1},$$

the last inclusion following from Lemma C-1.59(i) with $K = \zeta^{c-i}$ and $H = \zeta^{c-i+1}$.

If $\gamma_{c+1} = \{1\}$, we prove by induction on $j$ that $\gamma_{c+1-j} \subseteq \zeta^j$ (this is the same inclusion as in the following statement: set $j = c - i$). If $j = 0$, then $\gamma_{c+1} = \{1\} = \zeta^0$. For the inductive step, if $\gamma_{c+1-j} \subseteq \zeta^j$, then the Third Isomorphism Theorem gives a surjective homomorphism $G/\gamma_{c+1-j} \to G/\zeta^j$. Now $[\gamma_{c-j}, G] = \gamma_{c+1-j}$, so that Lemma C-1.59(ii) gives $\gamma_{c-j}/\gamma_{c+1-j} \subseteq Z(G/\gamma_{c+1-j})$. But Exercise C-1.66 on page 49 (if $A \subseteq Z(G)$ and $f\colon G \to H$ is surjective, then $f(A) \subseteq Z(H)$), we have

$$\gamma_{c-j}\zeta^j \subseteq Z(G/\zeta^j) = \zeta^{j+1}/\zeta^j.$$

Therefore, the inclusion always holds. •

We restate Proposition C-1.61 in words.

**Corollary C-1.62.** *The lower central series of a group $G$ reaches $\{1\}$ if and only if the upper central series reaches $G$.*

**Proposition C-1.63.**

(i) *If $G$ is nilpotent and $G \neq \{1\}$, then $Z(G) \neq \{1\}$.*

(ii) *$S_3$ is a solvable group which is not nilpotent. Hence, if $H \lhd G$ and $G/H$ are nilpotent, then $G$ may not be nilpotent.*

**Proof.**

(i) Since $G \neq \{1\}$, it is nilpotent of class $c \geq 1$, so that $\gamma_{c+1}(G) = \{1\}$ and $\gamma_c(G) \neq \{1\}$. By Proposition C-1.61, $\{1\} \neq \gamma_c(G) \subseteq \zeta^1(G) = Z(G)$.

(ii) Since $Z(S_3) = \{(1)\}$, it follows from (i) that $S_3$ is not nilpotent. •

Since $S_3$ is not nilpotent, an extension of one nilpotent group by another need not be nilpotent. On the other hand, Exercise C-1.70 on page 49 says that if $G/Z(G)$ is nilpotent, then $G$ is nilpotent. There is a generalization due to P. Hall. First, it is easy to see that if $H \lhd G$, then its commutator subgroup $H'$ is also normal in $G$. Hall proved that if $H \lhd G$ and $G/H'$ are nilpotent, then $G$ is nilpotent (see Robinson [**181**], p. 134).

**Example C-1.64.**

(i) Finite unitriangular groups $\mathrm{UT}(n, \mathbb{F}_q)$ are nilpotent.

(ii) The unitriangular group $\mathrm{UT}(n, \mathbb{Z})$ is nilpotent.

(iii) The ***Heisenberg group*** $H_R$, where $R = \mathbb{R}$ or $R = \mathbb{C}$, is the $3\times3$ unitriangular group

$$H = H_R = \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix},$$

where $a, b, c \in R$. This group arises when discussing quantum mechanics.

(iv) If $G$ is a group, denote $\{g \in G : g \neq 1\}$ by $G^{\#}$. The following type of group arises in representation theory. A finite group $G$ is a ***Frobenius group*** if it has a subgroup $H$ such that $H \cap xHx^{-1} = \{1\}$ for every $x \notin H$. The subset $N$ of $G$, defined by $N = G - \bigcup_{x\in G} xH^{\#}x^{-1}$, is called the ***Frobenius kernel***; it turns out that $N$ is a (normal) subgroup of $G$. A theorem of Thompson (see Robinson [**181**], p. 306) says that the Frobenius kernel is nilpotent.

**Theorem C-1.65.** *Every finite p-group is nilpotent.*

**Proof.** Recall Theorem C-1.22; every finite $p$-group has a nontrivial center. If, for some $i$, we have $\zeta^i(G) \subsetneq G$, then $Z(G/\zeta^i(G)) \neq \{1\}$, and so $\zeta^i(G) \subsetneq \zeta^{i+1}(G)$. As $G$ is finite, there must be an integer $c$ with $\zeta^c(G) = G$; that is, $G$ is nilpotent. •

Here are several interesting characterizations of finite nilpotent groups.

**Theorem C-1.66.** *The following statements for a finite group $G$ are equivalent.*

(i) *$G$ is nilpotent.*

(ii) *Every subgroup $H$ of $G$ is **subnormal**: there are subgroups $G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_m = H$ with $G_i \lhd G_{i-1}$ for all $i \leq m$.*

(iii) *$G$ satisfies the **normalizer condition**: if $H \subsetneq G$, then $H \subsetneq N_G(H)$.*

(iv) *If $H$ is a maximal subgroup of $G$, then $H \lhd G$.*

(v) *$G$ is the direct product of its Sylow subgroups.*

**Proof.**

(i) $\Rightarrow$ (ii). Note, for all $i$, that $H\zeta^i$ is a subgroup of $G$, for $\zeta^i \lhd G$. Consider the series

$$H = H\zeta^0 \subseteq H\zeta^1 \subseteq H\zeta^2 \subseteq \cdots \subseteq H\zeta^c = G.$$

But $H\zeta^i \lhd H\zeta^{i+1}$ for all $i$ because $\zeta^{i+1}/\zeta^i = Z\left(G/\zeta^i(G)\right)$. Therefore, $H$ is a subnormal subgroup of $G$.

(ii) $\Rightarrow$ (iii). Since $H$ is subnormal, there is a series

$$H = H_0 \subseteq H_1 \subseteq \cdots \subseteq H_m = G$$

with $H_j \lhd H_{j+1}$ for all $j$. Since $H$ is a proper subgroup of $G$, there is some $j$ with $H \neq H_j$; if $k$ is the smallest such $j$, then $H = H_{k-1} \lhd H_k$, and so $H_k \subseteq N_G(H)$.

(iii) $\Rightarrow$ (iv). Since $H$ is a proper subgroup of $G$, we have $H \subsetneq N_G(H)$; that is, $H \subsetneq N_G(H) \subseteq G$. Since $H$ is a maximal subgroup of $G$, we have $N_G(H) = G$; that is, $H \lhd G$.

(iv) $\Rightarrow$ (v). It suffices to show that every Sylow subgroup $P$ is a normal subgroup of $G$, for then Proposition C-1.38 will show that $G$ is the direct product of its Sylow subgroups. If a Sylow subgroup $P$ is not normal, then its normalizer $N_G(P)$ is a proper subgroup of $G$ and, hence, it is contained in a maximal subgroup, say, $M$. By hypothesis, $M \lhd G$. But this contradicts Corollary C-1.34, which says that if $N = N_G(P)$, then $N = N_G(N)$.

(v) $\Rightarrow$ (i). Every $p$-group is nilpotent, and the direct product of nilpotent groups is nilpotent. •

This theorem is not true for infinite groups.

The following subgroup of a group $G$ is analgous to the *Jacobson radical* of a ring.

**Definition.** The ***Frattini subgroup*** $\Phi(G)$ of a group $G$ is defined to be the intersection of all the maximal subgroups of $G$. If $G$ has no maximal subgroups, define $\Phi(G) = G$.[12]

It is clear that $\Phi(G)$ char $G$, and so $\Phi(G) \lhd G$.

**Definition.** An element $x \in G$ is a ***nongenerator*** if can be deleted from any generating set of $G$; that is, if $G = \langle x, Y \rangle$, then $G = \langle Y \rangle$.

**Theorem C-1.67.** *The Frattini subgroup $\Phi(G)$ of any group $G$ is the set of all its nongenerators.*

**Proof.** Let $x$ be a nongenerator of $G$, and let $M$ be a maximal subgroup of $G$. If $x \notin M$, then $G = \langle x, M \rangle = M$, a contradiction. Therefore, $x \in M$ for all $M$; that is, $x \in \Phi(G)$.

Conversely, assume that $z \in \Phi(G)$ and $G = \langle z, Y \rangle$. If $\langle Y \rangle \neq G$, then there is a maximal subgroup $M$ with $\langle Y \rangle \subseteq M$. But $z \in M$, so that $G = \langle z, Y \rangle \subseteq M$, a contradiction. •

The next theorem was proved in 1885.

**Theorem C-1.68 (Frattini).** *Let $G$ be a finite group.*

(i) *$\Phi(G)$ is nilpotent.*

(ii) *If $G$ is a finite $p$-group, then $\Phi(G) = G'G^p$, where $G'$ is the commutator subgroup and $G^p$ is the subgroup of $G$ generated by all the $p$th powers.*

(iii) *If $G$ is a finite $p$-group, then $G/\Phi(G)$ is a vector space over $\mathbb{F}_p$.*

**Proof.**

(i) Let $P$ be a Sylow $p$-subgroup of $\Phi(G)$ for some prime $p$. Since $\Phi(G) \lhd G$, Lemma C-1.56 (the Frattini Argument) gives $G = \Phi(G)N_G(P)$. But $\Phi(G)$ consists of nongenerators, and so $G = N_G(P)$; that is, $P \lhd G$ and, hence,

[12] If $G$ is finite, then $G$ does have maximal subgroups; if $G$ is infinite, it may not have any maximal subgroups. For example, consider the additive group of rationals $\mathbb{Q}$. Since $\mathbb{Q}$ is abelian, any maximal subgroup $H$ would be normal, and so $\mathbb{Q}/H$ would be a simple abelian group; that is, $\mathbb{Q}/H \cong \mathbb{Z}_p$ for some prime $p$. But it is easy to see that $\mathbb{Q}$ has no finite nontrivial images.

$P \lhd \Phi(G)$. Therefore, $\Phi(G)$ is the direct product of its Sylow subgroups, by Proposition C-1.38, and so it is nilpotent, by Theorem C-1.66.

(ii) If $M$ is a maximal subgroup of $G$, where $G$ is now a $p$-group, then Theorem C-1.66 gives $M \lhd G$. But $G/M$ is a simple $p$-group (since $M$ is maximal), so that $G/M$ is abelian of order $p$. Hence, $G' \subseteq M$ and $G^p \subseteq M$. Therefore, $G'G^p \subseteq \Phi(G)$.

For the reverse inclusion, observe that $G/G'G^p$ is a finite abelian group of exponent $p$; that is, it is a vector space over $\mathbb{F}_p$. Clearly, $\Phi(G/G'G^p) = \{1\}$. If $H \lhd G$ and $H \subseteq \Phi(G)$, then it is easy to check that $\Phi(G)$ is the inverse image (under the natural map $G \to G/G'G^p$) of $\Phi(G/H)$ (for maximal subgroups correspond). It follows that $\Phi(G) = G'G^p$.

(iii) Since $\Phi(G) = G'G^p$, the quotient group $G/\Phi(G)$ is an abelian group of exponent $p$; that is, it is a vector space over $\mathbb{F}_p$. •

**Definition.** A subset $X$ of a group $G$ is a ***minimal generating set*** if $G = \langle X \rangle$ but no proper subset of $X$ generates $G$.

There is a competing candidate for a minimal generating set of a finite group $G$: namely, a generating set of least cardinality. Notice that these two notions can be distinct. For example, let $G = \langle a \rangle \times \langle b \rangle$, where $a$ has order 2 and $b$ has order 3. Now $X = \{a, b\}$ is a minimal generating set, for no proper subset generates $G$. On the other hand, $G = \langle ab \rangle \cong \mathbb{Z}_6$, and the generating sets $\{ab\}$ and $\{a, b\}$ have different cardinalities. The next theorem shows that these notions coincide when $G$ is a finite $p$-group.

If $p$ is a prime and $G$ is a finite *abelian* $p$-group, then $\Phi(G) = pG$.

**Theorem C-1.69 (Burnside Basis Theorem).** *If $G$ is a finite $p$-group, then $G/\Phi(G)$ is a vector space over $\mathbb{F}_p$, and its dimension is the minimum number of generators of $G$.*

**Proof.** If $\{x_1, \ldots, x_n\}$ is a minimal generating set of $G$, then the family of cosets $\{\bar{x}_1, \ldots, \bar{x}_n\}$ spans the vector space $G/\Phi(G)$, where $\bar{x}$ denotes the coset $x\Phi(G)$. If this family is linearly dependent, then one of them, say, $\bar{x}_1$, lies in $\langle \bar{x}_2, \ldots, \bar{x}_n \rangle$. Thus, there is $y \in \langle x_2, \ldots, x_n \rangle \subseteq G$ with $x_1 y^{-1} \in \Phi(G)$. Now $\{x_1 y^{-1}, x_2, \ldots, x_n\}$ generates $G$, so that $G = \langle \bar{x}_2, \ldots, \bar{x}_n \rangle$, by Theorem C-1.67, and this contradicts minimality. Therefore, $n = \dim G/\Phi(G)$, and all minimal generating sets have the same cardinality.

Conversely, if $x \notin \Phi(G)$, then $\bar{x} \neq 0$ in the vector space $G/\Phi(G)$, and so it is part of a basis, say, $\bar{x}, \bar{x}_2, \ldots, \bar{x}_n$. If $x_i$ represents the coset $\bar{x}_i$ for $i \geq 2$, then $G = \langle \Phi(G), x, x_2, \ldots, x_n \rangle = \langle x, x_2, \ldots, x_n \rangle$. Moreover, $\{x, x_2, \ldots, x_n\}$ is a minimal generating set, for the cosets of a proper subset do not generate $G/\Phi(G)$. •

For deeper results about solvable and nilpotent groups, we suggest the book of Lennox and Robinson [**140**]. See also the "Canadian notes" of P. Hall [**91**], which are also contained in Hall's collected works [**92**].

## Exercises

∗ **C-1.66.** Prove that if $A \subseteq Z(G)$ and $f\colon G \to H$ is surjective, then $f(A) \subseteq Z(H)$.

**C-1.67.** If $G$ is a finite nilpotent group and $x, y \in G$ have coprime orders, prove that $x$ and $y$ commute.

**C-1.68.** Let $G$ be a finite nilpotent group.

(i) If $G$ is a nilpotent group and $N$ is a nontrivial normal subgroup of $G$, prove that $N \cap Z(G) \neq \{1\}$.

(ii) Prove that a minimal normal subgroup $M$ of a nilpotent group $G$ is contained in $Z(G)$.

**C-1.69.** Let $\mathfrak{A}$ denote the class of all abelian groups, $\mathfrak{N}$ the class of all nilpotent groups, and $\mathfrak{S}$ the class of all solvable groups. Prove that $\mathfrak{A} \subseteq \mathfrak{N} \subseteq \mathfrak{S}$ and that each of the inclusions is strict; that is, there is a nilpotent group that is not abelian, and there is a solvable group that is not nilpotent.

∗ **C-1.70.** If $G$ is a group with $G/Z(G)$ nilpotent of class $c$, prove that $G$ is nilpotent of class $c + 1$.

**C-1.71.** (i) If $H$ and $K$ are nilpotent normal subgroups of a group $G$, prove that $HK$ is also a nilpotent normal subgroup.

**Hint.** Let $H$ and $K$ be nilpotent of class $c$ and $d$, respectively, and let $S = HK$. Prove, by induction on $c + d$, that $S$ is nilpotent. For the inductive step, use Exercise C-1.70 in considering $S/Z(H)$ and $S/Z(K)$.

(ii) Prove that every finite group $G$ has a unique maximal nilpotent normal subgroup. It is called the ***Fitting subgroup***, and it is denoted by $\mathrm{Fit}(G)$.

(iii) If $G$ is a finite group, prove that $\mathrm{Fit}(G)$ char $G$, and conclude that $\mathrm{Fit}(G) \lhd G$.

∗ **C-1.72.** (i) Let $X$ be a finite $G$-set, and let $H \subseteq G$ act transitively on $X$. Prove that $G = HG_x$ for each $x \in X$.

(ii) Show that the Frattini argument follows from (i).

**C-1.73.** For each $n \geq 1$, let $G_n$ be a finite $p$-group of class $n$. Define $H$ to be the subgroup of $\prod_{n\geq 1} G_n$ with $g_n = 1$ for large $n$; that is, only finitely many coordinates of $(g_1, g_2, \ldots)$ are distinct from 1. Prove that $H$ is an infinite $p$-group which is not nilpotent.

**C-1.74.** If $G$ is a group and $g, x \in G$, write $g^x = xgx^{-1}$.

(i) Prove, for all $x, y, z \in G$, that $[x, yz] = [x, y][x, z]^y$ and $[xy, z] = [y, z]^x[x, z]$.

(ii) (**Jacobi Identity**) If $x, y, z \in G$ are elements in a group $G$, define

$$[x, y, z] = [x, [y, z]].$$

Prove that

$$[x, y^{-1}, z]^y[y, z^{-1}, x]^z[z, x^{-1}, y]^x = 1.$$

**C-1.75.** If $H, K, L$ are subgroups of a group $G$, define

$$[H, K, L] = \big\langle \big\{ [h, k, \ell] : h \in H, k \in K, \ell \in L \big\} \big\rangle.$$

(i) Prove that if $[H, K, L] = \{1\} = [K, L, H]$, then $[L, H, K] = \{1\}$.

(ii) (**Three Subgroups Lemma**) If $N \lhd G$ and $[H,K,L][K,L,H] \subseteq N$, prove that
$$[L,H,K] \subseteq N.$$

(iii) Prove that if $G$ is a group with $G = G'$, then $G/Z(G)$ is centerless.
**Hint.** If $\pi\colon G \to G/Z(G)$ is the natural map, define $\zeta^2(G) = \pi^{-1}(Z(G/Z(G)))$. Use the Three Subgroups Lemma with $L = \zeta^2(G)$ and $H = K = G$.

(iv) Prove, for all $i, j$, that $[\gamma_i(G), \gamma_j(G)] \subseteq \gamma_{i+j}(G)$.

## C-1.4. Projective Unimodular Groups

The Jordan–Hölder Theorem shows that simple groups can be viewed as building blocks of finite groups. As a practical matter, we can often use this fact to reduce a problem about finite groups to a problem about finite simple groups. Now the only simple groups we have seen so far are cyclic groups of prime order and the alternating groups $A_n$ for $n \geq 5$. We will display more simple groups in this section.

### ▪ General Linear Group $\mathrm{GL}(n,k)$

Recall that if $k$ is a field and $V$ is an $n$-dimensional vector space over $k$, then the ***general linear group*** $\mathrm{GL}(V)$ is the group of all nonsingular linear transformations $T\colon V \to V$ with composition as its binary operation. Given a basis $X = v_1, \ldots, v_n$ of $V$, we can assign an $n \times n$ matrix $A = [\alpha_{ij}]$ to $T$, namely, the matrix whose $j$th column $\alpha_{1j}, \ldots, \alpha_{nj}$ is the coordinate list of $T(v_j)$:
$$T(v_j) = \sum_{i=1}^{n} \alpha_{ij} v_i = \alpha_{1j} v_1 + \cdots + \alpha_{nj} v_n.$$
(In Part 1, we showed the dependence of $A$ on $T$ and $X$ by denoting $A$ by ${}_X[T]_X$.) The function $T \mapsto {}_X[T]_X = A$ is a group isomorphism $\mathrm{GL}(V) \to \mathrm{GL}(n,k)$. When $k = \mathbb{F}_q$ is a finite field with $q$ elements, we often denote $\mathrm{GL}(n,k)$ by $\mathrm{GL}(n,q)$.

Can we display a composition series for the finite group $\mathrm{GL}(n,q)$? We do know two normal subgroups of it: the ***special linear group*** $\mathrm{SL}(n,q)$:
$$\mathrm{SL}(n,q) = \{\text{all } \textbf{\textit{unimodular}} \text{ matrices } A\} = \{\text{all } n \times n \ A : \det(A) = 1\};$$
the center
$$Z(n,q) = \{\text{all } n \times n \ \textbf{\textit{scalar}} \text{ matrices } \alpha I : \alpha \in \mathbb{F}_q^{\#}\},$$
where $\mathbb{F}_q^{\#}$ is the multiplicative group of nonzero elements in $\mathbb{F}_q$ and $I$ is the $n \times n$ identity matrix. If
$$\mathrm{SZ}(n,q) = Z(n,q) \cap \mathrm{SL}(n,q),$$
then a normal series of $\mathrm{GL}(n,q)$ is
$$\mathrm{GL}(n,q) \supseteq \mathrm{SL}(n,q) \supseteq \mathrm{SZ}(n,q) \supseteq \{I\}.$$
Now $\det\colon \mathrm{GL}(n,q) \to \mathbb{F}_q^{\#}$ is easily seen to be a homomorphism. Given a composition series of the cyclic group $\mathbb{F}_q^{\#}$, the Correspondence Theorem interpolates normal subgroups between $\mathrm{GL}(n,q)$ and $\mathrm{SL}(n,q)$. The factor group
$$\mathrm{PSL}(n,q) = \mathrm{SL}(n,q)/\,\mathrm{SZ}(n,q)$$

is called the ***projective unimodular group***; more generally, for any vector space $V$, define

$$\mathrm{PSL}(V) = \mathrm{SL}(V)/\mathrm{SZ}(V).$$

We are going to see that almost all $\mathrm{PSL}(V)$ are simple, which will show, in particular, that we can compute a composition series for $\mathrm{GL}(n,q)$.

The next result begins by recalling Corollary A-7.41 in Part 1: the center of the general linear group consists of the scalar matrices.

**Proposition C-1.70.** *The center of* $\mathrm{SL}(n,q)$ *is* $\mathrm{SZ}(n,q)$.

**Proof.** We first show that the center of $\mathrm{GL}(n,k)$, where $k$ is a field, is the subgroup of the scalar matrices. Every scalar matrix lies in the center; for the reverse inclusion, let $T\colon k^n \to k^n$ not be scalar. Then there is $v \in k^n$ with $v, Tv$ linearly independent. Extend $v, Tv$ to a basis $X = v, Tv, u_3, \ldots, u_n$ of $k^n$. It is easy to see that $X' = v, v+Tv, u_3, \ldots, u_n$ is also a basis. Define $S\colon k^n \to k^n$ by $S(Tv) = v+Tv$ while $S$ fixes $v, u_3, \ldots, u_n$. Note that $S$ is nonsingular, for $X$ and $X'$ are both bases. Now $S$ and $T$ do not commute, for $TS(v) = Tv$ while $ST(v) = v + Tv$. Hence, $T \notin Z(n,k)$.

Clearly, $\mathrm{SZ}(n,q)$ is contained in the center of $\mathrm{SL}(n,q)$. For the reverse inclusion, note that the matrix of $S$ with respect to the basis $X$ is lower triangular with only 1's on its diagonal (its northwest $2 \times 2$ block is $\left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$), and so $\det(S) = 1$. Thus, for every nonscalar matrix, there is a unimodular matrix not commuting with it. Hence, every matrix in the center of SL is also scalar. Thus, the center of $\mathrm{SL}(n,q)$ is $\mathrm{SZ}(n,q)$. •

**Remark.** We have just seen that the center of SL is $\mathrm{SZ} = \mathrm{SL} \cap Z(\mathrm{GL})$, but it is not true in general that $N \subseteq G$ implies $Z(N) = N \cap Z(G)$ (even when $N$ is normal). We always have $N \cap Z(G) \subseteq Z(N)$, but this inclusion may be strict. For example, if $G = S_3$ and $N = A_3 \cong \mathbb{Z}_3$, then $Z(A_3) = A_3$ while $A_3 \cap Z(S_3) = \{1\}$. ◀

Here are the orders of these groups.

**Theorem C-1.71.** *Let* $q = p^m$.

(i) $|\mathrm{GL}(n,q)| = (q^n - 1)(q^n - q)(q^n - q^2)\cdots(q^n - q^{n-1})$.

(ii) $|\mathrm{SL}(n,q)| = (q^n - 1)(q^n - q)(q^n - q^2)\cdots(q^n - q^{n-1})/(q-1)$.

(iii) *If* $d = \gcd(n, q-1)$, *then*

$$|\mathrm{SZ}(n,q)| = d$$

*and*

$$|\mathrm{PSL}(n,q)| = (q^n - 1)(q^n - q)\cdots(q^n - q^{n-1})/d(q-1).$$

**Proof.**

(i) Theorem C-1.45.

(ii) There is an exact sequence

$$\{I\} \to \mathrm{SL}(n,q) \to \mathrm{GL}(n,q) \xrightarrow{\det} \mathbb{F}_q^{\#} \to \{1\},$$

where $\mathbb{F}_q^{\#}$ is the multiplicative group of nonzero elements of the field $\mathbb{F}_q$. Hence, $|\,\mathrm{GL}\,|/|\,\mathrm{SL}\,| = q-1$, and

$$|\,\mathrm{SL}(n,q)| = |\,\mathrm{GL}(n,q)|/(q-1).$$

(iii) Every $A$ in $\mathrm{SZ}(n,q)$, being scalar, has the form $\alpha I$, so that $\det(A) = 1 = \alpha^n$. We claim that if $\alpha$ is any nonzero element in $k$, then $\alpha^n = 1$ if and only if $\alpha^d = 1$. Since $d \mid n$, we see that $\alpha^d = 1$ implies $\alpha^n = 1$. Conversely, there are integers $r$ and $s$ with $d = rn + s(q-1)$. Thus,

$$\alpha^d = \alpha^{rn+s(q-1)} = \alpha^{rn}\alpha^{s(q-1)} = \alpha^{rn},$$

because $\alpha^{q-1} = 1$. Hence, $\alpha^n = 1$ implies $\alpha^d = 1$. It follows that $\mathrm{SZ}(n,q) = \{\alpha \in \mathbb{F}_q : \alpha^n = 1\} = \{\alpha \in \mathbb{F}_q : \alpha^d = 1\}$. Thus, if $\beta$ is a generator of $\mathbb{F}_q^{\#}$ (which is a cyclic group), then $\mathrm{SZ}(n,q) \cong \langle \beta^{n/d} \rangle$. Therefore, $|\,\mathrm{SZ}(n,q)| = d$.

The last equality follows from (ii) and $|\,\mathrm{PSL}\,| = |\,\mathrm{SL}\,|/|\,\mathrm{SZ}\,|$. •

## ▪ Simplicity of $\mathrm{PSL}(2,q)$

In this section we focus on the groups $\mathrm{PSL}(2,q)$, proving their simplicity for (almost all) $\mathbb{F}_q$. In the next section, we will prove simplicity of $\mathrm{PSL}(n,q)$ for all $n \geq 3$; later in this chapter we will give a second proof using *multiple transitivity*.

We begin by computing the order of $\mathrm{PSL}(2,q)$.

**Lemma C-1.72.** $|\,\mathrm{SL}(2,q)| = (q+1)q(q-1)$.

**Proof.** By Theorem C-1.71(iii), $|\,\mathrm{SL}(2,q)| = (q^2-1)(q^2-q)/d(q-1)$. But $q^2-1 = (q+1)(q-1)$. •

**Theorem C-1.73.**

$$|\mathrm{PSL}(2,q)| = \begin{cases} \frac{1}{2}(q+1)q(q-1) & \text{if } q = p^m \text{ and } p \text{ is an odd prime,} \\ (q+1)q(q-1) & \text{if } q = 2^m. \end{cases}$$

**Proof.** By Lemma C-1.72 we have $|\,\mathrm{SL}(2,q)| = (q+1)q(q-1)$. Since $\mathrm{PSL} = \mathrm{SL}\,/\,\mathrm{SZ}$, we have $|\,\mathrm{PSL}(2,q)| = (q+1)q(q-1)/d$.

Now $\mathbb{F}_q^{\times}$ is a cyclic group of order $q-1$, by Theorem A-3.59 in Part 1. If $q$ is odd, then $q-1$ is even, and the cyclic group $\mathbb{F}_q^{\times}$ has a unique subgroup of order 2.

For any field $k$, if $\gamma \in k$ satisfies $\gamma^2 = 1$, then $\gamma = \pm 1$. In particular, if $k = \mathbb{F}_q$, where $q$ is a power of 2, then $\mathbb{F}_q$ has characteristic 2 and $\gamma^2 = 1$ implies $\gamma = 1$. Hence, $\mathrm{SZ}(2,q) = \{I\}$, and so $\mathrm{PSL}(2,2^m) = \mathrm{SL}(2,2^m)$.

Therefore, $|SZ(2,q)| = 2$ if $q$ is a power of an odd prime, and $|SZ(2,q)| = 1$ if $q$ is a power of 2. •

Here are some matrices that play the same role for $n \times n$ linear groups that 3-cycles play for the alternating groups (see Exercise C-1.14).

For the moment, *transvections* (defined below) are matrices; later, we will consider related linear transformations which are also called transvections.

**Definition.** An $n \times n$ ***transvection*** over a field $k$ is an $n \times n$ matrix of the form

$$B_{ij}(\lambda) = I + L_{ij}(\lambda),$$

where $i \neq j$, $\lambda \in k$ is nonzero, $I$ is the identity matrix, and $L_{ij}(\lambda)$ is the matrix having $i, j$ entry $\lambda$ and all other entries 0.

In particular, a $2 \times 2$ ***transvection*** is a matrix of the form

$$B_{12}(\lambda) = \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad B_{21}(\lambda) = \begin{bmatrix} 1 & 0 \\ \lambda & 1 \end{bmatrix}.$$

Note that $\det(B_{ij}(\lambda)) = 1$; that is, transvections are unimodular. Note also that $B_{ij}(\lambda)^{-1} = B_{ji}(-\lambda)$, so that the inverse of a transvection is also a transvection.

We have seen transvections before: $B_{ij}(\lambda)$ is just an elementary matrix of type II which adds $\lambda$ ROW $j$ to ROW $i$. The matrix of the linear transformation $S$ in the proof of Proposition C-1.70 is a transvection.

**Lemma C-1.74.** *Let $k$ be a field. If $A = [\alpha_{ij}] \in \mathrm{GL}(n, q)$ and $\det(A) = \delta$, then*

$$A = UD,$$

*where $U$ is a product of transvections and $D(\delta) = \mathrm{diag}\{1, \ldots, 1, \delta\}$.*

**Proof.** The proof is essentially Gaussian elimination. We show, by induction on $t \leq n - 1$, that an $n \times n$ matrix $A = [\alpha_{ij}]$ can be transformed by a sequence of elementary row operations of type II (which add a multiple of one row to another row) into a matrix of the form

$$A_t = \begin{bmatrix} I_t & * \\ 0 & C \end{bmatrix},$$

where $I_t$ is the $t \times t$ identity matrix and $C = [\gamma_{rs}]$ is an $(n - t) \times (n - t)$ matrix.

For the base step, note that the first column of $A$ is not 0, for $A$ is nonsingular. Adding some row to ROW 2 if necessary, we may assume that $\alpha_{21} \neq 0$. Add $\alpha_{21}^{-1}(1 - \alpha_{11})$ ROW 2 to ROW 1 to get 1 in the upper left corner (i.e., in the $(1, 1)$ position), and now add suitable multiples of ROW 1 to the lower rows so that all entries in the first column below ROW 1 are 0.

For the inductive step, we may assume that $A$ has been transformed into $A_t$ displayed above. Note that the matrix $C$ is nonsingular, for $0 \neq \det(A) = \det(I_t)\det(C) = \det(C)$. If $C$ has at least two rows, we may further assume, as in the base step, that its upper left corner $\gamma_{t+1,t+1} = 1$ (this involves only ROW $(t + 1)$ and rows below it, so that the upper rows 1 through $t$ are not disturbed). Thus, adding suitable multiples of ROW $(t + 1)$ to the rows beneath it yields a matrix $A_{t+1}$.

We may now assume that $A$ has been transformed into

$$A_{n-1} = \begin{bmatrix} I_{n-1} & * \\ 0 & \delta \end{bmatrix},$$

where $\delta \in k$ is nonzero. Finally, add multiples of the bottom row to higher rows so that the higher entries in the last column are 0; that is, we have obtained $D(\delta)$.

In terms of matrix multiplication, we have shown that there is a matrix $P$ which is a product of elementary matrices of type II, that is, $P$ is a product of transvections, such that $PA = D(\delta)$. Therefore, $A = P^{-1}D(\delta)$ is the desired factorization, for $U = P^{-1}$ is a product of transvections (because the inverse of a transvection is also a transvection). •

**Corollary C-1.75.** $\mathrm{SL}(n, q)$ *is generated by transvections.*

**Proof.** If $A \in \mathrm{SL}(n, k)$, then Lemma C-1.74 gives a factorization $A = UD$, where $U$ is a product of transvections and $D = D(\delta)$, where $\delta = \det(A)$. Since $A \in \mathrm{SL}(n, k)$, we have $\delta = \det(A) = 1$, and so $A = U$. •

We now focus on $2 \times 2$ matrices.

**Lemma C-1.76.** *Let $N$ be a normal subgroup of $\mathrm{SL}(2, q)$. If $N$ contains a transvection $B_{12}(\lambda)$ or $B_{21}(\lambda)$, then $N = \mathrm{SL}(2, q)$.*

**Proof.** Note first that if $U = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$, then $\det(U) = 1$ and $U \in \mathrm{SL}(2, q)$; since $N$ is a normal subgroup, $UB_{12}(\lambda)U^{-1}$ also lies in $N$. But $UB_{12}(\lambda)U^{-1} = B_{21}(-\lambda)$, from which it follows that $N$ contains a transvection of the form $B_{12}(\lambda)$ if and only if it contains a transvection of the form $B_{21}(-\lambda)$. Since SL is generated by the transvections, it suffices to show that every transvection $B_{12}(\lambda)$ lies in $N$.

The following conjugate of $B_{12}(\lambda)$ lies in $N$ (because $N$ is normal):

$$\begin{bmatrix} \alpha & \beta \\ 0 & \alpha^{-1} \end{bmatrix} \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha^{-1} & -\beta \\ 0 & \alpha \end{bmatrix} = \begin{bmatrix} 1 & \lambda\alpha^2 \\ 0 & 1 \end{bmatrix} = B_{12}(\lambda\alpha^2).$$

Define

$$G = \{0\} \cup \{\mu \in \mathbb{F}_q : \mu \text{ is nonzero and } B_{12}(\mu) \in N\}.$$

We have just shown that $\lambda\alpha^2 \in G$ for all $\alpha \in \mathbb{F}_q$. It is easy to check that $G$ is a subgroup of the additive group of $\mathbb{F}_q$ and, hence, it contains all the elements of the form $u = \lambda(\alpha^2 - \beta^2)$, where $\alpha, \beta \in \mathbb{F}_q$. We claim that $G = \mathbb{F}_q$, which will complete the proof.

If $q$ is odd, then each $\tau \in \mathbb{F}_q$ is a difference of squares:

$$\tau = (\tfrac{\tau+1}{2})^2 - (\tfrac{\tau-1}{2})^2.$$

In particular, if $\mu \in \mathbb{F}_q$, then there are $\alpha, \beta \in \mathbb{F}_q$ with $\lambda^{-1}\mu = \alpha^2 - \beta^2$. Hence, $\mu = \lambda(\alpha^2 - \beta^2) \in G$ and $G = \mathbb{F}_q$.

If $q$ is even, then the function $\mu \mapsto \mu^2$ is an injection $\mathbb{F}_q \to \mathbb{F}_q$ (for if $\mu^2 = \sigma^2$, then $0 = \mu^2 - \sigma^2 = (\mu - \sigma)^2$, and $\mu = \sigma$). The Pigeonhole Principle says that this function is surjective, and so every element $\mu$ has a square root in $\mathbb{F}_q$. In particular, there is $\alpha \in \mathbb{F}_q$ with $\lambda^{-1}\mu = \alpha^2$, and $\mu = \lambda\alpha^2 \in G$. •

We need one more technical lemma before giving the main result.

**Lemma C-1.77.** *Let $N$ be a normal subgroup of $\mathrm{SL}(2,q)$. If $A \in N$ is similar to $R = \begin{bmatrix} \alpha & \beta \\ \psi & \omega \end{bmatrix} \in \mathrm{GL}(2,q)$, then there is a nonzero $\tau \in \mathbb{F}_q$ with $N$ containing*

$$\begin{bmatrix} \alpha & \tau^{-1}\beta \\ \tau\psi & \omega \end{bmatrix}.$$

**Proof.** By hypothesis, there is a matrix $P \in \mathrm{GL}(2,q)$ with $R = PAP^{-1}$. By Lemma C-1.74, there is a matrix $U \in \mathrm{SL}$ and a diagonal matrix $D = \mathrm{diag}\{1, \tau\}$ with $P^{-1} = UD$. Therefore, $A = UDRD^{-1}U^{-1}$; since $N \lhd \mathrm{SL}$, we have $DRD^{-1} = U^{-1}AU \in N$. But

$$DRD^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \tau \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \psi & \omega \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \tau^{-1} \end{bmatrix} = \begin{bmatrix} \alpha & \tau^{-1}\beta \\ \tau\psi & \omega \end{bmatrix}. \quad \bullet$$

The next theorem was proved by Jordan in 1870 for $q$ prime. In 1893, after Cole had discovered a simple group of order 504, Moore recognized Cole's group as $\mathrm{PSL}(2,8)$, and he then proved the simplicity of $\mathrm{PSL}(2,q)$ for all prime powers $q \geq 4$.

We are going to use Corollary A-7.38 in Part 1: two $2 \times 2$ matrices $A$ and $B$ over a field $k$ are similar (that is, there exists a nonsingular matrix $P$ with $B = PAP^{-1}$) if and only if they both arise from a single linear transformation $\varphi\colon k^2 \to k^2$ relative to two choices of bases of $k^2$. Of course, these statements are true if we replace 2 by $n$. Recall that two nonsingular $n \times n$ matrices $A$ and $B$ are similar if and only if they are conjugate elements in the group $\mathrm{GL}(n,k)$.

**Theorem C-1.78 (Jordan–Moore).** *The groups $\mathrm{PSL}(2,q)$ are simple for all prime powers $q \geq 4$.*

**Remark.** By Theorem C-1.73, $|\mathrm{PSL}(2,2)| = 6$ and $|\mathrm{PSL}(2,3)| = 12$, so that neither of these groups is simple (see Exercise C-1.77 on page 58). ◀

**Proof.** It suffices to prove that a normal subgroup $N$ of $\mathrm{SL}(2,q)$ that contains a matrix not in the center $\mathrm{SZ}(2,q)$ must be all of $\mathrm{SL}(2,q)$.

Suppose that $N$ contains a matrix $A = \begin{bmatrix} \alpha & 0 \\ \beta & \alpha^{-1} \end{bmatrix}$, where $\alpha \neq \pm 1$; that is, $\alpha^2 \neq 1$. If $B = B_{21}(1)$, then $N$ contains the commutator $BAB^{-1}A^{-1} = B_{21}(1-\alpha^{-2})$, which is a transvection because $1 - \alpha^{-2} \neq 0$. Therefore, $N = \mathrm{SL}(2,q)$, by Lemma C-1.76.

To complete the proof, we need only show that $N$ contains a matrix whose top row is $[\alpha\ 0]$, where $\alpha \neq \pm 1$. By hypothesis, there is some matrix $M \in N$ that is not a scalar matrix. Let $J\colon (\mathbb{F}_q)^2 \to (\mathbb{F}_q)^2$ be the linear transformation given by $J(v) = Mv$, where $v$ is a $2 \times 1$ column vector. If $J(v) = \gamma_v v$ for all $v$, where $\gamma_v \in \mathbb{F}_q$, then the matrix of $J$ relative to any basis of $(\mathbb{F}_q)^2$ is a diagonal matrix. In this case, $M$ is similar to a diagonal matrix $D = \mathrm{diag}\{\alpha, \beta\}$, and Lemma C-1.77 says that $D \in N$. Since $M \notin \mathrm{SZ}(2,q)$, we must have $\alpha \neq \beta$. But $\alpha\beta = \det(M) = 1$, and so $\alpha \neq \pm 1$. Therefore, $D$ is a matrix in $N$ of the desired form.

In the remaining case, there is a vector $v$ with $J(v)$ not a scalar multiple of $v$. We saw, in Example A-7.32 in Part 1, that $M$ is similar to a matrix of the form

$\left[\begin{smallmatrix} 0 & \alpha \\ 1 & \beta \end{smallmatrix}\right]$, and we must have $\alpha = -1$ because $M$ has determinant 1. Lemma C-1.77 now says that there is some nonzero $\tau \in \mathbb{F}_q$ with

$$D = \begin{bmatrix} 0 & -\tau^{-1} \\ \tau & \beta \end{bmatrix} \in N.$$

If $T = \mathrm{diag}\{\alpha, \alpha^{-1}\}$ (where $\alpha$ will be chosen in a moment), then the commutator

$$V = (TDT^{-1})D^{-1} = \begin{bmatrix} \alpha^2 & 0 \\ \tau\beta(\alpha^{-2} - 1) & \alpha^{-2} \end{bmatrix} \in N.$$

We are done if $\alpha^2 \neq \pm 1$, that is, if there is some nonzero $\alpha \in \mathbb{F}_q$ with $\alpha^4 \neq 1$. If $q > 5$, then such an element $\alpha$ exists, for the polynomial $x^4 - 1$ has at most four roots in a field. If $q = 4$, then every $\alpha \in \mathbb{F}_4$ is a root of the equation $x^4 - x$; that is, $\alpha^4 = \alpha$. Hence, if $\alpha \neq 1$, then $\alpha^4 \neq 1$.

Only the case $q = 5$ remains. The entry $\beta$ in $D$ shows up in the lower left corner $\tau\beta(\alpha^{-2} - 1)$ of the commutator $V$. There are two subcases depending on whether $\beta \neq 0$ or $\beta = 0$. In the first subcase, choose $\alpha = 2$ so that $\alpha^{-2} = 4 = \alpha^2$ and $\mu = (4-1)\tau\beta = 3\tau\beta \neq 0$. Now $N$ contains $V^2 = B_{21}(-2\mu)$, which is a transvection because $-2\mu = -6\tau\beta = 4\tau\beta \neq 0$. Finally, if $\beta = 0$, then $D$ has the form

$$D = \begin{bmatrix} 0 & -\tau^{-1} \\ \tau & 0 \end{bmatrix}.$$

Conjugating $D$ by $B_{12}(\psi)$ for $\psi \in \mathbb{F}_5$ gives a matrix $B_{12}(\psi)DB_{12}(-\psi) \in N$ (remember that $B_{12}(\psi)^{-1} = B_{12}(-\psi)$) whose top row is

$$[\tau\psi \quad -\tau\psi^2 - \tau^{-1}].$$

If we choose $\psi = 2\tau^{-1}$, then the top row is $[2\ 0]$, and the proof is complete. •

It is true that $\mathrm{PSL}(2, k)$ is a simple group for every infinite field $k$, but our proof for $k = \mathbb{F}_q$ (in Lemma C-1.76) imposes a condition on $k$.

**Definition.** A field $k$ is ***perfect*** if its characteristic is 0 or it has characteristic $p$ and every $a \in k$ has a $p$th root in $k$.

If a field $k$ has characteristic $p$, then the (Frobenius) map $\mathrm{Fr}\colon k \to k$ sending $b \mapsto b^p$ is always injective. If $k$ is finite, then the Pigeonhole Principle says that Fr is an isomorphism, so that every finite field is perfect. An example of an infinite field which is not perfect is $k = \mathbb{F}_p(x)$, for the indeterminate $x$ does not have a $p$th root in $k$.

**Corollary C-1.79.** *If $k$ is a perfect field, then* $\mathrm{PSL}(2, k)$ *is simple*; *in particular,* $\mathrm{PSL}(2, k)$ *is an infinite simple group if $k$ is a field of characteristic* 0.

**Proof.** In the various lemmas leading up to the proof of simplicity of $\mathrm{PSL}(2, q)$, the finiteness of $\mathbb{F}_q$ was used either in computing orders of groups or in the proof of Lemma C-1.76. In that proof, we used the fact that every element in $\mathbb{F}_{2^m}$ has a square root; we did not assume the existence of roots otherwise. •

Here are the first few orders of these simple groups:

$$\begin{aligned} |\operatorname{PSL}(2,4)| &= 60, \\ |\operatorname{PSL}(2,5)| &= 60, \\ |\operatorname{PSL}(2,7)| &= 168, \\ |\operatorname{PSL}(2,8)| &= 504, \\ |\operatorname{PSL}(2,9)| &= 360, \\ |\operatorname{PSL}(2,11)| &= 660. \end{aligned}$$

It is known that these numbers are the only orders of nonabelian simple groups which are under 1000. Some of these, namely, 60 and 360, coincide with orders of alternating groups. The next proposition shows that $\operatorname{PSL}(2,4) \cong A_5 \cong \operatorname{PSL}(2,5)$, and Exercise C-1.78 on page 58 shows that $\operatorname{PSL}(2,9) \cong A_6$. We shall see in the next section that $A_8$ and $\operatorname{PSL}(3,4)$ are nonisomorphic simple groups of the same order, namely, $\frac{1}{2}8! = 20,160$.

**Proposition C-1.80.** *If $G$ is a simple group of order 60, then $G \cong A_5$.*

**Proof.** It suffices to show that $G$ has a subgroup $H$ of index 5, for then Theorem C-1.2, the representation on the cosets of $H$, gives a homomorphism $\varphi : G \to S_5$ with $\ker \varphi \subseteq H$. As $G$ is simple, the proper normal subgroup $\ker \varphi$ is equal to $\{1\}$, and so $G$ is isomorphic to a subgroup of $S_5$ of order 60. By Exercise C-1.17(ii) on page 15, $A_5$ is the only subgroup of $S_5$ of order 60, and so $G \cong A_5$.

Suppose that $P$ and $Q$ are Sylow 2-subgroups of $G$ with $P \cap Q \neq \{1\}$; choose $x \in P \cap Q$ with $x \neq 1$. Now $P$ has order 4, hence is abelian, and so $4 \mid |C_G(x)|$, by Lagrange's Theorem. Indeed, since both $P$ and $Q$ are abelian, the subset $P \cup Q$ is contained in $C_G(x)$, so that $|C_G(x)| \geq |P \cup Q| > 4$. Therefore, $|C_G(x)|$ is a proper multiple of 4 which is also a divisor of 60: either $|C_G(x)| = 12$, $|C_G(x)| = 20$, or $|C_G(x)| = 60$. The second case cannot occur lest $C_G(x)$ have index 3, and representing $G$ on its cosets would show that $G$ is isomorphic to a subgroup of $S_3$; the third case cannot occur lest $x \in Z(G) = \{1\}$. Therefore, $C_G(x)$ is a subgroup of $G$ of index 5, and we are done in this case. We may now assume that every two Sylow 2-subgroups of $G$ intersect in $\{1\}$.

A Sylow 2-subgroup $P$ of $G$ has $r = [G : N_G(P)]$ conjugates, where $r = 3, 5$, or 15. Now $r \neq 3$ ($G$ has no subgroup of index 3). We show that $r = 15$ is not possible by counting elements. Each Sylow 2-subgroup contains three nonidentity elements. Since any two Sylow 2-subgroups intersect trivially (as we have seen above), their union contains $15 \times 3 = 45$ nonidentity elements. Now a Sylow 5-subgroup of $G$ must have six conjugates (the number $r_5$ of them is a divisor of 60 satisfying $r_5 \equiv 1 \bmod 5$). But Sylow 5-subgroups are cyclic of order 5, so that the intersection of any pair of them is $\{1\}$, and so the union of them contains $6 \times 4 = 24$ nonidentity elements. We have exceeded the number of elements in $G$, and so this case cannot occur. •

**Corollary C-1.81.** $\operatorname{PSL}(2,4) \cong A_5 \cong \operatorname{PSL}(2,5)$.

**Proof.** All three groups are simple and have order 60. •

## Exercises

**C-1.76.** Give a composition series for $\mathrm{GL}(2,5)$ and list its factor groups.

∗ **C-1.77.** (i) Prove that $\mathrm{PSL}(2,2) \cong S_3$.

(ii) Prove that $\mathrm{PSL}(2,3) \cong A_4$.

∗ **C-1.78.** Prove that any simple group $G$ of order 360 is isomorphic to $A_6$. Conclude that $\mathrm{PSL}(2,9) \cong A_6$.

**Hint.** Let $N_G(P)$ be the normalizer of a Sylow 5-subgroup $P$. Prove that there is an element $\alpha$ of order 3 such that $A = \langle N_G(P), \alpha \rangle$ has order 120. Since $N_G(P) \cong D_{10}$, by Exercise C-1.43(iii) on page 33, we have $|A| = 30$, 60, or 360.

**C-1.79.** (i) Prove that $\mathrm{SL}(2,5)$ is not solvable.

(ii) Show that a Sylow 2-subgroup of $\mathrm{SL}(2,5)$ is isomorphic to $\mathbf{Q}$, the quaternion group of order 8.

(iii) Prove that the Sylow $p$-subgroups of $\mathrm{SL}(2,5)$ are cyclic if $p$ is an odd prime. Conclude, for every prime divisor $p$ of $|\mathrm{SL}(2,5)|$, that all the Sylow $p$-subgroups of $\mathrm{SL}(2,5)$ have a unique subgroup of order $p$.

**C-1.80.** Prove that $\mathrm{GL}(2,7)$ is not solvable.

∗ **C-1.81.** (i) Prove that $\mathrm{SL}(2,q)$ is the commutator subgroup of $\mathrm{GL}(2,q)$ for all prime powers $q \geq 4$.

(ii) What is the commutator subgroup of $\mathrm{GL}(2,q)$ when $q = 2$ and when $q = 3$?

**C-1.82.** Let $\pi$ be a primitive element of $\mathbb{F}_8$.

(i) What is the order of $\left[\begin{smallmatrix} \pi & 0 \\ 1 & \pi \end{smallmatrix}\right]$ considered as an element of $\mathrm{GL}(2,8)$?

(ii) What is the order of $\left[\begin{smallmatrix} \pi & 0 & 0 \\ 1 & \pi & 0 \\ 0 & 1 & \pi \end{smallmatrix}\right]$ considered as an element of $\mathrm{GL}(3,8)$?

**Hint.** Show that if $N = \left[\begin{smallmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right]$, then $N^2 = \left[\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{smallmatrix}\right]$ and $N^3 = 0$, and use the Binomial Theorem to show that if $A = \left[\begin{smallmatrix} \pi & 0 & 0 \\ 1 & \pi & 0 \\ 0 & 1 & \pi \end{smallmatrix}\right]$, then $A^m = \pi^m I + m\pi^{m-1}N + \binom{m}{2}\pi^{m-2}N^2$.

### ▪ Simplicity of $\mathrm{PSL}(n,q)$

In 1870, Jordan proved that $\mathrm{PSL}(n,p)$ is simple for all $n \geq 3$ and all primes $p$. In 1897, Dickson proved that $\mathrm{PSL}(n,q)$ is simple for all $n \geq 3$ and all prime powers $q = p^m$. In contrast to our proof of the simplicity of $\mathrm{PSL}(2,q)$, the proof here of the simplicity of $\mathrm{PSL}(n,q)$ for all $n \geq 3$ follows that of E. Artin, using more of the geometry present in vector spaces.

**Lemma C-1.82.** *Let $V$ be an $n$-dimensional vector space over a field $k$, let $H$ be a hyperplane in $V$ (that is, $H$ is an $(n-1)$-dimensional subspace), and let $T\colon V \to V$ be a nonsingular linear transformation fixing $H$ pointwise. Assume that $w \in V$ and $w \notin H$.*

(i) *Every $v \in V$ has a unique expression of the form $v = \lambda w + h$, where $\lambda \in k$ and $h \in H$. In particular,*

$$T(w) = \mu_0 w + h_0,$$

*where $\mu_0 \in k$ and $h_0 \in H$. Moreover, for every $v \in V$, there is $h' \in H$ with*

$$T(v) = \mu_0 v + h'.$$

(ii) *The scalar $\mu_0$ depends only on $T$ and not on $w$. Denote this scalar by*

$$\mu_0 = \mu(T).$$

(iii) *If $\mu_0 = 1$ and $v = \lambda w + h$ as in* (i)*, then*

$$T(v) = v + \lambda h_0.$$

*Moreover, if $T \neq 1_V$, then every eigenvector of $T$ lies in $H$.*

**Proof.**

(i) Since $H$ is a hyperplane and $w \notin H$, the subspace $\langle w \rangle = \{\lambda w : \lambda \in k\}$ is a complement of $H$:

$$V = \langle w \rangle \oplus H.$$

Thus, each $v \in V$ has a unique expression as stated.

If $v \in V$, then $v = \lambda w + h$. Since $T$ fixes $H$ pointwise,

$$\begin{aligned} T(v) &= \lambda T(w) + h \\ &= \lambda(\mu_0 w + h_0) + h \\ &= \mu_0 \lambda w + \lambda h_0 + \mu_0 h - \mu_0 h + h \\ &= \mu_0(\lambda w + h) + (\lambda h_0 - \mu_0 h + h) \\ &= \mu_0 v + h', \end{aligned}$$

where $h' = \lambda h_0 - \mu_0 h + h \in H$. Note that if $\mu_0 = 1$, then $h' = \lambda h_0$.

(ii) Let us see that $\mu_0$ does not depend on $w$. Choose $w' \in V$ with $w' \notin H$. As in (i), $T(w') = \mu_0' w' + h_0'$, where $\mu_0' \in k$ and $h_0' \in H$ and, taking $v = w'$, $T(w') = \mu_0 w' + h''$ for some $h'' \in H$. Thus, $(\mu_0' - \mu_0)w' = h'' - h_0' \in \langle w' \rangle \cap H = \{0\}$, so that $\mu_0' = \mu_0$.

(iii) Every nonzero vector in $H$ is an eigenvector of $T$ (for the eigenvalue 1), because $T$ fixes $H$ pointwise; are there any others? By (i), $T(v) = \mu_0 v + h'$, where $h' \in H$. But we noted in the proof of (i) that if $\mu_0 = 1$, then $h' = \lambda h_0$, and so

$$T(v) = v + \lambda h_0. \quad \bullet$$

**Definition.** Let $H$ be a hyperplane in a vector space $V$, and let $T \in \mathrm{GL}(V)$ fix $H$ pointwise. If $\mu(T) \neq 1$, then $T$ is called a ***dilatation***; if $\mu(T) = 1$ and $T \neq 1_V$, then $T$ is called a ***transvection***.

The next result gives matrix versions of dilatations and transvections. It shows, in particular, that if $\dim(V) = n \geq 2$, then the present definition of transvection agrees with the definition used in the preceding section.

**Theorem C-1.83.** *Let $H$ be a hyperplane in a vector space $V$ over a field $k$, and let $T \in \mathrm{GL}(V)$ fix $H$ pointwise.*

(i) *If $T$ is a dilatation, then there is a basis $X$ of $V$ so that the matrix of $T$ relative to $X$ is $D(\mu) = \mathrm{diag}\{1, \ldots, 1, \mu\}$.*

(ii) *If $T$ is a transvection, then there is a basis $X$ of $V$ such that the matrix of $T$ relative to $X$ is $B_{21}(1)$.*

**Proof.** We write $\mu = \mu(T)$ in this proof.

(i) We seek $g \in H$ with $T(w+g) = \mu(w+g)$. Now $T(w) = \mu w + h_0$, so that $\mu w + h_0 + g = \mu w + \mu g$ and $(1-\mu)g = h_0$. But $T$ is a dilatation, so $\mu \neq 1$, and we can solve for $g$. Hence, $w+g$ is a eigenvector of $T$ for $\mu$. If $h_1, \ldots, h_{n-1}$ is a basis of $H$, then $h_1, \ldots, h_{n-1}, w+g$ is a basis of $V$, and the matrix of $T$ relative to this basis is $\mathrm{diag}\{1, \ldots, 1, \mu\}$.

(ii) By Lemma C-1.82(iii), $T$ has no eigenvectors outside of $H$. Since $h_0 \in H$ is nonzero, it is part of a basis of $H$, say, $h_0, h_2, \ldots, h_{n-1}$, and adjoining $w$ gives the basis $w, h_0, h_2, \ldots, h_{n-1}$ of $V$. Relative to this basis, $T = B_{21}(1)$, for $T(w) = w + h_0$, and so $T$ is a transvection. •

Let $E = e_1, \ldots, e_n$ be the standard basis of $k^n$. The linear transformation $T\colon k^n \to k^n$ arising from the matrix transvection $B_{ij}(\lambda)$ is given by

$$T(e_\ell) = \begin{cases} e_\ell & \text{if } \ell \neq i, \\ e_i + \lambda e_j & \text{if } \ell = i. \end{cases}$$

Now $H = \langle e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n \rangle$ is a hyperplane fixed pointwise by $T$. If we define $w = e_i + \lambda e_j$, then $T(w) = w + \lambda e_j$; note that $\lambda e_j \in H$, so that $\mu(T) = 1$ and $T$ is a transvection.

**Lemma C-1.84.** *All (matrix) transvections in $\mathrm{GL}(n,q)$ are conjugate in $\mathrm{GL}(n,q)$.*

**Proof.** It suffices to prove that there is $P \in \mathrm{GL}(n,q)$ with $B_{ij}(\lambda) = PB_{21}(1)P^{-1}$, for then any two matrix transvections are similar to $B_{21}(1)$ and, hence, are similar to each other.

Let $T\colon k^n \to k^n$ be the linear transformation arising from a matrix transvection $B_{ij}(\lambda)$. As usual, if $E$ is the standard basis of $k^n$, then ${}_E[T]_E = B_{ij}(\lambda)$. Now Theorem C-1.83 shows that there is a basis $X$ of $k^n$ with respect to which the matrix of $T$ is $B_{21}(1)$; that is, ${}_X[T]_X = B_{21}(1)$. By Corollary A-7.38 in Part 1, there is a nonsingular matrix $P$, namely $P = {}_E[1_{k^n}]_X$, with

$$B_{ij}(\lambda) = P\,{}_X[T]_X P^{-1} = PB_{21}(1)P^{-1}. \quad \bullet$$

Here is a more geometric description of transvections. Let $H$ be a hyperplane in a vector space $V$ and let $T\colon V \to V$ be a transvection fixing $H$ pointwise. In Lemma C-1.82(iii), we saw, for every $v = \lambda w + h \in V$, that

$$T(v) = v + \lambda h_0.$$

The function $\varphi\colon V \to k$, given by $\varphi(v) = \lambda$, is a linear functional, and we may write

$$T(v) = v + \varphi(v)h_0.$$

Note that $\ker\varphi = H$.

**Notation.** Let $\varphi\colon V \to k$ be a linear functional with $\ker\varphi = H$. If $h_0 \in H$, define $[\varphi, h_0]\colon V \to V$ by

$$[\varphi, h_0]\colon v \mapsto v + \varphi(v)h_0.$$

**Lemma C-1.85.** *Let $V$ be a vector space over $k$.*

(i) $[\varphi, h_0]$ *is a transvection.*

(ii) *Given* $[\varphi, h_0]$ *and* $[\psi, \ell_0]$, *we have*

$$[\varphi, h_0] \circ [\varphi, \ell_0] = [\varphi, h_0 + \ell_0] \quad \textit{and} \quad [\varphi, h_0] \circ [\psi, h_0] = [\varphi + \psi, h_0].$$

(iii) *For all nonzero* $\lambda \in k$,

$$[\lambda\varphi, h_0] = [\varphi, \lambda h_0].$$

(iv) $[\varphi, h_0] = [\psi, \ell_0]$ *if and only if there is a nonzero* $\tau \in k$ *with*

$$\psi = \tau\varphi \quad \textit{and} \quad h_0 = \tau\ell_0.$$

(v) *If* $S \in \mathrm{GL}(V)$, *then*

$$S[\varphi, h_0]S^{-1} = [\varphi S^{-1}, Sh_0].$$

**Proof.** All are routine calculations. •

Lemma C-1.84 says that all transvections (which are necessarily unimodular) are conjugate in $\mathrm{GL}(n, q)$; that is, if $B$ and $B'$ are transvections, then there is some $P \in \mathrm{GL}(n, q)$ with $B' = PBP^{-1}$. The next theorem says that there is such a matrix $P$ having determinant 1.

**Theorem C-1.86.** *All transvections in* $\mathrm{SL}(V)$ *are conjugate in* $\mathrm{SL}(V)$ *if* $\dim(V) \geq 3$.

**Remark.** We have seen that any two transvections are conjugate in GL, but it is easy to see that

$$B_{12}(1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = B_{12}(-1)$$

are not conjugate in $\mathrm{SL}(2, 3)$ (indeed, these transvections are not conjugate in $\mathrm{SL}(2, k)$ for any field $k$ in which $-1$ is not a square). The assumption that $n \geq 3$ is thus essential. ◀

**Proof.** Let $[\varphi, h]$ and $[\psi, \ell]$ be transvections, so that $H = \ker\varphi$ and $L = \ker\psi$ are the hyperplanes fixed by each. Choose $v, u \in V$ with $\varphi(v) = 1 = \psi(u)$ (it follows that $v \notin H$ and $u \notin L$). There are bases $h, h_2, \ldots, h_{n-1}$ and $\ell, \ell_2, \ldots, \ell_{n-1}$ of $H$ and $L$, respectively, and adjoining $v$ and $u$ gives bases $X = v, h, h_2, \ldots, h_{n-1}$ and $Y = u, \ell, \ell_2, \ldots, \ell_{n-1}$ of $V$. If $T\colon V \to V$ takes $X$ to $Y$, then $T \in \mathrm{GL}(V)$ and

$$T(v) = v, \quad T(H) = L, \quad \text{and} \quad T(h) = \ell. \tag{1}$$

Let $\det(T) = \delta$. We now modify $T$ so it will be unimodular. Since $\dim(V) \geq 3$, we may assume, in the basis $X$, that $v, h, h_{n-1}$ are distinct. Define $S$ so that it agrees with $T$ on $X$ except that $S(h_{n-1}) = \delta^{-1}\ell_{n-1}$ (the matrix of $S$ is the matrix of $T$ except that its last column is multiplied by $\delta^{-1}$). Thus, $S$ has the same properties as $T$ while $\det(S) = 1$; that is, $S \in \mathrm{SL}(V)$.

By Lemma C-1.85, $S[\varphi, h]S^{-1} = [\varphi S^{-1}, S(h)] = [\varphi S^{-1}, \ell]$. But $\varphi S^{-1}$ and $\psi$ agree on the basis $Y$ of $V$, so that $\varphi S^{-1} = \psi$. Therefore, $[\varphi, h]$ and $[\psi, \ell]$ are conjugate in $\mathrm{SL}(V)$. •

**Corollary C-1.87.** *Let $N$ be a normal subgroup of $\mathrm{SL}(V)$, where $\dim(V) \geq 3$. If $N$ contains a transvection $U$, then $N = \mathrm{SL}(V)$.*

**Proof.** Corollary C-1.75 allows us to prove this for $\mathrm{SL}(n, k) \cong \mathrm{SL}(V)$. If $A \in \mathrm{SL}(n, k)$, then $A = B_1 \cdots B_m$, where each $B_t$ is a transvection. By hypothesis, for each $t$, there is $P_t \in \mathrm{SL}(n, k)$ with $B_t = P_t U P_t^{-1}$. Since $N \lhd \mathrm{SL}(n, k)$, each $B_t \in N$, and so $A \in N$. •

**Definition.** If $H$ is a hyperplane in a vector space $V$, define

$$\mathrm{Tran}(H) = \{1_V\} \cup \{\text{all transvections fixing } H \text{ pointwise}\}.$$

**Lemma C-1.88.** *Let $H$ be a hyperplane in a vector space $V$ over $k$.*

(i) *There is a linear functional $\varphi$ with $H = \ker\varphi$ so that*

$$\mathrm{Tran}(H) = \{1_V\} \cup \{[\varphi, h] : h \in H \text{ and } h \neq 0\}.$$

(ii) $\mathrm{Tran}(H)$ *is an abelian subgroup of* $\mathrm{SL}(V)$, *and* $\mathrm{Tran}(H) \cong H$.

(iii) *The centralizer* $C_{\mathrm{SL}}(\mathrm{Tran}(H)) = \mathrm{SZ}(V)\mathrm{Tran}(H)$.

**Proof.**

(i) We show first that $\varphi, \psi \in \mathrm{Tran}(H)$ are distinct from $1_V$, i.e., $\ker\varphi = H = \ker\psi$, if and only if there is a nonzero $\alpha \in k$ with $\psi = \alpha\varphi$. It is clear that $\psi = \alpha\varphi$ implies $\ker\varphi = \ker\psi$. Conversely, if $\ker\varphi = H = \ker\psi$, choose $w \in V$ with $w \notin H$. Since both $\varphi(w)$ and $\psi(w)$ are nonzero elements of $k$, there is $\alpha \neq 0$ with $\varphi(w) = \alpha\psi(w)$. Now every $v \in V$ can be written as $v = \lambda w + h$, where $\lambda \in K$ and $h \in H$. Hence,

$$\psi(v) = \psi(\lambda w + h) = \lambda\psi(w) = \lambda\alpha\varphi(w) = \alpha\varphi(v).$$

Now choose $[\varphi, h_0] \in \mathrm{Tran}(H)$. If $[\psi, \ell] \in \mathrm{Tran}(H)$, then $\psi$ fixes $H$ pointwise; by the above, there exists $\alpha \in k$ with $\psi = \alpha\varphi$. Hence, $[\psi, \ell] = [\alpha\varphi, \ell] = [\varphi, \alpha\ell]$.

(ii) By (i), we may assume that there is a linear functional $\varphi$ so that every transvection in $\mathrm{Tran}(H)$ has the form $[\varphi, h]$ for some nonzero $h \in H$. If $[\varphi, h], [\varphi, \ell] \in \mathrm{Tran}(H)$, then $[\varphi, -\ell] \in \mathrm{Tran}(H)$ and $[\varphi, h] \circ [\varphi, \ell] \in \mathrm{Tran}(H) = [\varphi, h + \ell]$. Since $[\varphi, h] \circ [\varphi, \ell] = [\varphi, h + \ell] = [\varphi, \ell + h]$, it follows that $\mathrm{Tran}(H)$ is an abelian subgroup of $\mathrm{SL}(V)$, and it is easy to see that $[\varphi, h] \mapsto h$ is an isomorphism $\mathrm{Tran}(H) \to H$.

(iii) Since $\mathrm{Tran}(H)$ is abelian, we have $\mathrm{SZ}(V)\mathrm{Tran}(H) \subseteq C_{\mathrm{SL}}(\mathrm{Tran}(H))$. For the reverse inclusion, assume that $S \in \mathrm{SL}(V)$ commutes with every $[\varphi, h]$; that is, for all $h \in H$, $S[\varphi, h]S^{-1} = [\varphi, h]$. Now Lemma C-1.85(v) says that $S[\varphi, h]S^{-1} = [\varphi S^{-1}, Sh]$. But part (iv) of that lemma says there is a nonzero $\alpha \in k$ with

$$\varphi S^{-1} = \alpha\varphi \quad \text{and} \quad Sh = \alpha^{-1}h. \tag{2}$$

Since $\alpha S$ fixes $H$ pointwise, it is either a dilatation or a transvection. If it is a transvection, then $\alpha S \in \mathrm{Tran}(H)$, and so $S = \alpha^{-1}(\alpha S) \in \mathrm{SZ}(V)\mathrm{Tran}(H)$. If $\alpha S$ is a dilatation, then it has an eigenvector $w \notin H$: there is $\mu \in k$ with $\alpha S(w) = \mu w$, where $1 \neq \mu = \det(\alpha S) = \alpha^n$ (for $\det(S) = 1$); hence, $Sw = \alpha^{n-1}w$. But $\varphi S^{-1}w = \varphi(\alpha^{-n+1}w) = \alpha^{-n+1}w\varphi(w)$, so that Eq. (2) gives $\varphi(w) = \alpha^n\varphi(w)$. Since $\varphi(w) \neq 0$ (because $w \notin H$), we reach the contradiction $\alpha^n = 1$. •

**Theorem C-1.89 (Jordan–Dickson).** *For every $n \geq 3$ and every field $k$, the group $\mathrm{PSL}(n, k)$ is simple.*

**Proof.** The proof will consider $\mathrm{SL}(V)$, where $\dim(V) = n$ instead of $\mathrm{SL}(n, k)$, that is, linear transformations instead of matrices. We show that if $N$ is a normal subgroup of $\mathrm{SL}(V)$ containing some $A \notin \mathrm{SZ}(V)$, then $N = \mathrm{SL}(V)$; by Corollary C-1.87, it suffices to show that $N$ contains a transvection.

Since $\mathrm{SL}(V)$ is generated by transvections, there exists a transvection $T$ which does not commute with $A$. Hence, the commutator $B = T^{-1}A^{-1}TA \neq I$. Note that $N \lhd \mathrm{SL}(V)$ gives $B \in N$. Thus,

$$B = T^{-1}(A^{-1}TA) = T_1T_2,$$

where both $T_1$ and $T_2$ are transvections. Now each $T_i = [\varphi_i, h_i]$, where $h_i \in H_i = \ker\varphi_i$; that is, for $i = 1, 2$,

$$T_i(v) = v + \varphi_i(v)h_i \quad \text{for all } v \in V.$$

Let $W$ be the subspace $\langle h_1, h_2\rangle$ of $V$, so that $\dim(W) \leq 2$. Since $\dim(V) \geq 3$, there is a hyperplane $L$ of $V$ containing $W$. We claim that $B(L) \subseteq L$. If $\ell \in L$, then

$$\begin{aligned} B(\ell) &= T_1T_2(\ell) = T_2(\ell) + \varphi_1(T_2(\ell))h_1 \\ &= \ell + \varphi_2(\ell)h_2 + \varphi_1(T_2(\ell))h_1 \in L + W \subseteq L. \end{aligned}$$

Next, we claim that $H_1 \cap H_2 \neq \{0\}$. This is surely true if $H_1 = H_2$. Otherwise, if $H_1 \neq H_2$, then $H_1 + H_2 = V$ (hyperplanes are maximal subspaces), so that $\dim(H_1 + H_2) = n$. Since

$$\dim(H_1) + \dim(H_2) = \dim(H_1 + H_2) + \dim(H_1 \cap H_2),$$

we have $\dim(H_1 \cap H_2) = n - 2 \geq 1$. Take $z \in H_1 \cap H_2$ with $z \neq 0$; then

$$B(z) = T_2T_2(z) = z.$$

We may assume that $B$ is not a transvection (or we are done). Therefore, $B \notin \mathrm{Tran}(L)$, which is wholly comprised of transvections. If $B = \alpha S$, where $S \in$

$\mathrm{Tran}(L)$, then $z$ is an eigenvector of $S$ (for $z = B(z) = \alpha S(z)$, and so $S(z) = \alpha^{-1}z$). As eigenvectors of transvections lie in the fixed hyperplane, we have $z \in L$; thus, $\alpha = 1$, giving the contradiction $S = B$. Therefore, $B \notin \mathrm{SZ}(V)\mathrm{Tran}(L) = C_{\mathrm{SL}}(\mathrm{Tran}(L))$; thus, there exists $U \in \mathrm{Tran}(L)$ not commuting with $B$:

$$UBU^{-1}B^{-1} \neq 1;$$

of course $C$, defined by $C = UBU^{-1}B^{-1}$, lies in $N$ (for $N$ is normal). If $\ell \in L$, then

$$C(\ell) = UBU^{-1}B^{-1}(\ell) = UB(B^{-1}(\ell)) = \ell,$$

because $B^{-1}(\ell) \in L$ and $U^{-1} \in \mathrm{Tran}(L)$ fixes $L$ pointwise. As the linear transformation $C$ fixes the hyperplane $L$, it is either a transvection or a dilatation. However, $C$ is not a dilatation because $\det(C) = 1$. Therefore, $C$ is a transvection in $N$. •

We will give another proof of the Jordan–Dickson Theorem in the next section.

The next result gives some interesting information about GL.

**Proposition C-1.90.** *If $V$ is a vector space over $k$, then the commutator subgroup of $\mathrm{GL}(V)$ is $\mathrm{SL}(V)$ unless $V$ is a two-dimensional vector space over $\mathbb{F}_2$.*

**Proof.** Now $\det\colon \mathrm{GL} \to k^{\#}$ has kernel SL; since the multiplicative group $k^{\#}$ is abelian, we have $(\mathrm{GL})' \subseteq \mathrm{SL}$. For the reverse inclusion, let $\pi\colon \mathrm{GL} \to \mathrm{GL}/(\mathrm{GL})'$ be the natural map. By Lemma C-1.84, all transvections are conjugate in GL. It follows that $\pi(T) = \pi(S)$ for all transvections $T$ and $S$; let $\delta$ denote their common value. Now $T = [\varphi, h]$ for some linear functional $\varphi$ and some $h \in H$. If we avoid the exceptional case in the statement, then $H$ contains a nonzero vector $v$ (which might be $h$) with $v + h \neq 0$. By Lemma C-1.85(ii), $[\varphi, h] \circ [\varphi, \ell] = [\varphi, h + \ell]$ (the latter is a transvection because $h + \ell \neq 0$). Applying $\pi$ to this equation gives $\delta^2 = \delta$ in $\mathrm{GL}/(\mathrm{GL})'$, whence $\delta = 1$. Thus, every transvection lies in $\ker \pi = (\mathrm{GL})'$. But SL is generated by the transvections, by Lemma C-1.74, and so $\mathrm{SL} \subseteq (\mathrm{GL})'$. •

If $V$ is a two-dimensional vector space over $\mathbb{F}_2$, then $\mathrm{GL}(V)$ is a genuine exception to the proposition. In this case,

$$\mathrm{GL}(V) = \mathrm{SL}(V) \cong \mathrm{SL}(2, 2) \cong S_3,$$

and $(\mathrm{SL})' = A_3$, a proper subgroup.

The first example of nonisomorphic simple groups having the same order was given by Schottenfels in 1900. Note that the order of $\mathrm{PSL}(3, 4)$ is

$$|\mathrm{PSL}(3, 4)| = (4^3 - 1)(4^3 - 4)(4^3 - 4^2)/3 \cdot 3 = 20,160,$$

for $n = 3$, $q = 4$, and $d = \gcd(3, 4 - 1) = 3$.

**Theorem C-1.91 (Schottenfels).** $\mathrm{PSL}(3, 4)$ *and $A_8$ are nonisomorphic simple groups of order $\frac{1}{2}8! = 20,160$.*

**Proof.** The permutations (1 2)(3 4) and (1 2)(3 4)(5 6)(7 8) lie in $A_8$ (they are even) and they have order 2. Since they have different cycle structure, they are not conjugate in $S_8$; hence, they are not conjugate in the subgroup $A_8$. We prove the theorem by showing that all elements of order 2 in $\mathrm{PSL}(3, 4)$ are conjugate.

A nonscalar matrix $A \in \mathrm{SL}(3,4)$ corresponds to an element of order 2 in $\mathrm{PSL}(3,4)$ if and only if $A^2$ is scalar, and $A^2$ is scalar if and only if $PAP^{-1}$ is scalar for every nonsingular matrix $P$. Thus, $A$ can be replaced by any matrix similar to it; that is, we may assume that $A$ is a rational canonical form, of which there are three types.

If $A$ is a direct sum of $1 \times 1$ companion matrices, then $A = \mathrm{diag}\{\alpha, \beta, \gamma\}$. But $A^2$ is scalar, so that $\alpha^2 = \beta^2 = \gamma^2$. Since $\mathbb{F}_4$ has characteristic 2, this gives $\alpha = \beta = \gamma$; that is, $A$ is scalar, a contradiction.

Assume $A$ is a $3 \times 3$ companion matrix, say,

$$A = \begin{bmatrix} 0 & 0 & \alpha \\ 1 & 0 & \beta \\ 0 & 1 & \gamma \end{bmatrix},$$

where $\alpha \neq 0$ because $A$ is nonsingular. But $A^2$ has entry $\alpha$ in position $(1,2)$, so that $A^2$ is not scalar.

We conclude that $A$ must be a direct sum of a $1 \times 1$ companion matrix and a $2 \times 2$ companion matrix, say,

$$A = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & 0 & \beta \\ 0 & 1 & \gamma \end{bmatrix}.$$

Now $\det(A) = 1 = \alpha\beta$ (remember that $-1 = 1$ in $\mathbb{F}_4$), so that $\beta = \alpha^{-1}$; as $A^2$ is scalar, we must have $\gamma = 0$. Thus,

$$A = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & 0 & \alpha^{-1} \\ 0 & 1 & 0 \end{bmatrix}.$$

If $\pi$ is a primitive element of $\mathbb{F}_4$, there are only three such matrices:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \pi & 0 & 0 \\ 0 & 0 & \pi^2 \\ 0 & 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} \pi^2 & 0 & 0 \\ 0 & 0 & \pi \\ 0 & 1 & 0 \end{bmatrix}.$$

Note that $A^2 = I$, $B^2 = \pi^2 I$, and $C^2 = \pi I$. It follows that if $M \in \mathrm{SL}(3,4)$ has order 2, that is, $M^2 = I$ (a stronger condition than $M^2$ being scalar), then $M$ is similar to $A$; that is, $M = PAP^{-1}$ for some $P \in \mathrm{GL}(3,4)$. In particular, $\pi^2 B$ and $\pi C$ have order 2, so there are $P, Q \in \mathrm{GL}(3,4)$ with

$$PAP^{-1} = \pi^2 B \quad \text{and} \quad QAQ^{-1} = \pi C.$$

Since the index $[\mathrm{GL}(3,4) : \mathrm{SL}(3,4)] = 3$ (for $\mathrm{GL}/\mathrm{SL} \cong \mathbb{F}_4^{\#} \cong \mathbb{Z}_3$) and since $\mathrm{diag}\{\pi, 1, 1\}$, which has determinant $\pi \neq 1$, commutes with $A$, Exercise C-1.85 allows us to assume that $P, Q \in \mathrm{SL}(3,4)$. It follows that $A$, $B$, $C$ become conjugate in $\mathrm{PSL}(3,4)$, as desired. •

Another proof of this theorem is described in Exercise C-1.83 below.

Around 1900, all simple Lie algebras over the complex numbers were classified by E. Cartan and Killing: there are four infinite families of these, along with five "sporadic" such families, belonging to no infinite family. Chevalley saw that

there are finite simple analogs of the simple Lie algebras: projective unimodular groups, symplectic groups, orthogonal groups, and unitary groups, for example, which are simple groups, while Ree, Steinberg, and Suzuki constructed others (the Suzuki groups $Sz(q)$, where $q$ is a power of 2, are the only finite nonabelian simple groups whose order is not divisible by 3). The ***Classification Theorem of Finite Simple Groups*** says that every finite simple group is cyclic of prime order, is an alternating group, lies in one of finitely many explicit infinite families of finite simple groups, collectively called the simple groups of ***Lie type***, or is one of 26 ***sporadic*** simple groups, the largest of which is the "Monster" (of order approximately $8 \times 10^{53}$). This theorem, completed around 2004, shows that there are infinitely many pairs of nonisomorphic simple groups having the same order, but there do not exist three nonisomorphic simple groups of the same order. (The paper of Cameron [**34**] gives many more consequences of the Classification Theorem to finite groups.) For a more detailed discussion of the Classification Theorem, see E. Artin [**8**], Carter [**37**] (as well as the chapter by Carter in the book edited by Kostrikin–Shafarevich [**128**]), Conway et al. [**44**], Dieudonné [**52**], and Gorenstein–Lyons–Solomon [**81**].

## Exercises

∗ **C-1.83.** Give another proof of Theorem C-1.91 by observing that (1 2 3)(4 5 6 7 8) is an element of order 15 in $A_8$, while $\mathrm{PSL}(3,4)$ contains no elements of order 15.

**Hint.** Use the Jordan canonical form.

**C-1.84.** Prove that $|\mathrm{PSL}(3,2)| = 168$, and show that $\mathrm{PSL}(3,2) \cong \mathrm{PSL}(2,7)$.

∗ **C-1.85.** Let $G$ be a finite group.

(i) For every $a, x \in G$, prove that $C_G(axa^{-1}) = aC_G(x)a^{-1}$.

(ii) If $H \subseteq G$ and $h \in H$, prove that $C_H(h) = C_G(h) \cap H$.

(iii) Let $H$ be a normal subgroup of prime index, and let $x \in H$ satisfy $C_H(x) \subsetneq C_G(x)$. If $y \in H$ is conjugate to $x$ in $G$, prove that $y$ is conjugate to $x$ in $H$.

## C-1.5. More Group Actions

We continue our discussion of group actions, but we begin by introducing projective geometry, for PSL acts on projective space. Aside from its value in other branches of mathematics (in particular, in algebraic topology and in algebraic geometry) as well as its intrinsic interest, we shall see that $\mathrm{PGL}(V) = \mathrm{GL}(V)/Z(V)$ and its subgroup[13] the *projective unimodular group* $\mathrm{PSL}(V) = \mathrm{SL}(V)/SZ(V)$ acts on projective space. This action leads to a second proof of the simplicity of the PSL's, using *multiple transitivity*.

[13] Note that $\mathrm{SL}(V)/\mathrm{SZ}(V) = \mathrm{SL}(V)/(SL(V) \cap Z(V)) \cong (\mathrm{SL}(V)\mathbb{Z}(V)\,\mathrm{SL}(V)) \subseteq \mathrm{GL}(V)/Z(V)$.

## ▪ Projective Geometry

We have seen in Part 1 that modern computers have influenced mathematics.[14] An earlier influence occurred about a century before the Renaissance when artists began drawing in *perspective*, that is, depicting objects on a two-dimensional surface so as to give the right impression of their height, width, depth, and position in relation to each other when viewed from the artist's or the viewer's eye. For example, we know that train tracks, which are parallel, seem to get closer as their distance from us grows. Indeed, it appears that parallel lines actually meet at the horizon. Figure C-1.5 is a sketch drawn by Alberti around 1435.

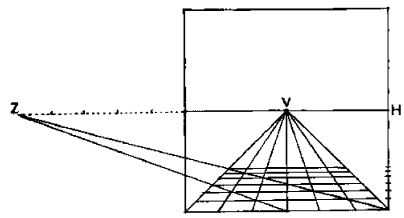



**Figure C-1.5.** Perspective of a tiled floor.

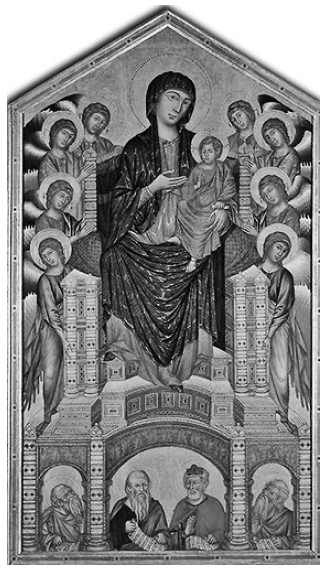


**Figure C-1.6.** Cimabue.

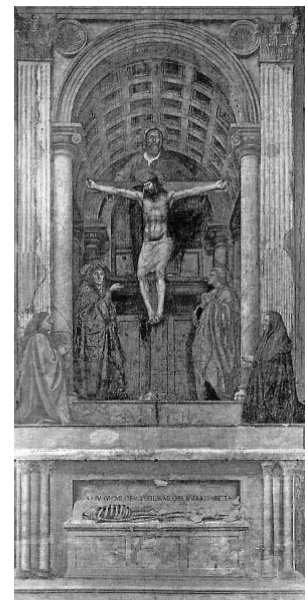


**Figure C-1.7.** Masaccio.

[14]There have always been outside influences on mathematics. For example, commerce, calendars (to know when to plant), areas (to compute taxes after annual Nile flooding), calculus (to aid naval navigation, among other applications), and physics.

In the Middle Ages, paintings were essentially flat. For example, Figure C-1.6, Cimabue's Madonna enthroned, painted around 1285, contrasts with Figure C-1.7, the Holy Trinity of Masaccio, painted around 1425, which appears more three dimensional.

There are two mathematical problems involved in painting the three-dimensional world on a flat canvas. First, what are "vanishing points" (points like v in Figure C-1.5) to which parallel lines on the plane of the tiled floor seem to converge. This problem generalizes to three dimensions, for painters are also interested in vanishing points in space. Second, find the relation between the actual length of an object and its length in the picture depending on its distance behind the plane of the canvas. These specific problems were solved by Brunelleschi around 1413, although the first written account is due to Alberti in 1435. Further theorems were found by Kepler in 1604, Desargues in 1639, and Pascal in 1640. Ultimately, these ideas led to the creation in the early 1800s of ***projective geometry***, by Grassmann, Möbius, Plücker, Poncelet, and von Staudt (see Neumann–Stoy–Thompson [**167**] and Stillwell [**214**]).

We consider only the first problem: how do we attach the horizon to the plane? Now every line in $\mathbb{R}^2$ is parallel to a (unique) line through the origin. Each such line $\ell$ through the origin meets the unit circle $S^1 \subseteq \mathbb{R}^2$ in two antipodal points, $u$ and $-u$, either one of which determines $\ell$. Since we want to attach only one (vanishing) point to the family of all lines parallel to $\ell$, we identify each point on $S^1$ with its antipode. Now the interior of $S^1$ is homeomorphic to $\mathbb{R}^2$, and we may view a semicircle, say, $\{(\cos\theta, \sin\theta) : 0 \le \theta < \pi\}$, as the horizon. Actually, once we identify the endpoints $(\cos 0, \sin 0)$ and $(\cos\pi, \sin\pi)$, we have a circle: indeed, we can view this circle as arising from the real line $\mathbb{R}$ by adjoining a point we call $\infty$. This idea extends to 3-space: every line in space is parallel to a unique line through the origin which intersects the 2-sphere $S^2 \subseteq \mathbb{R}^3$ in two antipodal points. Since the interior of $S^2$ is homeomorphic to $\mathbb{R}^3$, our previous discussion does generalize. Indeed, this idea extends to $S^n \subseteq \mathbb{R}^{n+1}$ for all $n \ge 0$. Here is the usual definition of $\mathbb{P}_n(\mathbb{R})$, ***real projective $n$-space***, used in algebraic topology: it is the quotient space

$$\mathbb{P}_n(\mathbb{R}) = S^n/\sim,$$

where $x \sim y$ (for $x, y \in S^n$) if $x$ and $y$ are antipodal.

Our problem now is to define $\mathbb{P}_n(k)$ for any field $k$. The key idea is to add a coordinate, which will distinguish ***affine points***, that is, points in $k^n$, from ***points at infinity*** (the term used nowadays instead of *vanishing points*).[15]

If $V$ is a vector space over a field $k$, define an equivalence relation on $V - \{0\}$ by $v \sim v'$ if there exists $t \in k^\#$ with $v' = tv$ (we write $k^\#$ to denote the set of nonzero elements in $k$). Denote the equivalence class of $v$ by

$$[v] = \{tv : t \in k^\#\}.$$

[15] In the midst of my writing this section, I ran across Nigel Hitchin's lovely notes on a course he had taught on projective geomety, and he kindly allowed me to borrow from them.

If $V$ is finite-dimensional and $v \in V$ has coordinates $v = (x_0, \ldots, x_n)$ with respect to some basis, then we write

$$[v] = [x_0, \ldots, x_n] = \{(tx_0, \ldots, tx_n) : t \in k^{\#}\},$$

and we call $[x_0, \ldots, x_n]$ the ***homogeneous coordinates*** of $v = (x_0, \ldots, x_n)$. If $x_0 \neq 0$, then we can recapture the last $n$ coordinates of $v$:

$$[v] = [x_0, \ldots, x_n] = [1, x_1/x_0, \ldots, x_n/x_0].$$

Note that homogeneous coordinates $[x_0, \ldots, x_n]$ of $(x_0, \ldots, x_n)$ do detect whether $x_i = 0$ or $x_i \neq 0$ for any $i$. In particular, we are going to distinguish points $[x_0, \ldots, x_n]$ with $x_0 = 0$ and those with $x_0 \neq 0$.

As the construction of $\mathbb{P}_2(\mathbb{R})$ suggests, we define projective space over an arbitrary field $k$ as the set of all lines through the origin in $k^n$.

**Definition.** Let $V$ be an $(n+1)$-dimensional vector space over a field $k$. The ***projective space of dimension*** $n$ ***over*** $k$ is defined by

$$\mathbb{P}(V) = \mathbb{P}_n(k) = \{[v] : v \in V - \{0\}\}.$$

If $U \subseteq V$ is a subspace of $V$, then $\mathbb{P}(U)$ is the ***projective subspace*** defined by

$$\mathbb{P}(U) = \{[u] \in \mathbb{P}(V) : u \in U\}.$$

A ***projective hyperplane*** is a projective subspace $\mathbb{P}(U)$ if $\dim(U) = \dim(V) - 1$.

For example, a ***projective line*** $\mathbb{P}_1(k)$ is

$$\mathbb{P}_1(k) = \{[u] : u \in U - \{0\}\},$$

where $\dim(U) = 2$. When $k = \mathbb{C}$, then $\mathbb{P}_1(\mathbb{C})$ is the ***Riemann sphere***; that is, $\mathbb{C}$ is identified with $\mathbb{R}^2$ as usual, and $\mathbb{P}_1(\mathbb{C})$ is $S^2$, the (real) plane with $\infty$ adjoined.

**Example C-1.92.** If $k = \mathbb{F}_2$, then $k^2$ has four points: $O = (0,0)$, $a = (1,0)$, $b = (0,1)$, and $c = (1,1)$, and six lines, as in Figure C-1.8.

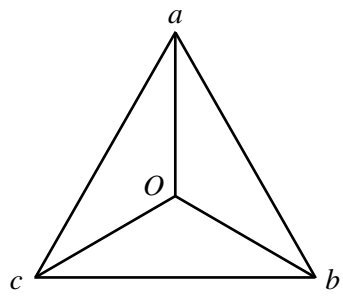



**Figure C-1.8.** Affine plane.

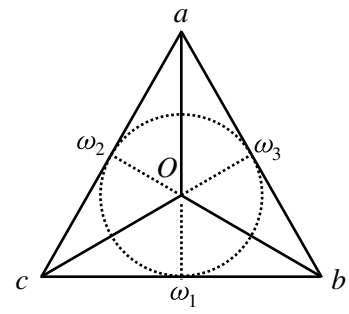



**Figure C-1.9.** $\mathbb{P}_2(\mathbb{F}_2)$.

There are three sets of parallel lines in $k^2$: $Oa$ and $bc$; $Ob$ and $ac$; $Oc$ and $ab$. The projective plane $\mathbb{P}_2(\mathbb{F}_2)$ (sometimes called the ***Fano plane***) is obtained by adding new points $\omega_1$, $\omega_2$, and $\omega_3$ forcing parallel lines to meet. There are seven

lines: the six lines in the affine plane $k^2$ (each lengthened with some $\omega_i$) plus the line at infinity $\omega_1, \omega_2, \omega_3$. ◀

**Proposition C-1.93.** *For every field $k$ and all $n \geq 1$, there is a disjoint union*

$$\mathbb{P}_n(k) = k^n \cup \mathbb{P}_{n-1}(k).$$

**Proof.** If $U = k^{n+1}$, then $U_0$, defined by

$$U_0 = \{[u] = [1, x_1, \ldots, x_n] \in U - \{0\}\} \subseteq \mathbb{P}_n(k),$$

can obviously be identified with $k^n$, while its complement $\{[u] = [0, x_1, \ldots, x_n] \in U - \{0\}\} \subseteq \mathbb{P}_n(k)$ can be identified with $\mathbb{P}_{n-1}(k)$. •

Thus, a line $\ell = \{tu = t(x_1, \ldots, x_n)\}$ in $k^n$ acquires an extra point in $\mathbb{P}(V)$, namely, $[0, x_1, \ldots, x_n]$; the projective line corresponding to $\ell$ is $\mathbb{P}(U)$, where $U$ is the two-dimensional subspace

$$U = \langle (1, x_1, \ldots, x_n), (0, x_1, \ldots, x_n) \rangle \subseteq V.$$

**Corollary C-1.94.** *If $k$ is a finite field with $q = p^m$ elements, then $\mathbb{P}_2(k)$ has exactly $q^2 + q + 1$ elements and each projective line in it has exactly $q + 1$ elements.*

**Proof.** Proposition C-1.93 says that $|\,\mathbb{P}_2(k)| = q^2 + |\,\mathbb{P}_1(k)|$. But a line $\ell$ in $k^2$ has $q$ points, and its corresponding projective line has one more point. •

**Proposition C-1.95.** *Let $\mathbb{P}(V)$ be the projective space arising from an $n$-dimensional vector space $V$ over a field $k$, where $n \geq 2$.*

(i) *If $[x], [y]$ are distinct elements of $\mathbb{P}(V)$, then $x, y$ are linearly independent.*

(ii) *Given distinct points $[u], [v] \in \mathbb{P}(V)$, there is a unique projective line containing them.*

(iii) *If $\dim(V) = 3$, then distinct lines $\mathbb{P}(U)$ and $\mathbb{P}(U')$ in the projective plane $\mathbb{P}(V)$ intersect in a unique point.*

**Proof.**

(i) If $u, v$ are not linearly independent in $V$, then there is a nonzero scalar $t \in k$ with $v = tu$ (or $u = tv$), which gives a contradiction:

$$[u] = [tv] = [v].$$

(ii) By (i), $U = \langle u, v \rangle$ is a two-dimensional subspace of $V$, and so $\mathbb{P}(U)$ is a projective line in $\mathbb{P}(V)$ containing $[u]$ and $[v]$.[16]

To prove uniqueness, suppose there is a two-dimensional subspace $U'$ of $V$ with $\mathbb{P}(U') = \mathbb{P}(U)$. Since $[u], [v] \in \mathbb{P}(U')$, we have $u, v \in U'$, and so $U = \langle u, v \rangle \subseteq U'$. As $\dim(U) = 2 = \dim(U')$, it follows that $U = U'$ and $\mathbb{P}(U) = \mathbb{P}(U')$.

[16] Here is a more visual proof. Given a point $u \in k^n$, define the ***pencil at*** $u$ to be the family of all lines passing through $u$; there is a way to view projective lines as pencils. Given pencils at $u$ and $u'$, with $u \neq u'$, then their intersection is the projective point arising from the line joining $u$ and $u'$.

(iii) Now

$$\dim(V) \geq \dim(U + U') = \dim(U) + \dim(U') - \dim(U \cap U');$$

that is, $3 \geq 2 + 2 - \dim(U \cap U')$, so that $\dim(U \cap U') \geq 1$. Since $U$ and $U'$ are two-dimensional, we have $\dim(U \cap U') \leq 2$ with equality if and only if $U = U'$. But $U \neq U'$, so that $\dim(U \cap U') = 1$, and $\mathbb{P}(U \cap U')$ is a projective point in $\mathbb{P}(U) \cap \mathbb{P}(U')$. Uniqueness follows from part (i): were there two points of intersection, then $\mathbb{P}(U) = \mathbb{P}(U')$. •

Two parallel lines $\ell$ and $\ell'$ in $\mathbb{R}^2$ have equations

$$y - mx - b = 0 \qquad \text{and} \qquad y - mx - b' = 0,$$

and their corresponding lines in $\mathbb{P}_2(\mathbb{R})$ intersect in $[0, 1, -m]$. We have forced parallel lines in the plane to meet.

**Remark.** When investigating magic squares in 1782, Euler considered the following combinatorial problem. Suppose there are 36 officers of 6 ranks and from 6 regiments. If the regiments are numbered 1 through 6 and the ranks are captain, major, lieutenant, ..., then each officer has a double label; for example, captain 3 or major 4. Euler asked whether there is a $6 \times 6$ formation of these officers so that each row and each column contains exactly one officer of each rank and one officer from each regiment. More generally, call two $n \times n$ Latin squares[17] $[a_{ij}]$ and $[b_{ij}]$ with entries in $\{1, 2, \ldots, n\}$ ***orthogonal*** if all the entries in $[c_{ij}]$ are distinct, where $c_{ij} = (a_{ij}, b_{ij})$.[18] Thus, Euler asked whether there exists a pair of orthogonal $6 \times 6$ Latin squares. It turns out that the answer is no. But it also turns out that a positive answer suggests a combinatorial version of a projective plane.

**Definition.** A ***projective plane of order*** $n$ is a set $X$ with $|X| = n^2 + n + 1$ together with a family of subsets, called ***lines***, each having exactly $n + 1$ points, such that every two points determine a unique line.

Call a family $A_1, \ldots, A_m$ of $n \times n$ Latin squares over $\{1, 2, \ldots, n\}$ ***orthogonal*** if $A_r$ and $A_s$ are orthogonal for all $r \neq s$.

**Theorem**. *If $n \geq 3$, then there exists a projective plane of order $n$ if and only if there exists an orthogonal family of $n - 1$ $n \times n$ Latin squares.*

**Proof.** Ryser [**195**], p. 92. •

If $k = \mathbb{F}_q$, then Corollary C-1.94 gives $|\,\mathbb{P}_2(k)| = q^2 + |\,\mathbb{P}_1(k)| = q^2 + q + 1$, for each projective line has $q + 1$ points. Thus, $\mathbb{P}_2(\mathbb{F}_q)$ is a projective plane of order $q$. Projective planes of order $q$ need not be unique; for example, it is known that there are four projective planes of order 9, only one of which arises from $\mathbb{F}_9$. Euler proved that if $n \not\equiv 2 \bmod 4$, then there exists an orthogonal pair of $n \times n$ Latin squares,

[17] A ***Latin square over*** $\{1, 2, \ldots, n\}$ is an $n \times n$ matrix with no integer $i$ occurring twice in any row or in any column.

[18] The matrix $[c_{ij}]$ defining orthogonality involves two types of elements, say, rank and regiment, which Euler denoted by $c_{\alpha,\beta}$, using two alphabets. He called the matrix $[c_{\alpha,\beta}]$ a ***Graeco-Latin square***, and this is the origin of the term *Latin square*.

and so $n = 6$ was the first number not covered by his theorem (this is why he asked about 36 officers). Tarry (1901) was the first to prove there is no orthogonal pair of $6 \times 6$ Latin squares; it follows that there is no projective plane of order 6 (which requires an orthogonal family of 5 Latin squares). Euler conjectured that there does not exist an orthogonal pair of $n \times n$ matrices if $n \equiv 2 \bmod 4$. In 1958, Parker disproved Euler's conjecture by constructing an orthogonal pair of $10 \times 10$ Latin squares (see the cover of [**185**]). The problem of finding all $n$ for which there exists a projective plane of order $n$ remains open today.

Here is the best result about nonexistence of projective planes now known.

**Theorem (Bruck–Ryser).** *If $n \equiv 1 \bmod 4$ or $n \equiv 2 \bmod 4$ and $n$ is not a sum of two squares, then there does not exist a finite projective plane of order $n$.*

**Proof.** See Ryser [**195**], p. 111. •

The contrapositive of the Bruck–Ryser Theorem is also interesting. If $n \equiv 1 \bmod 4$ or $n \equiv 2 \bmod 4$ and a projective plane of order $n$ does exist, then $n$ is a sum of two squares. The first few $n \geq 3$ with $n \equiv 1$ or $2 \bmod 4$ are

$$5, \quad 6, \quad 9, \quad 10, \quad 13, \quad 14, \quad 17, \quad 18, \quad 21 \quad 22.$$

Some of these are primes or prime powers,[19] so that projective planes of these orders must exist; these numbers are indeed sums of two squares:

$$5 = 1 + 4, \qquad 9 = 0 + 9, \qquad 13 = 4 + 9, \qquad 17 = 1 + 16.$$

However, there are no projective planes of order 6, 14, 21, or 22, for these numbers are not sums of two squares. The smallest $n$ not covered by the Bruck–Ryser Theorem is $n = 10$ ($10 \equiv 2 \bmod 4$ and $10 = 1 + 9$). Using a massive amount of computer calculation, Clement Lam showed, in 1988, that no projective plane of order 10 exists. As of this writing, it is unknown whether a projective plane of order 12 exists. ◀

We now show that linear groups act on projective space. We introduced projective unimodular groups $\mathrm{SL}(V)/\mathrm{SZ}(V)$ by considering composition series of general linear groups. But there is a geometric reason for dividing out the center $\mathrm{SZ}(V)$, for the action of GL takes account of homogeneous coordinates of points in projective space.

**Theorem C-1.96.** *Let $V$ be an $(n+1)$-dimensional vector space over a field $k$.*

(i) *If $W$ is a vector space over $k$, then every nonsingular linear transformation $T\colon V \to W$ determines a unique function (called a **projectivity**)*

$$\mathbb{P}(T)\colon\ \mathbb{P}(V) \to \mathbb{P}(W);$$

*namely, $\mathbb{P}(T)\colon [v] \mapsto [Tv]$.*

[19] Fermat's Two Squares Theorem (see Cuoco–Rotman [**47**], p. 342) says that an odd prime $p$ is a sum of two squares if and only if $p \equiv 1 \bmod 4$. Since there exists a projective plane of order $p$, the Bruck–Ryser Theorem implies Fermat's Theorem. In fact, the Bruck–Ryser Theorem says that $p^e$ is a sum of two squares for all $e \geq 1$ if $p \equiv 1 \bmod 4$.

(ii) *If $T, S\colon V \to W$ are nonsingular linear transformations, then $\mathbb{P}(T) = \mathbb{P}(S)$ if and only if $S = \lambda T$ for some $\lambda \in k^{\#}$.*

(iii) *Define the* ***projective general linear group*** *by*

$$\mathrm{PGL}(V) = \mathrm{GL}(V)/Z(V),$$

*where $Z(V)$ is the center of $\mathrm{GL}(V)$. Then $\mathrm{PGL}(V)$ acts faithfully on $\mathbb{P}(V)$.*

(iv) *Define*

$$\mathrm{PGL}(n+1,k) = \mathrm{GL}(n+1,k)/Z(n+1,k),$$

*where $Z(n+1,k)$ is the center of $\mathrm{GL}(n+1,k)$. Then $\mathrm{PGL}(n+1,k)$ acts on $\mathbb{P}_n(k)$.*

**Proof.**

(i) If $U$ is a subspace of $V$, then $T(U)$ is a subspace of $W$; moreover, $T$ nonsingular implies $\dim(U) = \dim(TU)$. In particular, if $\dim(U) = 1$, then $\dim(TU) = 1$, and so $[u] = \mathbb{P}(U)$ implies $[Tu] = \mathbb{P}(TU)$; that is, $\mathbb{P}(T)\colon \mathbb{P}(V) \to \mathbb{P}(W)$ is a well-defined function.

(ii) If $S = \lambda T$ for some $\lambda \in k^{\#}$, then for all $[v] \in \mathbb{P}(V)$ we have

$$\mathbb{P}(S)\colon [v] \mapsto [Sv] = [\lambda Tv] = [Tv] = \mathbb{P}(T)[v],$$

and so $\mathbb{P}(S) = \mathbb{P}(T)$.

Conversely, if $\mathbb{P}(S) = \mathbb{P}(T)$, then $[Sv] = [Tv]$ for all $v \in V - \{0\}$; that is, for each nonzero $v \in V$, there is $\lambda \in k^{\#}$, depending on $v$, with $S(v) = \lambda T(v)$. In particular, if $v_0, \dots, v_n$ is a basis of $V$, then there are $\lambda_i \in k^{\#}$, for all $i$, with

$$Sv_i = \lambda_i T v_i.$$

But $v = \sum_0^n \alpha_i v_i$ for $\alpha_i \in k$, so that

$$Sv = \sum_0^n \alpha_i S v_i = \sum_0^n \alpha_i \lambda_i T v_i.$$

On the other hand,

$$Sv = \lambda T v = \sum_0^n \lambda \alpha_i T v_i.$$

As $T$ is nonsingular, the list $Tv_0, \dots, Tv_n$ is linearly independent; hence, $\lambda\alpha_i T v_i = \lambda_i \alpha_i T v_i$ and $\lambda_i = \lambda$ for all $i$. It follows that $S = \lambda T$.

(iii) The map $\pi\colon T \mapsto \mathbb{P}(T)$ is a homomorphism $\mathrm{GL}(V) \to S_{\mathbb{P}(V)}$ (the symmetric group), and part (ii) says that $\ker \pi$ is the subgroup of scalar transformations. But Corollary A-7.41 in Part 1 says that $\ker \pi = Z(V)$, the center of $\mathrm{GL}(V)$, so that $\mathrm{PGL}(V) = \mathrm{GL}(V)/Z(V)$ acts on $\mathbb{P}(V)$.

To see that $\mathbb{P}(V)$ is a faithful $\mathrm{PSL}(V)$-set, it suffices to show that if $T \in \mathrm{GL}(V)$ fixes $\mathbb{P}(V)$ pointwise, then $T$ is a scalar transformation; that is, if $[Tv] = [v]$ for every $v \in V$, then $T = \lambda 1_V$ for some nonzero $\lambda \in k$. If $v_1, \dots, v_n$ is a basis of $V$, then $[Tv_i] = [v_i]$ for all $i$: there are nonzero $\lambda_i \in k$ with $Tv_i = \lambda_i v_i$. We claim, if $i \neq j$, that $\lambda_i = \lambda_j$. By hypothesis,

$[T(v_i + v_j)] = [v_i + v_j]$; that is, $T(v_i + v_j) = \mu(v_i + v_j) = \mu v_i + \mu v_j$ for some nonzero $\mu \in k$. But $T(v_i + v_j) = \lambda_i v_i + \lambda_j v_j$. Therefore, $\lambda_i = \mu = \lambda_j$.

(iv) This follows from part (iii) if we choose a basis of $V$ and write linear transformations in $\mathrm{GL}(V)$ as matrices and elements of $\mathbb{P}(V)$ in terms of homogeneous coordinates. •

**Example C-1.97.** By Theorem C-1.96, $\mathrm{PGL}(2, k)$ acts on the projective line $\mathbb{P}_1(k)$. A nonsingular linear transformation $T\colon k^2 \to k^2$ corresponds to a nonsingular matrix

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

and $\mathbb{P}(T)\colon [x_0, x_1] \mapsto [A(x_0, x_1)] = [ax_0 + bx_1, cx_0 + dx_1]$. But $cx_0 + dx_1 \neq 0$, since $A$ is nonsingular, and $[ax_0 + bx_1, cx_0 + dx_1] = [(ax_0 + bx_1)/(cx_0 + dx_1), 1]$, so that $\mathrm{PGL}(2, k)$ acts on $\mathbb{P}_1(k)$ as linear fractional transformations. ◀

Further study of projective geometry continues with duality, theorems of Desargues and of Pappus, and the introduction of coordinates to projective space (see Artin [**8**] or Coxeter–Moser [**46**]).

## ▪ Multiple Transitivity

Although some results hold in greater generality, we assume for the remainder of this section that all groups $G$ and all $G$-sets $X$ are finite.

Recall that if $x \in X$, where $X$ is a $G$-set, then its *orbit*, denoted by $\mathcal{O}(x)$, is $\{gx : g \in G\}$, and its *stabilizer*, denoted by $G_x$, is the subgroup $\{g \in G : gx = x\}$. A $G$-set $X$ is *transitive* if $\mathcal{O}(x) = X$ for some (and hence every) $x \in X$; that is, for each $x, y \in X$, there exists $g \in G$ with $y = gx$. The *degree* of $X$ is $|X|$.

Here are two useful elementary results.

**Proposition C-1.98.** *If $X$ is a $G$-set and $x \in X$, then $X - \{x\}$ is a $G_x$-set.*

**Proof.** If $h \in G_x$ and $y \neq x$ (that is, $y \in X - \{x\}$), then $hy \neq x$, for every $h \in G$ permutes $X$. Hence, $hy \in X - \{x\}$. •

**Proposition C-1.99.** *Let $X$ be a transitive $G$-set, and let $x, y \in X$. If $y = tx$ for some $t \in G$, then $G_x$ and $G_y$ are conjugate: $G_y = tG_x t^{-1}$.*

**Proof.** If $g \in G_x$, then $gx = x$ and $tgt^{-1}y = tgx = tx = y$; hence, $tG_x t^{-1} \subseteq G_y$. The reverse inclusion is similar. •

Recall that an action $\alpha\colon G \times X \to X$ gives a homomorphism $G \to S_X$; namely, $g \mapsto \alpha_g\colon x \to gx$. This action is called *faithful* if $G \to S_X$ is injective.

**Theorem C-1.100.** *If $G$ is a finite group, $X$ is a transitive $G$-set of degree $n$, and $x \in X$, then*

$$|G| = n|G_x|.$$

*Moreover, if $X$ is faithful, then $|G_x|$ is a divisor of $(n-1)!$.*

**Proof.** By Theorem C-1.16, we have $|\mathcal{O}(x)| = [G : G_x]$, so that $|G| = n|G_x|$.

If $X$ is faithful, then $G$ is isomorphic to a subgroup of $S_X$. Now $X - \{x\} \subseteq X$ is a $G_x$-set by Proposition C-1.98; since $G_x \subseteq G$, it follows that $X - \{x\}$ is a faithful $G_x$-set. Thus, $G_x$ can be viewed as a subgroup of $S_{X-\{x\}} \cong S_{n-1}$, and so $|G_x|$ divides $(n-1)!$. •

Some $G$-sets are more transitive than others.

**Definition.** Let $X$ be a $G$-set of degree $n$. If $r \leq n$, then $X$ is ***r*-transitive** if, for every pair of $r$-tuples $(x_1, \ldots, x_r)$ and $(y_1, \ldots, y_r)$ having distinct entries in $X$, there exists $g \in G$ with $gx_i = y_i$ for all $i = 1, \ldots, r$.

The ***stabilizer*** $G_{x_1,\ldots,x_r}$ of $x_1, \ldots, x_r$ in $X$ is defined by

$$G_{x_1,\ldots,x_r} = \{g \in G : gx_i = x_i \text{ for all } i = 1, \ldots, r\}.$$

Clearly, 1-transitivity is transitivity. If $r \geq 2$, then every $r$-transitive $G$-set $X$ is $(r-1)$-transitive. We say that $r$-transitive $G$-sets are ***multiply transitive*** if $r \geq 2$. In particular, 2-transitive sets are called ***doubly transitive***, 3-transitive sets are called ***triply transitive***, and so forth.

Note that if $X$ is a $G$-set and $B = \{x_1, \ldots, x_r\} \subseteq X$, then

$$G_B = \{g \in G : gx_i \in B \text{ for all } i = 1, \ldots, r\}$$

may be distinct from $G_{x_1,\ldots,x_r}$; in general, $G_{x_1,\ldots,x_r} \subseteq G_B$, but the inequality may be strict.

The most obvious example of a multiply transitive $G$-set occurs when $G$ is the symmetric group $S_n$: if $X = \{1, \ldots, n\}$, then $X$ is an $n$-transitive $S_n$-set. If $G$ is the alternating group $A_n$, where $n \geq 3$, then $X$ is an $(n-2)$-transitive $A_n$-set (see Proposition C-1.106).

**Example C-1.101.** Mathieu found the first sporadic simple groups in the nineteenth century; there are five such groups: $M_{11}$, $M_{12}$, $M_{22}$, $M_{23}$, $M_{24}$, and they are multiply transitive (the subscript gives the degree of the set on which they act). Now $M_{22}$ is 3-transitive, $M_{11}$ and $M_{23}$ are 4-transitive, and $M_{12}$ and $M_{24}$ are 5-transitive. (See Rotman [**188**], pp. 286–293, for proofs of these results, and pp. 293–306 for a discussion of the relation of Mathieu groups to ***Steiner systems***.) ◀

**Example C-1.102.** In Example C-1.9, we saw that if $k$ is a field, $f(x) \in k[x]$ is a polynomial with no repeated roots, and if $E/k$ is a splitting field of $f$, then $G = \mathrm{Gal}(E/k)$ acts faithfully on the set $X = \{\alpha_1, \ldots, \alpha_n\}$ of the roots of $f$. Moreover, $f$ is irreducible if and only if $X$ is a transitive $G$-set.

Now $f$ factors in $k(\alpha_1)$, say, $f(x) = (x - \alpha_1)f_1(x)$. It is easy to see that $G_1 = \mathrm{Gal}(E/k(\alpha_1)) \subseteq G$ is the stabilizer of $\{\alpha_1\}$, and so $G_1$ acts on $X - \{\alpha_1\}$; in fact, $G_1$ is the Galois group of $f_1$. Thus, $G$ acts doubly transitively on $X$ if and only if $f \in k[x]$ and $f_1 \in k(\alpha_1)[x]$ are irreducible. Iterating, $X$ is 3-transitive if $f_2$ is irreducible, where $f = (x-\alpha_1)(x-\alpha_2)f_2$. It can be proved, using the Classification Theorem of Finite Simple Groups, that there are no faithful $r$-transitive $G$-sets for $r > 5$ unless $G$ is a symmetric or an alternating group (see Cameron [**34**];

Cameron's article also gives other applications of the Classification Theorem to previously unsolved conjectures about finite groups). It follows that if a Galois group $G$ is $r$-transitive for $r \geq 6$, then $G$ is symmetric or alternating. ◀

**Lemma C-1.103.** *Let $X$ be a $G$-set. If $r \geq 2$, then $X$ is an $r$-transitive $G$-set if and only if, for each $x \in X$, the $G_x$-set $X - \{x\}$ is $(r-1)$-transitive.*

**Proof.** Let $X$ be an $r$-transitive $G$-set. If $(y_1, \dots y_{r-1})$ and $(z_1, \dots z_{r-1})$ are $(r-1)$-tuples of distinct elements in $X - \{x\}$, then there is $g \in G$ with $gx = x$ and $gy_i = z_i$ for all $i = 1, \dots, r-1$. By Proposition C-1.98, $X - \{x\}$ is a $G_x$-set; since $gx = x$, we have $g \in G_x$, and so $X - \{x\}$ is an $(r-1)$-transitive $G_x$-set.

Conversely, let $(x_1, \dots x_r)$ and $(y_1, \dots y_r)$ be $r$-tuples of distinct elements in $X$. By hypothesis, there is $g \in G_{x_r}$ (so $gx_r = x_r$) with $gx_i = y_i$ for all $i < r$, and there is $h \in G_{y_1}$ with $hy_j = y_j$ for all $j < r$ (of course, $hy_1 = y_1$ since $h \in G_{y_1}$). The composite $hg$ sends $x_i \to y_i$ for all $i = 1, \dots, r$. Thus, $X$ is $r$-transitive. •

We now generalize Theorem C-1.100.

**Theorem C-1.104.** *If $X$ is an $r$-transitive $G$-set of degree $n$, then*

$$|G| = n(n-1)(n-2)\cdots(n-r+1)|G_{x_1,\dots,x_r}|$$

*for every choice of $r$ distinct elements $x_1, \dots, x_r$ in $X$. Moreover, if $X$ is faithful, then $|G_{x_1,\dots,x_r}|$ divides $(n-r)!$.*

**Proof.** The proof is by induction on $r$, the base step being Theorem C-1.100. Now $G_{x_1}$ acts $(r-1)$-transitively on $X - \{x_1\}$, by Lemma C-1.103. Induction completes the proof. •

**Definition.** An $r$-transitive $G$-set $X$ is ***sharply $r$-transitive*** if only the identity fixes $r$ distinct elements of $X$.

**Theorem C-1.105.** *The following conditions are equivalent for a faithful $r$-transitive $G$-set $X$ of degree $n$.*

(i) *$X$ is sharply $r$-transitive.*

(ii) *If $(x_1, \dots, x_r)$ and $(y_1, \dots, y_r)$ are $r$-tuples of distinct elements of $X$, then there is a unique $g \in G$ with $gx_i = y_i$ for all $i$.*

(iii) $|G| = n(n-1)\cdots(n-r+1)$.

(iv) *The stabilizer of any $r$ distinct elements of $X$ is trivial.*

*If $r \geq 2$, then these conditions are equivalent to the following:*

(v) *For every $x \in X$, the $G_x$-set $X - \{x\}$ is sharply $(r-1)$-transitive.*

**Proof.** All verifications are routine; the proofs (i) ⇒ (ii) ⇒ (iii) ⇒ (iv) ⇒ (i) are left as exercises for the reader. Finally, the reader should prove that (v) is equivalent to any of (i) through (iv). •

**Proposition C-1.106.** *The symmetric group $S_n$ acts sharply $n$-transitively on $X = \{1, \dots, n\}$ for every $n \geq 1$.*

*If $n \geq 3$, the alternating group $A_n$ acts sharply $(n-2)$-transitively on $X$.*

**Proof.** The first statement is obvious, for $S_n$ contains every permutation of $X$. We prove the second statement by induction on $n \geq 3$. If $n = 3$, then $A_3 = \langle (1\ 2\ 3) \rangle$ acts sharply transitively on $X = \{1, 2, 3\}$. If $n > 3$, then $(A_n)_i$, the stabilizer of $i$, where $1 \leq i \leq n$, is isomorphic to $A_{n-1}$; by induction, it acts sharply $(n-3)$-transitively on $X - \{i\}$, and Theorem C-1.105 completes the proof. •

## ▪ PSL Redux

We now give a second proof of simplicity of the projective unimodular groups, using a simplicity criterion due to Iwasawa. The action of PSL on the projective line is not as useful as when PSL acts on a projective space of higher dimension, and so the reader may prefer the previous proof of the simplicity of $\mathrm{PSL}(2, k)$.

The following notion arises in representation theory.

**Definition.** If $X$ is a $G$-set, then a ***block*** is a subset $B \subseteq X$ such that, for every $g \in G$, either $gB = B$ or $gB \cap B = \varnothing$.

A $G$-set $X$ is ***primitive*** if it is transitive and it contains no ***nontrivial blocks***; that is, $X$ contains no block other than $\varnothing$, $X$, or one-point subsets.

**Theorem C-1.107.** *Every doubly transitive $G$-set $X$ is primitive.*

**Proof.** If $X$ has a nontrivial block $B$, then there are elements $x, y, z \in X$ with $x, y \in B$ and $z \notin B$. Since $X$ is doubly transitive, there is $g \in G$ with $gx = x$ and $gy = z$. Hence, $x \in B \cap gB$ and $B \neq gB$, a contradiction. •

Here is a characterization of primitive $G$-sets.

**Theorem C-1.108.** *Let $X$ be a transitive $G$-set. Then $X$ is primitive if and only if, for each $x \in X$, the stabilizer $G_x$ is a maximal subgroup of $G$.*

**Proof.** If $G_x$ is not maximal, there is a subgroup $U$ with $G_x \subsetneq U \subseteq G$; we show that $Ux = \{gx : G \in U\}$ is a nontrivial block. If $g \in G$ and $Ux \cap gUx \neq \varnothing$, then $ux = gu'x$ for $u, u' \in U$. Since $u^{-1}gu'$ fixes $x$, we have $u^{-1}gu' \in G_x \subsetneq U$, and so $g \in U$. Hence, $Ux$ is a block, for $gUx = Ux$.

We now show that $Ux$ is a nontrivial block. Clearly, $Ux$ is nonempty. Choose $g \in G$ with $g \notin U$. If $Ux = X$, then for every $y \in X$, there is $u \in U$ with $y = ux$; in particular, $gx = ux$ for some $u \in U$, Therefore, $g^{-1}u \in G_x \subsetneq U$ and $g \in U$, a contradiction. Finally, if $Ux$ is a singleton, then $U \subseteq G_x$, contradicting $G_x \subsetneq U$. Therefore, $X$ is not primitive.

Assume that every $G_x$ is a maximal subgroup of $G$ and that there exists a nontrivial block $B$ in $X$. Define a subgroup $S$ of $G$:

$$S = \{g \in G : gB = B\}.$$

Choose $x \in B$. If $gx = x$, then $x \in B \cap gB$, and so $gB = B$ (because $B$ is a block); hence, $G_x \subseteq S$. Since $B$ is nontrivial, there is $y \in B$ with $y \neq x$. Transitivity provides $g \in G$ with $gx = y$; hence, $y \in B \cap gB$, and so $gB = B$. Thus, $g \in S$ while $g \notin G_x$; that is, $G_x \subsetneq S$. If $S = G$, then $gB = B$ for all $g \in G$, and this contradicts $X \neq B$ being a transitive $G$-set. Therefore, $G_x \subsetneq S \subsetneq G$, contradicting the maximality of $G_x$. •

**Lemma C-1.109.** *Let $X$ be a $G$-set and let $x, y \in G$.*

(i) *If $H \subseteq G$, then $Hx \cap Hy \neq \varnothing$ implies $Hx = Hy$.*

(ii) *If $H \lhd G$, then for every $x \in X$ the subset $Hx$ is a block of $X$.*

**Proof.**

(i) It is easy to see that $Hx = Hy$ if and only if $y \in Hx$. If $Hx \cap Hy \neq \varnothing$, then there are $h, h' \in H$ with $hy = h'x$. Hence, $y = h^{-1}h'x \in Hx$ and $Hy = Hx$.

(ii) Assume that $gHx \cap Hx \neq \varnothing$. Since $H \lhd G$, we have $gHx \cap Hx = Hgx \cap Gx$. There are $h, h' \in H$ with $hgx = h'x$, and so $gx = h^{-1}h'x \in Hx$. Therefore, $gHx = Hx$. •

**Theorem C-1.110.**

(i) *If $X$ is a faithful primitive $G$-set of degree $n \geq 2$, if $H \lhd G$, and if $H \neq \{1\}$, then $X$ is a transitive $H$-set.*

(ii) *$n$ divides $|H|$.*

**Proof.**

(i) Lemma C-1.109 shows that $Hx$ is a block for every $x \in X$. Since $X$ is primitive, either $Hx = \varnothing$ (obviously impossible), $Hx = \{x\}$, or $Hx = X$. If $Hx = \{x\}$ for some $x \in X$, then $H \subseteq G_x$. But if $g \in G$, then normality of $H$ gives $H = gG_xg^{-1} \subseteq G_{gx}$. Since $X$ is transitive, $H \subseteq \bigcap_{y \in X} G_y = \{1\}$, for $X$ is faithful, and this is a contradiction. Therefore, $Hx = X$ and $X$ is a transitive $H$-set.

(ii) This follows from Theorem C-1.100. •

The simplicity criterion we will use involves the following group-theoretic property.

**Definition.** A group $G$ is ***perfect*** if $G = G'$.

For example, every simple nonabelian group is perfect: since the commutator subgroup $G'$ is a normal subgroup, either $G' = \{1\}$ or $G' = G$. The first case is ruled out because we are assuming that $G$ is not abelian.

Here are some examples of perfect groups that are not simple.

**Proposition C-1.111.** *If $V$ is a vector space over a field $k$, then the groups* $\mathrm{SL}(V)$ *are perfect unless* $\dim(V) = 2$ *and $k = \mathbb{F}_2$ or $k = \mathbb{F}_3$.*

**Proof.** Suppose first that some transvection $T \in \mathrm{SL}(V)$ is a commutator, say, $T = [M, N] = MNM^{-1}N^{-1}$ for some $M, N \in \mathrm{SL}(V)$. If $T'$ is another transvection, then $T$ and $T'$ are conjugate in $\mathrm{GL}(V)$ (Lemma C-1.84); there is $U \in \mathrm{GL}(V)$ with $T' = T^U$, where we write $T^U = UTU^{-1}$. Therefore, $T' = T^U = [M, N]^U = [M^U, N^U]$. But both $M^U$ and $N^U$ lie in $\mathrm{SL}(V)$ because $\mathrm{SL} \lhd \mathrm{GL}$, and so $T'$ also is a commutator. Since $\mathrm{SL}(V)$ is generated by transvections, it follows that $\mathrm{SL}(V)$ is perfect.

It remains to show that there exists a transvection $T \in \mathrm{SL}(V) \cong \mathrm{SL}(n,k)$ which is a commutator. If $n \geq 3$ and $e_1, \ldots, e_n$ is the standard basis of $V \cong k^n$, define $T \in \mathrm{GL}(V)$ by $T(e_i) = e_i$ for all $i \neq 3$ and $T(e_3) = e_3 - e_2 - e_1$. The matrix of $T$ is

$$T = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that $T = [\varphi, h]$, where $h = -e_2 - e_1$ and $\varphi$ is the linear functional which selects the third coordinate of a vector $v = (\lambda_1, \ldots, \lambda_n)$; that is, $\varphi(v) = \lambda_3$. Define $M = B_{13}(-1)$, and define $N$ by

$$Ne_1 = -e_2, \quad Ne_2 = e_1, \quad Ne_i = e_i \quad \text{for all } i \geq 3.$$

Now $[M,N] = MNM^{-1}N^{-1} = T$; we illustrate this in the $3 \times 3$ case:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Finally, if $n = 2$ and $|k| \geq 4$, there exists $\lambda \in k$ with $\lambda^2 \neq 1$. But

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \lambda^2 - 1 \\ 0 & 1 \end{bmatrix};$$

thus, the transvection $B_{12}(\lambda^2 - 1)$ is a commutator. •

**Theorem C-1.112 (Iwasawa).** *Let $G$ be a perfect group and let $X$ be a faithful primitive $G$-set. If there is $x \in X$ and an abelian normal subgroup $U$ in $G_x$ whose conjugates $\{gUg^{-1} : g \in G\}$ generate $G$, then $G$ is a simple group.*

**Proof.** Let $H \neq \{1\}$ be a normal subgroup of $G$. By Theorem C-1.110, $H$ acts transitively on $X$. By hypothesis, each $g \in G$ has the form $g = \prod g_i u_i g_i^{-1}$, where $g_i \in G$ and $u_i \in U$. Now $G = HG_x$, by Exercise C-1.72 on page 49, so that $g_i = h_i s_i$ for each $i$, where $h_i \in H$ and $s_i \in G_x$. Normality of $U$ in $G_x$ now gives

$$g = \prod h_i s_i u_i s_i^{-1} h_i^{-1} \in HUH \subseteq HU,$$

because $H$ lies in the subgroup $HU$, and so $G = HU$. Since $U$ is abelian, $G/H = HU/H \cong U/(H \cap U)$ is abelian, and $H \supseteq G' = G$. Therefore, $G$ is simple. •

Theorem C-1.114 below holds for every infinite field $k$; the proof of Corollary C-1.79 is valid for *perfect fields* $k$ (all finite fields are perfect), but it does not apply to all infinite fields. We now prove the simplicity of $\mathrm{PSL}(V)$ by showing that it satisfies the hypotheses of Iwasawa's Theorem.

**Proposition C-1.113.** *$\mathbb{P}(V)$ is a faithful primitive $\mathrm{PSL}(V)$-set for every vector space $V$ over a field $k$.*

**Proof.** By Theorem C-1.96(iii), $\mathbb{P}(V)$ is a faithful $\mathrm{PSL}(V)$-set.

If $([x],[y])$ and $([x'],[y'])$ are ordered pairs of distinct elements of $\mathbb{P}(V)$, then $x, y$ and $x', y'$ are linearly independent, by Proposition C-1.95(i). Each of these independent lists can be extended to a basis of $V$, say, $x, y, z_3, \ldots, z_n$ and $x', y', z_3', \ldots, z_n'$. There exists $g \in \mathrm{GL}(V)$ with $gx = x'$, $gy = y'$, and $gz_i = z_i'$ for all $i \geq 3$. Hence,

$\mathbb{P}(g)[x] = [x']$ and $\mathbb{P}(g)[y] = [y']$. If $\det(g) = \lambda$, define $h \in \mathrm{GL}(V)$ by $hx = \lambda^{-1}x'$, $hy = y'$, and $hz_i = z_i'$ for all $i \geq 3$. Then $\det(h) = 1$, so that $h \in \mathrm{SL}(V)$, $\mathbb{P}(h)[x] = [\lambda^{-1}x'] = [x']$ and $\mathbb{P}(h)[y] = [y']$. Therefore, $\mathrm{PSL}(V)$ acts doubly transitively on $\mathrm{PSL}(V)$ and hence is primitive, by Theorem C-1.107. •

**Theorem C-1.114.** *For every field $k$, $\mathrm{PSL}(n,k)$ is a simple group if $(n,k) \neq (2,\mathbb{F}_2)$ or $(n,k) \neq (2,\mathbb{F}_3)$.*

**Proof.** We use Iwasawa's Theorem. Let $G = \mathrm{PSL}(V)$, where $V$ is a vector space over a field $k$. By Proposition C-1.113, $\mathbb{P}(V)$ is a faithful doubly transitive, hence primitive, $\mathrm{PSL}(V)$-set.

Choose a nonzero $h$ in $V$ and define a subgroup $U$ of the stabilitzer $G_{[h]}$ by

$$U = \{\mathbb{P}([\varphi, h]) : \varphi(h) = 0\} \cup \{1\}.$$

Applying $\mathbb{P}$ to the formula $[\varphi, h][\psi, h] = [\varphi+\psi, h]$ of Lemma C-1.85(ii), we see that $U$ is abelian. By part (v) of this lemma, if $S \in \mathrm{GL}(V)$, then $S[\varphi, h]S^{-1} = [\varphi S^{-1}, Sh]$; this is true, in particular, for $S \in \mathrm{SL}(V)$. Now $\mathbb{P}(S) \in G_{[h]}$ if and only if $Sh = \lambda h$ for some $\lambda \in k$. But $[\psi, \lambda h] = [\lambda\psi, h]$, by Lemma C-1.85(iii), and this shows that $U \lhd G_{[h]}$.

We now show that the conjugates of $U$ generate $G$; it suffices to show that every $\mathbb{P}([\psi, h'])$ is a conjugate of some $\mathbb{P}([\varphi, h])$ (for $\mathrm{SL}(V)$ is generated by transvections, each of which has the form $[\varphi, h]$). Choose $S \in \mathrm{SL}(V)$ with $Sh = h'$; then, for any $[\varphi, h]$,

$$\mathbb{P}(S)\,\mathbb{P}([\varphi, h])\,\mathbb{P}(S)^{-1} = \mathbb{P}([\varphi S^{-1}, h']),$$

by Lemma C-1.85. As $\varphi$ varies over all linear functionals annihilating $h$, the linear functional $\varphi S^{-1}$ varies over all those annihilating $h' = Sh$. In particular, given $\psi$, there exists $\varphi$ with $\varphi(h) = 0$ and $\psi = \varphi S^{-1}$.

Finally, we show that $G = \mathrm{PSL}(V)$ is perfect. By Proposition C-1.111, the groups $\mathrm{SL}(V)$ are perfect, with two exceptions. But if $G$ is a perfect group, so is any homomorphic image $M$. Let $f\colon G \to M$ be surjective. If $y \in M$, there exists $x \in G$ with $f(x) = y$. Since $G$ is perfect, $x$ is a product of commutators and, hence, $y = f(x)$ is also a product of commutators.

We conclude that $\mathrm{PSL}(V)$ is simple, with two exceptions. •

## Exercises

**C-1.86.** Let $G$ be a finite group, and let $X$ be an $r$-transitive $G$-set of degree $n$.

(i) Assume that $G$ has a normal subgroup $H$ for which there is no $h \neq 1$ in $H$ with $h(x) = x$. Prove that $r \leq 4$.

(ii) If $r = 4$, prove that $G \cong \mathbf{V}$ and $|X| = 4$.

**C-1.87.** Let $X$ be a faithful $r$-transitive $G$-set where $r \geq 4$. Prove that if $G_x$ is simple for some $x \in X$, then $G$ is simple.

**C-1.88.** Use the exercises above to prove that $A_n$ is simple for all $n \geq 5$.

∗ **C-1.89.** If $H$ and $K$ are (not necessarily distinct) subgroups of a group $G$, define an $(H-K)$-***double coset*** to be a subset of $G$ of the form

$$HxK = \{hxk : x \in G, \quad h \in H, \quad k \in K\}.$$

(i) Prove that the family of all $(H-K)$-double cosets is a partition of $G$.

(ii) Generalize Lagrange's Theorem as follows. If $G$ is a finite group and $G = \bigcup_{i=1}^{n} Hx_iK$, prove that

$$[G : K] = \sum_{i=1}^{n} [H : H \cap x_i K x_i^{-1}].$$

∗ **C-1.90.** Let $G = \mathrm{GL}(n, k)$, where $k$ is a field. Define $B$ to be the subgroup of all upper triangular matrices and $H$ the subgroup of all diagonal matrices.

(i) Prove that $G$ is generated by $B \cup N$, where $N = N_G(H)$.

(ii) Prove that $N$ is the subgroup of all ***monomial matrices***, i.e., matrices with exactly one nonzero element in each row and column.

(iii) Prove that $H = B \cap N$ and that $H \lhd N$.

(iv) Prove that $W = N/H$ is generated by $S = \{w_1H, \ldots, w_{n-1}H\}$, where $w_i$ interchanges two adjacent rows of a diagonal matrix.

(v) If $w_iH \in S$ and $wH \in W$ (where $w \in N$), then $w_iBw \subseteq Bw_iwB \cup BwB \subseteq G$.

(vi) No $w_i$ normalizes $B$.

**Remark.** Every group of Lie type contains subgroups $B$ and $N$ which satisfy analogs of the properties in Exercise C-1.90. The ordered pair $(B, N)$ is called a $(B, N)$ ***pair***, and it is used to prove simplicity of these groups. ◀

## C-1.6. Free Groups and Presentations

How can we describe a group? By Cayley's Theorem, a finite group $G$ of order $n$ is isomorphic to a subgroup of the symmetric group $S_n$, and so finite groups can always be defined as subgroups of $S_n$ generated by certain permutations. An example of this kind of construction can be found in Exercise C-1.57 on page 42. Carmichael posed this exercise in the 1930s ([**35**], p. 39), before the era of high-speed computers, and he expected his readers to solve it by hand.

A second way of describing a group $G$ is by replacing $S_n$ with $\mathrm{GL}(n, k)$, where $n \geq 2$ and $k$ is a field. Every group of order $n$ can be imbedded in $\mathrm{GL}(n, k)$ because the group of all $n \times n$ permutation matrices (whose entries are 0's and 1's) is a subgroup of $\mathrm{GL}(n, k)$ isomorphic to $S_n$. A given group $G$ of order $n$ can often be imbedded in $\mathrm{GL}(m, k)$ for $m < n$ if we use entries in $k$ other than 0 and 1; for example, the quaternions $\mathbf{Q}$ can be described as a subgroup of $\mathrm{GL}(2, \mathbb{C})$ of order 8 (page 156 in Part 1). For relatively small groups, descriptions in terms of permutations or matrices are useful, but when $n$ is large, such descriptions are impractical. A group of order 18 can get lost in a group of order $18! \sim 6.4 \times 10^{16}$.

We can also describe groups as being generated by elements subject to certain relations. For example, the dihedral group $D_8$ can be characterized as a group of

order 8 that can be generated by two elements $a$ and $b$ such that $a^4 = 1 = b^2$ and $bab = a^{-1}$. It is necessary to specify the order in this description; if we do not insist that the order is 8, then each of the groups $\mathbf{V}$, $\mathbb{Z}_2$, and $\{1\}$ is also generated by two elements satisfying the relations. We know that the dihedral group $D_8$ exists, for we displayed it in Example C-1.11.

Consider the following definition.

**Definition.** The ***generalized quaternions*** $\mathbf{Q}_n$ is a group of order $2^n$ having generators $a, b$ and relations $a^{2^{n-1}} = 1$, $bab^{-1}a = 1$, $b^{-2}a^{2^{n-2}} = 1$.

If $n = 3$, then $\mathbf{Q}_3$ is the quaternion group $\mathbf{Q}$. Is there a generalized quaternion group of order 16? If $n \geq 4$, does a generalized quaternion group exist? In the 1880s, von Dyck invented *free groups*, the key to answering such questions.

Here is a modern definition of free group, mimicking the freeness property of free modules given in Part 1. That definition, in turn, is modeled on Theorem A-7.28 in Part 1, the fundamental result about bases of vector spaces enabling us to describe linear transformations by matrices.

**Definition.** Let $X$ be a subset[20] of a group $F$, and let $i\colon X \to F$ be the inclusion. Then $F$ is a ***free group*** with ***basis*** $X$ if, for every group $G$ and every function $f\colon X \to G$, there exists a unique homomorphism $\varphi\colon F \to G$ with $\varphi(x) = f(x)$ for all $x \in X$; that is, $\varphi i = f$,

$$\begin{array}{ccc} F & & \\ {\scriptstyle i}\uparrow & \searrow^{\varphi} & \\ X & \xrightarrow[f]{} & G. \end{array}$$

A free group $F$ with basis $X$ may also be denoted by $F(X)$; in particular, if $X = \{x_1, \ldots, x_n\}$, we may denote $F = F(X)$ by $F(x_1, \ldots, x_n)$.

Thus, if $F$ is free with basis $X$ and $G$ is any group, a *function* $f\colon X \to G$ specifies values $f(x) \in G$ for every $x \in X$, and $f$ extends to a unique *homomorphism* $\varphi\colon F \to G$.

For example, suppose a free group $F = F(x, y)$ exists; let $N$ be the normal subgroup of $F$ generated by $\left\{x^{2^{n-1}},\ yxy^{-1}x,\ y^{-2}x^{2^{n-2}}\right\}$. It is clear that $F/N$ is generated by two elements $a = xN$ and $b = yN$ which satisfy the relations defining the generalized quaternion group $\mathbf{Q}_n$. However, it is not clear if $F/N = \mathbf{Q}_n$, for we do not know if $|F/N| = 2^n$ (it is not obvious whether $F/N \neq \{1\}$, nor is it even obvious whether $F/N$ is finite). We will prove, in Proposition C-1.129, that $|F/N| = 2^n$.

## ▪ Existence and Uniqueness of Free Groups

Even though the idea behind a construction of free groups is natural, we will be careful in implementing it.

[20] The subset $X$ may be infinite. When $X$ is finite, it is convenient, as in our earlier discussion of bases of vector spaces and bases of free abelian groups, to define bases as *lists* (ordered sets) in $F$ rather than as mere subsets.

**Definition.** Let $X$ be a set (called an ***alphabet***), let $X^{-1}$ be a set disjoint from $X$, let $x \mapsto x^{-1}$ be a bijection $X \to X^{-1}$, and let 1 be a symbol not in $X \cup X^{-1}$. If $n$ is a positive integer, a ***word*** $w$ ***on*** $X$ is either 1 or an $n$-tuple $w = (x_1^{e_1}, x_2^{e_2}, \dots, x_n^{e_n})$, where $x_i \in X$ for all $i$, $e_i = \pm 1$, and $x_i^1 = x_i$.[21] The symbol 1 is called the ***empty word***. (The empty word 1 is never a "letter" of a word $w = (x_1^{e_1}, \dots, x_n^{e_n})$ because $x_i^{e_i} \in X \cup X^{-1}$ for all $i$.)

The ***length*** of a word $w = (x_1^{e_1}, x_2^{e_2}, \dots, x_n^{e_n})$ is defined to be $n$; the ***length*** of the empty word $w = 1$ is defined to be 0. Denote the length of any word $u$ by

$$|u| = n \geq 0.$$

The ***inverse*** of a nonempty word $w = (x_1^{e_1}, x_2^{e_2}, \dots, x_n^{e_n})$ is

$$w^{-1} = (x_1^{e_1}, x_2^{e_2}, \dots, x_n^{e_n})^{-1} = (x_n^{-e_n}, \dots, x_2^{-e_2}, x_1^{-e_1}).$$

The empty word 1 is defined to be its own inverse.

It follows that $(u^{-1})^{-1} = u$ for every word $u$.

Every word on $X$ has a unique spelling, for an $n$-tuple $w$ and an $m$-tuple $w'$ in a cartesian product of copies of $X \cup X^{-1}$ are equal if and only if $n = m$ and, for each $i$, the $i$th letter $x_i^{e_i}$ of $w$ equals the $i$th letter of $w'$. Notice that $(x, x^{-1})$ is a word of length 2; in particular, it is not the empty word 1, which has length 0.

Recall that a *semigroup* is a set having an associative binary operation, and a *monoid* is a semigroup having a (two-sided) identity element 1.

**Definition.** Given a set $X$, let

$$\mathcal{W}(X)$$

denote the set of all words on the alphabet $X$ (if $X = \varnothing$, then $\mathcal{W}(X)$ consists only of the empty word). Define a binary operation $\odot$ on $\mathcal{W}(X)$, called ***juxtaposition***, as follows. First, define $1 \odot v = v = v \odot 1$ for every word $v$. Next, given nonempty words $u = (x_1^{e_1}, \dots, x_n^{e_n})$ and $v = (y_1^{f_1}, \dots, y_m^{f_m})$, where $y_j \in X$ and $f_j = \pm 1$, define $u \odot v$ by

$$u \odot v = (x_1^{e_1}, \dots, x_n^{e_n}, y_1^{f_1}, \dots, y_m^{f_m}).$$

Note that $|u \odot v| = |u| + |v|$. Juxtaposition is associative: if $w = (z_1^{g_1}, \dots, z_r^{g_r})$, then both $(u \odot v) \odot w$ and $u \odot (v \odot w)$ are $(n+m+r)$-tuples whose $k$th letters are equal for all $k$. Thus, $\mathcal{W}(X)$ is a (noncommutative) monoid whose identity is the empty word.

The monoid $\mathcal{W}(X)$ is our first approximation to a construction of a free group with basis $X$. That $x \odot x^{-1} \neq 1$ shows that $\mathcal{W}(X)$ is not a group.

**Definition.** A ***subword*** of a nonempty word $w = (x_1^{e_1}, \dots, x_n^{e_n}) \in \mathcal{W}(X)$ is either the empty word or a word of the form $u = (x_r^{e_r}, \dots, x_s^{e_s})$, where $1 \leq r \leq s \leq n$.

[21] Everyone denotes words by $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$ instead of $(x_1^{e_1}, x_2^{e_2}, \dots, x_n^{e_n})$, but we shall use $n$-tuples now because I think this is clearer. Later, we shall use the common notation.

Thus, if $(x_r^{e_r}, \dots, x_s^{e_s})$ is a subword of $w$, then $w = A \odot (x_r^{e_r}, \dots, x_s^{e_s}) \odot B$, where $A$ and $B$ are (possibly empty) subwords of $w$.

The most important words are *reduced* words.

**Definition.** A ***pm-word*** is a word of the form $(x^e, x^{-e})$ for some $x \in X$ and $e = \pm 1$ (*pm* abbreviates plus/minus).

A word $w \in \mathcal{W}(X)$ is ***reduced*** if either $w = 1$ or $w$ has no *pm*-subwords.

Note that every subword of a reduced word is itself reduced.

**Lemma C-1.115.** *If $w = (x_1^{e_1}, \dots, x_n^{e_n})$ is a nonempty reduced word, then $w$ has a unique factorization*

$$w = x_1^{e_1} \odot \cdots \odot x_n^{e_n}.$$

**Proof.** That $x_1^{e_1} \odot \cdots \odot x_n^{e_n} = (x_1^{e_1}, \dots, x_n^{e_n})$ is proved by induction on $n \geq 1$; uniqueness holds because the expression $w = (x_1^{e_1}, \dots, x_n^{e_n})$ in $\mathcal{W}(X)$ is unique. •

**Definition.** If $w \in \mathcal{W}(X)$ is not reduced, then $w = A \odot (x^e, x^{-e}) \odot B$, where $x \in X$ and $e = \pm 1$. If $w_1 = A \odot B$, then we say that $w \to w_1$ is an ***elementary cancellation***. If a word $w$ is not reduced, then a ***reduction*** of $w$ is a finite sequence of elementary cancellations

$$w \to w_1 \to \cdots \to w_r$$

with $w_r$ reduced. If $w$ is already reduced, then $w \to w_r$ is a ***reduction***, where $w_r = w$.

For example, there is a reduction $w = (x, x^{-1}) \to 1$; more generally, there is a reduction $w \odot w^{-1} \to \cdots \to 1$. Given an elementary cancellation $w \to w_1$, we have $|w_1| = |w| - 2$, so that $|w_1| < |w|$. However, the number of *pm*-subwords of $w_1$ may not be less than that of $w$: for example, consider $w = u \odot u^{-1}$.

**Lemma C-1.116.** *If $w \in \mathcal{W}(X)$, then either $w$ is reduced or there is a reduction*

$$w \to w_1 \to \cdots \to w_r,$$

*where $w_r$ is reduced.*

**Proof.** We use induction on $|w| \geq 0$. If $w$ is reduced, there is nothing to prove. If $w$ is not reduced and $w \to w_1$ is an elementary cancellation, then $|w_1| = |w| - 2$. By induction, either $w_1$ is reduced or there is a reduction $w_1 \to \cdots \to w_r$, and so $w \to w_1 \to \cdots \to w_r$ is a reduction of $w$. •

A word $w$ that is not reduced can have many reductions.

It is natural to try to define a free group with basis $X$ as the set of all *reduced* words in $\mathcal{W}(X)$, with juxtaposition as the binary operation, but this is not good enough, for this set is not closed: $u$ and $v$ reduced does not imply that $u \odot v$ is reduced. The obvious way to fix this is to change the operation from juxtaposition to juxtaposition followed by a reduction. The following lemma shows that this new binary operation is well-defined.

**Lemma C-1.117.** *Let $X$ be a set and let $w \in \mathcal{W}(X)$. If*

$$w \to w_1 \to w_2 \to \cdots \to w_r \text{ and } w \to w'_1 \to w'_2 \to \cdots \to w'_q$$

*are reductions, then $w_r = w'_q$.*

**Proof.** The proof is by induction on $|w| \geq 0$. The base step $|w| = 0$ is obviously true, for then $w = 1$. For the inductive step $|w| \geq 1$, we may assume that $w$ is not reduced (or we are done). There are two cases for $w_1$ and $w'_1$: the canceled $pm$-subwords are disjoint or they overlap.

In more detail, the first case has $w = A \odot (p, p^{-1}) \odot B \odot (q, q^{-1}) \odot C$ (where $B = 1$ is allowed), and $w_1 = A \odot B \odot (q, q^{-1}) \odot C$, while $w'_1 = A \odot (p, p^{-1}) \odot B \odot C$. If $z = A \odot B \odot C$, then there are elementary cancellations $w_1 \to z$ and $w'_1 \to z$. Choose a reduction $z \to \cdots \to w''_d$. Since $|w_1| < |w|$, the inductive hypothesis applies to the reductions $w_1 \to w_2 \to \cdots \to w_r$ and $w_1 \to z \to \cdots \to w''_d$, and so $w_r = w''_d$. Similarly, the inductive hypothesis applies to $w'_1 \to z \to \cdots \to w''_d$ and $w'_1 \to w'_2 \to \cdots \to w'_q$, so that $w''_d = w'_q$. Therefore, $w_r = w'_q$.

The second case has $w = A \odot (p^e, p^{-e}) \odot p^e \odot B$, and $w_1 = A \odot p^e \odot B$ (we have canceled $(p^e, p^{-e})$), while $w'_1 = A \odot p^e \odot B$ (we have canceled $(p^{-e}, p^e)$). Thus, we have $w_1 = w'_1$ here. Hence, the reduction $w'_1 \to w'_2 \to \cdots \to w'_q$ is $w_1 \to w'_2 \to \cdots \to w'_q$. Comparing this reduction with $w_1 \to w_2 \to \cdots \to w_r$ gives $w'_q = w_r$. •

In light of Lemma C-1.117, all reductions of a given word $w \in \mathcal{W}(X)$ end with the same reduced word, say, $w_r$. We denote this reduced word by

$$\text{red}(w) = w_r.$$

**Corollary C-1.118.** *If $F(X)$ is the set of all reduced words on $X$, then*

$$uv = \text{red}(u \odot v)$$

*is a well-defined binary operation on $F(X)$.*

**Proof.** Immediate from Lemmas C-1.116 and C-1.117. •

If $u, v$ are reduced words for which $u \odot v$ is also reduced, then $uv = u \odot v$. It follows that if $u = (x_1^{e_1}, \ldots, x_n^{e_n})$ is reduced, then $u = x_1^{e_1} \cdots x_n^{e_n}$. In other words, since $F(X) \subseteq \mathcal{W}(X)$, its elements are $n$-tuples $(x_1^{e_1}, \ldots, x_n^{e_n})$. But having just introduced the binary operation $\text{red}(\odot)$ on $F(X)$, we are now allowed to use the simpler notation $x_1^{e_1} \cdots x_n^{e_n}$.

**Theorem C-1.119.** *If $X$ is a set, then the set $F(X)$ of all reduced words on $X$ with binary operation $uv = \text{red}(u \odot v)$ is a free group with basis $X$.*

**Proof.** It is easy to see that the empty word 1 is the identity element and that the inverses defined on page 83 satisfy the group axiom for inverses. Only associativity need be checked. Given reduced words $u, v, y$, define $w = u \odot v \odot y$ in $\mathcal{W}(X)$ (we do not need parentheses, for $\mathcal{W}(X)$ is a monoid and its multiplication is associative). But Lemma C-1.117 says that the reductions $w = (u \odot v) \odot y \to \cdots \to (uv) \odot y \to$

$\cdots \to w_r$ and $w = u \odot (v \odot y) \to \cdots \to u \odot (vy) \to \cdots \to w'_q$ have the same reduced ending: $w_r = w'_q$; that is, $(uv)y = u(vy)$. Therefore, $F(X)$ is a group.

Let $G$ be a group and let $f\colon X \to G$ be a function. If $u \in F(X)$, then $u$ is reduced, and it has a *unique* expression $u = x_1^{e_1} \cdots x_n^{e_n}$, by Lemma C-1.115. Define a function $\varphi\colon F(X) \to G$ by $\varphi(1) = 1$ and

$$\varphi(u) = \varphi(x_1^{e_1} \cdots x_n^{e_n}) = f(x_1)^{e_1} \cdots f(x_n)^{e_n}.$$

It suffices to prove that $\varphi$ is a homomorphism, for uniqueness of $\varphi$ will follow from $X$ generating $F(X)$. If $u, v \in F(X)$, we prove that $\varphi(uv) = \varphi(u)\varphi(v)$ by induction on $|u| + |v| \geq 0$. The base step $|u| + |v| = 0$ is true: if $u = 1 = v$, then $uv = 1$ and $\varphi(uv) = \varphi(1) = 1$; hence, $\varphi(uv) = 1 = \varphi(u)\varphi(v)$ in this case. In fact, $\varphi(1v) = \varphi(1)\varphi(v)$, so that we may assume that $|u| \geq 1$ in proving the inductive step. Write the reduced word $u = x_1^{e_1} u'$, where $u' = x_2^{e_2} \cdots x_n^{e_n}$, and note, since $u$ is reduced, that

$$\varphi(u) = f(x_1)^{e_1} f(x_2)^{e_2} \cdots f(x_n)^{e_n} = \varphi(x_1^{e_1})\varphi(u').$$

Now $uv = x_1^{e_1}(u'v)$; write the reduced word $u'v = z_1^{c_1} \cdots z_t^{c_t}$, where $z_1, \ldots, z_t \in X$. There are two cases.

If $x_1^{e_1} \neq z_1^{-c_1}$, then $x_1^{e_1} z_1^{c_1} \cdots z_t^{c_t}$ is reduced, and the formula defining $\varphi$ gives

$$\begin{aligned}\varphi(uv) &= \varphi(x_1^{e_1} u'v) = \varphi(x_1^{e_1} z_1^{c_1} \cdots z_t^{c_t}) \\ &= f(x_1)^{e_1} f(z_1)^{c_1} \cdots f(z_t)^{c_t} = \varphi(x_1^{e_1})\varphi(u'v) \\ &= \varphi(x_1^{e_1})\varphi(u')\varphi(v) = \varphi(u)\varphi(v)\end{aligned}$$

(the inductive hypothesis gives the penultimate equality).

If $x_1^{e_1} = z_1^{-c_1}$, then $uv = x_1^{e_1} u'v = z_2^{c_2} \cdots z_t^{c_t}$. Hence,

$$\begin{aligned}\varphi(uv) &= \varphi(x_1^{e_1} u'v) = \varphi(z_2^{c_2} \cdots z_t^{c_t}) \\ &= f(z_2)^{c_2} \cdots f(z_t)^{c_t} = [f(x_1)^{e_1} f(x_1)^{e_1}] f(z_2)^{c_2} \cdots f(z_t)^{c_t} \\ &= \varphi(x_1^{e_1})\varphi(u'v) = \varphi(x^e)\varphi(u')\varphi(v) = \varphi(u)\varphi(v)\end{aligned}$$

because $f(x_1)^{e_1} f(z_1)^{c_1} = 1$. Therefore, $\varphi$ is a homomorphism, and $F(X)$ is a free group with basis $X$. •

**Theorem C-1.120.** *For every group $G$, there is a free group $F$ with $G$ isomorphic to a quotient of $F$.*

**Proof.** View the group $G$ as a set, let $F$ be the free group with basis $G$ (whose existence has just been proved), and let $f = 1_G\colon G \to G$. Now view $G$ as a group; there is a homomorphism $\varphi\colon F \to G$ extending $f$. Clearly, $\varphi$ is surjective, and so $G \cong F/\ker\varphi$. •

We may interpret Theorem C-1.120 as saying that every group is encoded inside some free group.

Informally, we say that a group $G$ is free with basis $X$ if there are no relations among the reduced words on $X$. Here is the formal statement.

**Corollary C-1.121.** *Let $G$ be a group generated by a subset $X$. If every reduced word $g = x_1^{e_1} \cdots x_n^{e_n}$, where $x_i \in X$, $e_i = \pm 1$, and $n \geq 1$, is not equal to 1, then $G$ is free with basis $X$.*

**Proof.** Let $F$ be the free group with basis $Y$, where $Y$ is a set for which there exists a bijection $f\colon Y \to X$,

$$\begin{array}{ccc} F & \overset{\varphi}{\longrightarrow} & G \\ \uparrow & & \uparrow{\scriptstyle i} \\ Y & \xrightarrow[f]{} & X. \end{array}$$

Since $F$ is free with basis $Y$, there is a (unique) homomorphism $\varphi\colon F \to G$ with $\varphi(y) = if(x)$ for all $y \in Y$, where $i\colon X \to G$ is the inclusion. Now $\varphi$ is surjective, because $\varphi(Y) = X$ generates $G$; moreover, $\varphi$ is injective, for if $h = y_1^{e_1} \cdots y_n^{e_n}$ is a nontrivial reduced word on $Y$, then $h \notin \ker \varphi$ because, by hypothesis, $\varphi(h) = \varphi(y_1^{e_1} \cdots y_n^{e_n}) = x_1^{e_1} \cdots x_n^{e_n} \neq 1$. Therefore, $\varphi$ is an isomorphism and $G$ is free with basis $\varphi(Y) = X$. •

Here are sketches of some other proofs of the existence of free groups.

(i) There is a shorter proof of the existence of the free group with basis a given set $X$, due to Barr (see Montgomery–Ralston [**158**], pp. 2–5)). The Adjoint Functor Theorem gives a necessary and sufficient condition that a functor have an adjoint (see Mac Lane [**144**], pp. 116–119); in particular, the left adjoint of a forgetful functor creates free objects (see Exercise C-1.96 on page 91). We have not given this proof here because it does not describe the elements of $F(X)$ as words on $X$, and this description is essential in studying and using free groups.

(ii) Another construction is called the ***van der Waerden trick*** (see Rotman [**188**], p. 345). Let $R$ be the set of all reduced words on $X$ (of course, this is the underlying set of $F(X)$). For each $x \in X$, consider the functions $[x]\colon R \to R$ and $[x^{-1}]\colon R \to R$, defined as follows. If $\epsilon = \pm 1$, then

$$\begin{aligned} [x^\epsilon](x_1^{e_1}, \ldots, x_n^{e_n}) &= (x^\epsilon, x_1^{e_1}, \ldots, x_n^{e_n}) \quad \text{if } x^\epsilon \neq x_1^{e_1}, \\ [x^{-\epsilon}](x_1^{e_1}, \ldots, x_n^{e_n}) &= (x_2^{e_2}, \ldots, x_n^{e_n}) \qquad \text{if } x^\epsilon = x_1^{e_1}. \end{aligned}$$

It turns out that $[x]$ is a permutation of $R$ (its inverse is $[x^{-1}]$), and the subgroup $F$ of the symmetric group $S_R$ generated by $[X] = \{[x] : x \in X\}$ is a free group with basis $[X]$.

(iii) There are topological proofs. A ***pointed space*** is an ordered pair $(X, w)$, where $X$ is a topological space and $w \in X$ (see Example B-4.15 in Part 1); we call $w$ the ***basepoint***. A ***pointed map*** $f\colon (X, w) \to (Y, w')$ is a continuous map $f\colon X \to Y$ with $f(w) = w'$. The ***fundamental group*** $\pi_1(X, w)$ is the set of all (pointed) homotopy classes of pointed maps $(S^1, 1) \to (X, w)$, where $S^1$ is the unit circle $\{e^{2\pi i x} : x \in \mathbb{R}\}$ with basepoint $1 = e^0$. Given an indexed family of circles, $(S_i^1, w)_{i \in I}$, any two intersecting only in their common basepoint $w$, then their union $B_I$ (suitably topologized) is called a

***bouquet of** $|I|$ **circles***. For example, a figure 8 is a bouquet of two circles. Then the fundamental group $\pi_1(B_I, w)$ is a free group with basis a set of cardinality $|I|$ (Exercise C-1.116 on page 110).

(iv) If $X$ is a ***graph*** (a one-dimensional space constructed of edges and vertices), then $\pi_1(X, w)$ is also a free group (Serre [**202**], p. 23, where it is shown that every connected graph has the homotopy type of a bouquet of circles).

The free group $F = F(X)$ with basis $X$ that we constructed in Theorem C-1.119 is generated by $X$, but we have just observed that there are other constructions of free groups. Now $F(X)$ is generated by the basis $X$; does every basis of a free group $F$ generate $F$? Are any two free groups with basis $X$ isomorphic?

**Proposition C-1.122.**

(i) *Let $X$ be a basis of a free group $F$ and let $X^*$ be a basis of a free group $F^*$. If there is a bijection $f\colon X \to X^*$, then $F \cong F^*$; indeed, there is an isomorphism $\varphi\colon F \to F^*$ extending $f$.*

(ii) *If $F$ is a free group with basis $X$, then $F$ is generated by $X$.*

**Proof.**

(i) The following diagram, in which the vertical arrows are inclusions, will help the reader follow the proof:

$$\begin{array}{ccc} F & \underset{\varphi^*}{\overset{\varphi}{\rightleftarrows}} & F^* \\ \uparrow & & \uparrow \\ X & \underset{f^{-1}}{\overset{f}{\rightleftarrows}} & X^*. \end{array}$$

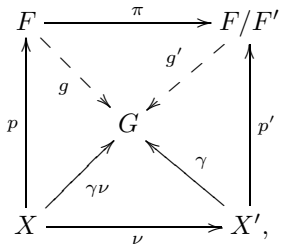
We may regard $f$ as having target $F^*$, because $X^* \subseteq F^*$; since $F$ is a free group with basis $X$, there is a homomorphism $\varphi\colon F \to F^*$ extending $f$. Similarly, there exists a homomorphism $\varphi^*\colon F^* \to F$ extending $f^{-1}$. It follows that the composite $\varphi^*\varphi\colon F \to F$ extends $1_X$. But the identity $1_F$ also extends $1_X$, so that uniqueness of the extension gives $\varphi^*\varphi = 1_F$. In the same way, we see that the other composite $\varphi\varphi^* = 1_{F^*}$, and so $\varphi$ is an isomorphism.

(ii) If $F(X)$ is the free group with basis $X$ constructed in Theorem C-1.119, then $X$ generates $F(X)$. By part (i), there is an isomorphism $\varphi\colon F(X) \to F$ with $\varphi(X) = X$ (take $f\colon X \to X$ to be the identity $1_X$). Since $X$ generates $F(X)$, $\varphi(X) = X$ generates $\operatorname{im}\varphi = F$; i.e., $X$ generates $F$. •

There is a notion of rank for free groups (as there is for free abelian groups), but we must first check that all bases of a free group have the same number of elements (which might be an infinite cardinal).

**Lemma C-1.123.** *If $F$ is a free group with basis $X$, then $F/F'$ is a free abelian group with basis $X' = \{xF' : x \in X\}$, where $F'$ is the commutator subgroup of $F$.*

**Proof.** We begin by noting that $X'$ generates $F/F'$; this follows from Proposition C-1.122(ii), which says that $X$ generates $F$. We prove that $F/F'$ is a free abelian group with basis $X'$: given an abelian group $G$ and a function $\gamma\colon X' \to G$, we show there is a (necessarily unique) homomorphism $g'\colon F/F' \to G$ extending $\gamma$. Consider the following diagram:




where $p$ and $p'$ are inclusions, $\pi$ is the natural map, and $\nu\colon x \mapsto xF'$. Let $g\colon F \to G$ be the unique homomorphism with $gp = \gamma\nu$ given by the definition of free group (for $\gamma\nu\colon X \to G$ is a function), and define $g'\colon F/F' \to G$ by $wF' \mapsto g(w)$ ($g'$ is well-defined because $G$ abelian forces $F' \subseteq \ker g$). Now $g'p' = \gamma$, for

$$g'p'\nu = g'\pi p = gp = \gamma\nu;$$

since $\nu$ is a surjection, it follows that $g'p' = \gamma$. Finally, $g'$ is the unique such map, for if $g''$ satisfies $g''p' = \gamma$, then $g'$ and $g''$ agree on the generating set $X'$, and hence they are equal. •

**Proposition C-1.124.** *Let $F$ be a free group with basis $X$. If $|X| = n$, then every basis of $F$ has $n$ elements.*

**Proof.** By Lemma C-1.123, $F/F'$ is a free abelian group of rank $n$. On the other hand, if $Y$ is a basis of $F$ and $|Y| = m$, then $F/F'$ is a free abelian group of rank $m$. By Corollary B-2.23 in Part 1, we have $m = n$. •

The reader may show, using Theorem B-2.13 in Part 1, that Proposition C-1.124 is true even if a basis is infinite: any two bases of a free group have the same cardinal. The following definition now makes sense.

**Definition.** The ***rank*** of a free group $F$, denoted by $\operatorname{rank}(F)$, is the number of elements in a basis of $F$.

Proposition C-1.122(i) can now be rephrased: two free groups are isomorphic if and only if they have the same rank. A free group $F$ of finite rank $n$ is often denoted by $F_n$; if $X = \{x_1, \ldots, x_n\}$ is a basis of $F$, we may also write $F_n(X)$ or $F(x_1, \ldots, x_n)$.

We are now going to prove that there is a rich supply of homomorphisms from free groups to finite groups. Afterward, we will give an example of the construction of a homomorphism in its proof.

**Proposition C-1.125.** *If $F$ is a free group, $g \in F$, and $g \neq 1$, then there is a finite group $K$ and a homomorphism $\varphi\colon F \to K$ with $\varphi(g) \neq 1$.*

**Proof.** Let $X$ be a basis of $F$, and write $g$ as a reduced word:

$$g = x_{i_1}^{e_1} \cdots x_{i_n}^{e_n},$$

where $x_{i_k} \in X$ and $e_k = \pm 1$ for $k = 1, \ldots, n$. There are $m \leq n$ distinct basis elements occurring in this word, say, $x_1, \ldots, x_j, \ldots, x_m$ (thus, each $x_j$ occurs at least once as some $x_{i_k}$). For each such basis element $x_j$, we are going to construct a permutation $\alpha_j \in S_{n+1}$. Consider the set of all positions $k$ where $x_j$ or $x_j^{-1}$ occurs. If $x_j$ occurs, define $\alpha_j(k) = k + 1$; if $x_i^{-1}$ occurs, define $\alpha_j(k + 1) = k$; that is, $\alpha_j^{-1}(k) = k+1$. In other words, $\alpha_j^{e_k}(k) = \alpha_{i_k}^{e_k}(k) = k+1$ for all $k$. We have defined $\alpha_j$ on a subset of $\{1, 2, \ldots, n+1\}$. Exercise A-4.2 on page 122 in Part 1 shows that each $\alpha_j$ can be completed to a permutation in $S_{n+1}$.

Since $F$ is free with basis $X$, we can define a homomorphism $\varphi\colon F \to S_{n+1}$ by specifying its values on the basis $X$. If $x \in X$ is an $x_j$ occurring in the spelling of $g$, define $\varphi(x_j) = \alpha_j$; if $x$ is not involved in $g$, define $\varphi(x) = 1$. We have

$$\varphi(g) = \alpha_{i_1}^{e_1} \cdots \alpha_{i_n}^{e_n}.$$

To apply $\varphi(g)$, remember that writing functions on the left forces us to evaluate a composite $\alpha\beta$ of two functions on some $k$ by $(\alpha\beta)(k) = \beta(\alpha(k))$. Thus,

$$\varphi(g)(1) = \alpha_{i_n}^{e_n} \cdots \alpha_{i_1}^{e_1}(1).$$

But each $\alpha_j^{e_k}(k) = k + 1$, so that $\varphi(g)\colon 1 \mapsto n + 1$. Therefore, $\varphi(g) \neq (1)$. •

We now illustrate the construction in the proof just given. Let $F = F(X)$ be the free group with basis $X$ and let

$$g = z^{-1}yyz^{-1}z^{-1}y^{-1}zy,$$

where $z, y \in X$ are distinct. Label the 8 positions from right to left: $z$ or $z^{-1}$ occurs in positions 2, 4, 5, 8 and $y$ or $y^{-1}$ occurs in positions 1, 3, 6, 7. Construct the permutations $\alpha_z, \alpha_y \in S_9$ in two stages. Here is the first stage for $\alpha_z$:

$$\alpha_z = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ & 3 & & & 4 & 5 & & & 8 \end{pmatrix}.$$

The second stage uses Exercise A-4.2 on page 122 in Part 1: we can assign values to the other integers between 1 and 9 to get a permutation in $S_9$. Here is the analogous construction for $\alpha_y$; its first stage is

$$\alpha_y = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & & & 3 & & 7 & 8 & & \end{pmatrix}.$$

Similarly, $\alpha_y$ can be completed to a permutation in $S_9$. The composite

$$\varphi(g) = \alpha_z^{-1}\alpha_y\alpha_y\alpha_z^{-1}\alpha_z^{-1}\alpha_y^{-1}\alpha_z\alpha_y$$

sends $1 \mapsto 2 \mapsto 3 \mapsto 4 \mapsto 5 \mapsto 6 \mapsto 7 \mapsto 8 \mapsto 9$, and so $\varphi(g) \neq (1)$.

**Definition.** A group $G$ is ***residually finite*** if the intersection of all its normal subgroups of finite index is trivial.

Residually finite groups arose (tacitly) in our discussion of infinite Galois extensions in Part 1. Given a group $G$, we defined its completion $\widehat{G}$ as an inverse limit $\varprojlim G/N$, where $N$ ranges over all normal subgroups of finite index. There is a map $G \to \widehat{G}$, and it is injective when $G$ is residually finite (see Exercise C-1.65 on page 42).

Every finite group is residually finite, but the additive group of rationals $\mathbb{Q}$ is not.

**Corollary C-1.126.** *Every free group is residually finite.*

**Proof.** It suffices to prove that if $g \neq 1$, then there is some normal subgroup $N$ of finite index with $g \notin N$. By Proposition C-1.125, there is a finite group $K$ and a homomorphism $\varphi\colon F \to K$ with $\varphi(g) \neq 1$. Hence, $g \notin \ker \varphi$. But $K$ finite forces $N = \ker \varphi$ to have finite index. •

## Exercises

∗ **C-1.91.** Let $F$ be a free group with basis $X$ and let $A \subseteq X$. Prove that if $N$ is the normal subgroup of $F$ generated by $A$, then $F/N$ is a free group.

∗ **C-1.92.** Let $F$ be a free group.

(i) Prove that $F$ has no elements of finite order (other than 1).

(ii) Prove that a free group $F$ is abelian if and only if $\operatorname{rank}(F) \leq 1$.
**Hint.** Map a free group of rank $\geq 2$ onto a nonabelian group.

(iii) Prove that if $\operatorname{rank}(F) \geq 2$, then $Z(F) = \{1\}$, where $Z(F)$ is the center of $F$.

**C-1.93.** Prove that a free group is solvable if and only if it is infinite cyclic.

**C-1.94.** If $G$ is a finitely generated group and $n$ is a positive integer, prove that $G$ has only finitely many subgroups of index $n$.

**Hint.** Consider homomorphisms $G \to S_n$.

**C-1.95.** (i) Prove that each of the generalized quaternion groups $\mathbf{Q}_n$ has a unique subgroup of order 2, namely, $\langle b^2 \rangle$, and this subgroup is the center $Z(\mathbf{Q}_n)$.

(ii) Prove that $\mathbf{Q}_n/Z(\mathbf{Q}_n) \cong D_{2^{n-1}}$.

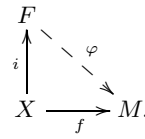
∗ **C-1.96.** (i) Prove that the forgetful functor $U\colon$ **Groups** $\to$ **Sets** (see Example B-4.15 in Part 1) is a functor.

(ii) Prove that $F\colon$ **Sets** $\to$ **Groups**, defined as follows, is a functor: $F\colon X \mapsto F(X)$ (here $X$ is a set and $F(X)$ is the free group with basis $X$); $F\colon f \mapsto \varphi$ (here $f\colon X \to Y$ is a function in **Sets** and $\varphi\colon F(X) \to F(Y)$ is the homomorphism determined by the function $X \to Y \to F(Y)$, where $Y \to F(Y)$ is the inclusion).

(iii) Prove that $(F, U)$ is an adjoint pair.

**C-1.97.** (i) If $X$ is a set, a word $w$ is ***positive*** if $w = 1$ or $w = (x_1^{e_1}, \ldots, x_n^{e_n})$, where all $e_i = 1$. Prove that the set $\mathcal{P}(X)$ of all positive words on $X$ is a submonoid of $\mathcal{W}(X)$.

(ii) Let $X$ be a subset of a monoid $F$, and let $i\colon X \to F$ be the inclusion. Then $F$ is a ***free monoid*** with ***basis*** $X$ if, for every monoid $M$ and every function $f\colon X \to M$, there exists a unique homomorphism $\varphi\colon F \to M$ with $\varphi(x) = f(x)$ for all $x \in X$,




Prove that $\mathcal{P}(X)$ is the free monoid with basis $X$.

## ▪ Presentations

Let us return to describing groups.

**Definition.** Let $X$ be a set, $F = F(X)$ the free group with basis $X$, and $R \subseteq F$ a set of words on $X$. A group $G$ has a ***presentation***,

$$G = (X \mid R),$$

if $G \cong F/N$, where $N$ is the normal subgroup of $F$ generated by $R$; that is, $N$ is the subgroup of $F$ generated by all conjugates of elements of $R$.

We call the set $X$ ***generators*** and the set $R$ ***relations***. (The term *generators* is now being used in a generalized sense, for $X$ is not a subset of $G$; the subset $\{xN\colon x \in X\}$ does generate $F/N$ in the usual sense.)

Theorem C-1.120 says that every group has a presentation.

**Example C-1.127.**

(i) A group has many presentations. For example, $G = \mathbb{Z}_6$ has presentations

$$(x \mid x^6) \quad \text{and} \quad (a, b \mid a^3, b^2, aba^{-1}b^{-1}).$$

This means that there are isomorphisms $\mathbb{Z}_6 \cong F(x)/\langle x^6 \rangle$ and $\mathbb{Z}_6 \cong F(a,b)/N$, where $N$ is the normal subgroup generated by $a^3, b^2, aba^{-1}b^{-1}$. The relation $aba^{-1}b^{-1}$ says that $a$ and $b$ commute. If we replace this commutator by $abab$, then we have a presentation of $S_3$, for now we have $bab = a^{-1}$. If we delete this relation, we obtain a presentation of the infinite *modular group* $M$ defined in Exercise A-4.88 on page 173 in Part 1.

(ii) The free group with basis $X$ has a presentation

$$(X \mid \varnothing).$$

A free group is so called precisely because it has a presentation with no relations (see Corollary C-1.121). ◀

A word on notation. Often, we write the relations in a presentation as equations. Thus, the relations

$$a^3, \quad b^2, \quad aba^{-1}b^{-1}$$

in the second presentation of $\mathbb{Z}_6$ may also be written as

$$a^3 = 1, \quad b^2 = 1, \quad ab = ba.$$

**Definition.** A group $G$ is ***finitely generated*** if it has a presentation $(X \mid R)$ with $X$ finite. A group $G$ is called ***finitely presented*** if it has a presentation $(X \mid R)$ in which both $X$ and $R$ are finite.

It is easy to see that a group $G$ is finitely generated if and only if there exists a finite subset $A \subseteq G$ with $G = \langle A \rangle$. Of course, every finitely generated free group is finitely presented. There do exist finitely generated groups that are not finitely presented (Rotman [**188**], p. 417).

A fundamental problem is how to determine whether two presentations give isomorphic groups. It can be proved that no algorithm can exist that solves this problem. Indeed, it is an undecidable problem whether a presentation defines the (trivial) group of order 1 (Rotman [**188**], p. 469).

**Definition.** Let $F$ be the free group with basis $X$, and let $R \subseteq F$. A group $G$ is of ***type*** $\mathbb{T}(X \mid R)$ if there is a surjective homomorphism $\varphi\colon F \to G$ with $\varphi(r) = 1$ for all $r \in R$.

For example, recall that the dihedral group $D_8$ is a group of order 8 that can be generated by two elements $a$ and $b$ such that $a^4 = 1 = b^2$ and $bab = a^{-1}$. Thus, $D_8$ has type $\mathbb{T}(a, b \mid a^4 = 1 = b^2, bab = a^{-1})$. Note, in saying $D_8$ has this type, that we have not mentioned that $D_8$ has order 8. More generally, a group $G$ with presentation $G = (X \mid R)$ obviously has type $\mathbb{T}(X \mid R)$, but the converse is false. For example, the trivial group $\{1\}$ has type $\mathbb{T}(X \mid R)$ for every ordered pair $(X \mid R)$.

Here is the connection between presentations and types.

**Theorem C-1.128 (von Dyck's Theorem).**

(i) *If groups $G$ and $H$ have presentations*

$$G = (X \mid R) \quad \text{and} \quad H = (X \mid R \cup S),$$

*then $H$ is a quotient of $G$. In particular, if $H$ is a group of type $\mathbb{T}(X \mid R)$, then $H$ is a quotient of $G$.*

(ii) *Let $G = (X \mid R)$ and let $H$ be a group of type $\mathbb{T}(X \mid R)$. If $G$ is finite and $|G| = |H|$, then $G \cong H$.*

**Proof.**

(i) Let $F$ be the free group with basis $X$. If $N$ is the normal subgroup of $F$ generated by $R$ and $K$ is the normal subgroup generated by $R \cup S$, then $N \subseteq K$. Recall the proof of the Third Isomorphism Theorem: the function $\psi\colon F/N \to F/K$, given by $\psi\colon fN \mapsto fK$, is a surjective homomorphism (with $\ker \psi = K/N$, so that $(F/N)/(K/N) \cong F/K$); that is, $H = F/K$ is a quotient of $G = F/N$. In particular, to say that $H$ has type $\mathbb{T}(X \mid R)$ is to say that it satisfies all the relations holding in $G$.

(ii) Since $G$ is finite, the Pigeonhole Principle says that the surjective homomorphism $\psi\colon G \to H$ in part (i) is an isomorphism. •

Note that if $G = (X \mid R)$ is a finite group, then von Dyck's Theorem implies that $|G| \geq |H|$ for every group $H$ of type $\mathbb{T}(X \mid R)$.

**Proposition C-1.129.** *For every $n \geq 3$, the generalized quaternion group $\mathbf{Q}_n$ exists: the group with presentation*

$$\mathbf{Q}_n = \big(a, b \mid a^{2^{n-1}} = 1, bab^{-1} = a^{-1}, b^2 = a^{2^{n-2}}\big)$$

*has order $2^n$.*

**Proof.** The cyclic subgroup $\langle a \rangle$ in $\mathbf{Q}_n$ has order at most $2^{n-1}$, because $a^{2^{n-1}} = 1$. The relation $bab^{-1} = a^{-1}$ implies that $\langle a \rangle \lhd \mathbf{Q}_n = \langle a, b \rangle$, so that $\mathbf{Q}_n/\langle a \rangle$ is generated by the image of $b$. Finally, the relation $b^2 = a^{2^{n-2}}$ shows that $|\mathbf{Q}_n/\langle a \rangle| \leq 2$. Hence,

$$|\mathbf{Q}_n| \leq |\langle a \rangle||\mathbf{Q}_n/\langle a \rangle| \leq 2^{n-1} \cdot 2 = 2^n.$$

We prove the reverse inequality by constructing a concrete group $H_n$ of type $\mathbb{T}(x, y \mid x^{2^{n-1}}, yxy^{-1}x, y^{-2}x^{2^{n-2}})$. Consider the complex matrices $A = \left[\begin{smallmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$, where $\omega$ is a primitive $2^{n-1}$th root of unity, and let $H_n = \langle A, B \rangle \subseteq \mathrm{GL}(2, \mathbb{C})$. We claim that $A$ and $B$ satisfy the necessary relations. For all $i \geq 1$,

$$A^{2^i} = \begin{bmatrix} \omega^{2^i} & 0 \\ 0 & \omega^{-2^i} \end{bmatrix},$$

so that $A^{2^{n-1}} = I$; indeed, $A$ has order $2^{n-1}$. Moreover, $B^2 = \left[\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right] = A^{2^{n-2}}$ and $BAB^{-1} = \left[\begin{smallmatrix} \omega^{-1} & 0 \\ 0 & \omega \end{smallmatrix}\right] = A^{-1}$. Notice that $A$ and $B$ do not commute; hence, $B \notin \langle A \rangle$, and so the cosets $\langle A \rangle$ and $B\langle A \rangle$ are distinct. Since $A$ has order $2^{n-1}$, it follows that

$$|H_n| \geq |\langle A \rangle \cup B\langle A \rangle| = 2^{n-1} + 2^{n-1} = 2^n.$$

By von Dyck's Theorem, $2^n \leq |H_n| \leq |\mathbf{Q}_n| \leq 2^n$. Therefore, $|\mathbf{Q}_n| = 2^n$, and Theorem C-1.128(ii) gives $\mathbf{Q}_n \cong H_n$. •

In Exercise A-4.66 on page 159 in Part 1, we gave a concrete construction of the dihedral group $D_{2n}$, and we can use that group—as in the last proof—to give a presentation of it.

**Proposition C-1.130.** *The dihedral group $D_{2n}$ has a presentation*

$$D_{2n} = (a, b \mid a^n = 1, b^2 = 1, bab = a^{-1}).$$

**Proof.** Let $D_{2n}$ denote the group defined by the presentation, and let $C_{2n}$ be the group of order $2n$ constructed in Exercise A-4.66 on page 159 in Part 1. By von Dyck's Theorem, $|D_{2n}| \geq |C_{2n}| = 2n$. We prove the reverse inequality. The cyclic subgroup $\langle a \rangle$ in $D_{2n}$ has order at most $n$, because $a^n = 1$. The relation $bab^{-1} = a^{-1}$ implies that $\langle a \rangle \lhd D_{2n} = \langle a, b \rangle$, so that $D_{2n}/\langle a \rangle$ is generated by the image of $b$. Finally, the relation $b^2 = 1$ shows that $|D_{2n}/\langle a \rangle| \leq 2$. Hence, $|D_{2n}| \leq |\langle a \rangle||D_{2n}/\langle a \rangle| \leq 2n$, and $|D_{2n}| = 2n$. Therefore, Theorem C-1.128(ii) gives $D_{2n} \cong C_{2n}$. •

In Section C-1.1, we classified the groups of order 7 or less. Since groups of prime order are cyclic, it was only a question of classifying the groups of orders 4 and 6. The proof we gave, in Proposition C-1.4, that every nonabelian group of

order 6 is isomorphic to $S_3$ was rather complicated, analyzing the representation of a group on the cosets of a cyclic subgroup. Here is a proof in the present spirit.

**Proposition C-1.131.** *If $G$ is a nonabelian group of order 6, then $G \cong S_3$.*

**Proof.** As in the proof of Proposition C-1.4, $G$ must contain elements $a$ and $b$ of orders 3 and 2, respectively. Now $\langle a \rangle \lhd G$, because it has index 2, and so either $bab^{-1} = a$ or $bab^{-1} = a^{-1}$. The first possibility cannot occur, because $G$ is not abelian. Therefore, $G$ has type $\mathbb{T}(a, b \mid a^3, b^2, bab = a^{-1})$, and so Theorem C-1.128(ii) gives $D_6 \cong G$ (of course, $D_6 \cong S_3$). •

We can now classify the groups of order 8.

**Theorem C-1.132.** *Every group $G$ of order 8 is isomorphic to*

$$D_8, \quad \mathbf{Q}, \quad \mathbb{Z}_8, \quad \mathbb{Z}_4 \oplus \mathbb{Z}_2, \quad \text{or} \quad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2.$$

*Moreover, no two of the displayed groups are isomorphic.*

**Proof.** If $G$ is abelian, then the Basis Theorem shows that $G$ is a direct sum of cyclic groups, and the Fundamental Theorem shows that the only such groups are those listed. Therefore, we may assume that $G$ is not abelian.

Now $G$ cannot have an element of order 8, lest it be cyclic, hence abelian; moreover, not every nonidentity element can have order 2, lest $G$ be abelian, by Exercise A-4.31 on page 138 in Part 1. We conclude that $G$ must have an element $a$ of order 4; hence, $\langle a \rangle$ has index 2, and so $\langle a \rangle \lhd G$. Choose $b \in G$ with $b \notin \langle a \rangle$; note that $G = \langle a, b \rangle$ because $\langle a \rangle$, having index 2, must be a maximal subgroup. Now $b^2 \in \langle a \rangle$, because $G/\langle a \rangle$ is a group of order 2, and so $b^2 = a^i$, where $0 \leq i \leq 3$. We cannot have $b^2 = a$ or $b^2 = a^3 = a^{-1}$ lest $b$ have order 8. Therefore, either

$$b^2 = a^2 \quad \text{or} \quad b^2 = 1.$$

Furthermore, $bab^{-1} \in \langle a \rangle$, by normality, and so $bab^{-1} = a$ or $bab^{-1} = a^{-1}$ (for $bab^{-1}$ has the same order as $a$). But $bab^{-1} = a$ says that $a$ and $b$ commute, which implies that $G$ is abelian. We conclude that $bab^{-1} = a^{-1}$. Therefore, there are only two possibilities:

$$a^4 = 1, \ b^2 = a^2, \ bab^{-1} = a^{-1} \quad \text{or} \quad a^4 = 1, \ b^2 = 1, \ bab^{-1} = a^{-1}.$$

The first equations give relations of a presentation for $\mathbf{Q}$, by Proposition C-1.129, while the second equations give relations of a presentation of $D_8$, by Proposition C-1.130. Now von Dyck's Theorem gives a surjective homomorphism $\mathbf{Q} \to G$ or $D_8 \to G$, as $|G| = 8$. Theorem C-1.128(ii) says that these homomorphisms must be isomorphisms.

Finally, Exercise A-4.69 on page 159 in Part 1 shows that $\mathbf{Q}$ and $D_8$ are not isomorphic. •

The reader may continue this classification of the groups $G$ of small order $|G| \leq 15$; the results are displayed in Table 2. By Corollary C-1.23, every group of order $p^2$, where $p$ is prime, is abelian, and so every group of order 9 is abelian. By the Fundamental Theorem of Finite Abelian Groups, there are only two such

| Order | Groups |
|---|---|
| 4 | $\mathbb{Z}_4$, $\mathbf{V}$ |
| 6 | $\mathbb{Z}_6$, $S_3$ |
| 8 | $\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, $D_8$, $\mathbf{Q}$ |
| 9 | $\mathbb{Z}_9$, $\mathbb{Z}_3 \times \mathbb{Z}_3$ |
| 10 | $\mathbb{Z}_{10}$, $D_{10}$ |
| 12 | $\mathbb{Z}_{12}$, $\mathbb{Z}_3 \times \mathbf{V}$, $D_{12}$, $A_4$, $T$ |
| 14 | $\mathbb{Z}_{14}$, $D_{14}$ |
| 15 | $\mathbb{Z}_{15}$ |

**Table 2.** Groups of small order.

groups up to isomorphism: $\mathbb{Z}_9$ and $\mathbb{Z}_3 \times \mathbb{Z}_3$. If $p$ is prime, then every group of order $2p$ is either cyclic or dihedral (Exercise C-1.98). Thus, there are only two groups of order 10 and only two groups of order 14. There are five groups of order 12 (Rotman [**188**], p. 84). Two of these are abelian: $\mathbb{Z}_{12} \cong \mathbb{Z}_4 \times \mathbb{Z}_3$ and $\mathbb{Z}_3 \times \mathbf{V}$; the nonabelian groups of order 12 are $D_{12} \cong S_3 \times \mathbb{Z}_2$, $A_4$, and a group $T$ having the presentation

$$T = \big(a, b \mid a^6 = 1, b^2 = a^3 = (ab)^2\big)$$

(Exercise C-1.99 realizes $T$ as a group of matrices). The group $T$, sometimes called a ***dicyclic group*** of type $(2, 2, 3)$, is an example of a *semidirect product* (we shall discuss semidirect products in the chapter on homological algebra). A group of order $pq$, where $p < q$ are primes and $q \not\equiv 1 \bmod p$, must be cyclic, and so there is only one group of order 15 [Rotman [**188**], p. 83]. There are fourteen nonisomorphic groups of order 16, so this is a good place to stop.

## Exercises

∗ **C-1.98.** If $p$ is prime, prove that every group $G$ of order $2p$ is either cyclic or isomorphic to $D_{2p}$.

**Hint.** By Cauchy's Theorem, $G$ must contain an element $a$ of order $p$, and $\langle a \rangle \lhd G$ because it has index 2.

∗ **C-1.99.** Let $G$ be the subgroup of $\mathrm{GL}(2, \mathbb{C})$ generated by $\left[\begin{smallmatrix} \omega & 0 \\ 0 & \omega^2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & i \\ i & 0 \end{smallmatrix}\right]$, where $\omega = e^{2\pi i/3}$ is a primitive cube root of unity.

(i) Prove that $G$ is a group of order 12 that is not isomorphic to $A_4$ or to $D_{12}$.

(ii) Prove that $G$ is isomorphic to the group $T$ in Table 2.

**C-1.100.** Prove that every finite group is finitely presented.

**C-1.101.** Compute the order of the group $G$ with the presentation

$$G = \big(a, b, c, d \mid bab^{-1} = a^2, bdb^{-1} = d^2, c^{-1}ac = b^2, dcd^{-1} = c^2, bd = db\big).$$

## C-1.7. Nielsen–Schreier Theorem

We are now going to prove one of the most fundamental results about free groups: every subgroup is also free. Nielsen proved, in 1921, that finitely generated subgroups of free groups are free.[22] Even if a free group $F$ has finite rank, Nielsen's proof does not show that every subgroup of $F$ is free, for subgroups of a finitely generated free group need not be finitely generated (Corollary C-1.138). The finiteness hypothesis was removed by Schreier in 1926, and the subgroup theorem is called the Nielsen–Schreier Theorem.

In the next section, we will discuss a second proof, found by Baer and Levi in 1933, which uses a correspondence between *covering spaces* $\widetilde{X}$ of a topological space $X$ and subgroups of its fundamental group $\pi_1(X)$. There are interesting variations of this idea. One such involves trees (which arise as *universal covering spaces* of connected graphs); Serre [**202**], p. 27, characterizes free groups by their action on trees, which he then uses to prove the Nielsen–Schreier Theorem ([**202**], p. 29). A second variation is due to Higgins [**99**], pp. 117–118.

There are now many other proofs of the Nielsen–Schreier Theorem, most quite intricate; we prefer the recent proof of Avinoam Mann [**148**], p. 16. Let us begin with an interesting proposition whose proof contains an idea we will use.

Recall the following definition.

**Definition.** Let $H$ be a subgroup of a group $G$. A (right) ***transversal*** of $H$ in $G$ is a subset of $G$ consisting of exactly one element $\tau(Hg) \in Hg$ for every right coset $Hg$, and with $\tau(H) = 1$.

**Proposition C-1.133.** *If $G$ is a finitely generated group, then every subgroup $H$ of finite index is also finitely generated.*

**Proof.** Let $G = \langle x_1, \ldots, x_d \rangle$, let $[G : H] = s$, and let $\{a_1 = 1, a_2, \ldots, a_s\}$ be a transversal $\tau$ of $H$. Let $g = x_{i_1}^{e_1} \cdots x_{i_n}^{e_n}$ be any element of $G$, where each $x_{i_j}$ is some $x_i$ and $e_{i_j} = \pm 1$. If $\tau(Hx_{i_1}^{e_1}) = a_{i_1}$, then there is $h_1 \in H$ with $h_1 x_{i_1}^{e_1} = a_{i_1}$; that is, $x_{i_1}^{e_1} a_{i_1}^{-1} = h_1^{-1} \in H$. Thus, $g = (x_{i_1}^{e_1} a_{i_1}^{-1}) a_{i_1} x_{i_2}^{e_2} \cdots x_{i_n}^{e_n}$. If $\tau(Ha_{i_1} x_{i_2}^{e_2}) = a_{i_2}$, then there is $h_2 \in H$ with $h_2 a_{i_1} x_{i_2}^{e_2} = a_{i_2}$; that is, $a_{i_1} x_{i_2}^{e_2} a_{i_2}^{-1} = h_2^{-1} \in H$, and so $g = (x_{i_1}^{e_1} a_{i_1}^{-1})(a_{i_1} x_{i_2}^{e_2} a_{i_2}^{-1}) a_{i_2} \cdots x_{i_n}^{e_n}$. Continuing in this way,

$$g = (x_{i_1}^{e_1} a_{i_1}^{-1})(a_{i_1} x_{i_2}^{e_2} a_{i_2}^{-1}) a_{i_2} \cdots (a_{i_{n-1}} x_{i_n}^{e_n} a_{i_n}^{-1}) a_{i_n},$$

where each $a_{i_j} x_{i_{j+1}}^{e_{j+1}} a_{i_{j+1}}^{-1} \in H$; that is, $g = h a_{i_n}$ for $h \in H$. In particular, if $g \in H$, then $h a_{i_n} \in H$ and $a_{i_n} = 1$ (because $\tau(H) = 1$). Therefore, every $g \in H$ is a product of elements in $H$ of the form $a_{i_j} x_{i_{j+1}}^{e_{j+1}} a_{i_{j+1}}^{-1}$ (rewrite the first factor $x_{i_1}^{e_1} a_{i_1}^{-1}$ as $1 x_{i_1}^{e_1} a_{i_1}^{-1}$).

[22] If $S$ is a finitely generated subgroup of a free group, then Nielsen's proof shows that $S$ is free by giving an algorithm, analogous to Gaussian elimination in linear algebra, that replaces a finite generating set with a basis of $S$ (Lyndon–Schupp [**142**], pp. 4–13). This theoretical algorithm has evolved into the *Schreier–Sims algorithm*, an efficient way to compute the order of a subgroup $H \subseteq S_n$ when a generating set of $H$ is given.

The subscripts $i_j$ occurring in the factors of the triple products depend on the element $g$, and they may change when $g$ is replaced by another element of $G$. Thus, we may rewrite triple products with simpler subscripts, say, $a_j x_k^e a_\ell^{-1}$. But there are only finitely many $a$'s and $x$'s, and so $H$ is generated by finitely many elements. •

**Remark.** The proof of Proposition C-1.133 shows that a set of generators of $H$ can be described in terms of a set of generators of the large group $G$ and a transversal of $H$ in $G$: it consists of triple products of the form $a_j x_k^e a_\ell^{-1}$. This generating set may be too big. Since $a_j x_k^{-1} a_\ell^{-1} = (a_\ell x_k a_j^{-1})^{-1}$, we may assume that the middle factor has exponent 1.

Another way to shrink the generating set is to throw out triple products equal to 1. For example, suppose that $a_{i_n} = x_{i_1}^{e_1} \cdots x_{i_n}^{e_n}$; if the initial segment $x_{i_1}^{e_1} \cdots x_{i_{n-1}}^{e_{n-1}}$ is also a coset representative in the transversal, call it $a_{i_{n-1}}$, then $a_{i_{n-1}} x_{i_n}^{e_n} = a_{i_n}$, and $a_{i_{n-1}} x_{i_n} a_{i_n}^{-1} = 1$. This procedure can be iterated if every initial segment $x_{i_1}^{e_1} \cdots x_{i_r}^{e_r}$ of $x_{i_1}^{e_1} \cdots x_{i_n}^{e_n}$, for $1 \le r < n$, lies in the transversal; now every triple product $a_{i_{j-1}} x_{i_j}^{e_j} a_{i_j}^{-1} = 1$, for $1 \le j < n$. ◀

Now consider a free group $F$ with basis $X$ and a subgroup $H$, not necessarily of finite index. To prove that $H$ is free, our task is to find a basis of it. We will find a set of generators of $H$ using a special transversal $\tau$ of $H$, and it will turn out that a subset of this generating set will be a basis of $H$.

**Definition.** Let $H$ be a subgroup of a free group $F$ with basis $X$. A transversal $\tau$ of $H$ in $F$ is a ***Schreier transversal*** if each $\tau(Hg) = x_{i_1}^{e_1} \cdots x_{i_n}^{e_n}$ is a reduced word on $X$ such that every initial segment $x_{i_1}^{e_1} \cdots x_{i_r}^{e_r}$ for $r \le n$ is also a coset representative; that is, $x_{i_1}^{e_1} \cdots x_{i_r}^{e_r} = \tau(Hg'_r)$ for some $g'_r \in F$.

Let $F$ be a free group with basis $X$, let $H$ be a subgroup of $F$, and let $\tau$ be a transversal of $H$ in $F$. If $a \in F$ and $x \in X$, define

$$\sigma(a, x) = \tau(Ha)\, x\, \tau(Hax)^{-1}.$$

Note that $\sigma(a, x) = 1$ if and only if $\tau(Ha)\, x = \tau(Hax)$.

Recall that if $F$ is a free group with basis $X$, then we defined the *length* $|u|$ of a word $u = x_{i_1}^{e_1} \cdots x_{i_n}^{e_n} \in F$ to be $n$ (remember that $u$ is reduced). Well-order the basis $X$ of $F$ as $x_0, x_1, \ldots$, and then well-order $X \cup X^{-1}$ by setting the inverse of each basis element $x_j$ to be the next term in the ordering: $x_0, x_0^{-1}, x_1, x_1^{-1}, \ldots$. Now well-order $F(X)$, whose elements are the reduced words on $X$, first by their length as words on $X$ and then lexicographically for words of the same length (this is essentially the degree-lexicographic well-ordering we introduced in Part 1 in our discussion of the division algorithm for polynomials in several variables).

**Lemma C-1.134.** *If $F$ is a free group with basis $X$ and $H$ is a subgroup of $F$, then there exists a Schreier transversal $\tau$ of $H$ in $F$.*

**Proof.** We may assume that the elements of $F$ have been well-ordered, as above. Define the ***length*** $|Hg|$ of a coset $Hg$ to be the first element $hg \in Hg$. We prove, by induction on the length $|Hg|$, that there is a representative $\tau(Hg) \in Hg$ such

that all its initial segments are representatives of cosets of shorter length. Begin by defining $\tau(H) = 1$. For the inductive step, let $|Hg| = n + 1$ and let $ux^e \in Hz$, where $e = \pm 1$ and $|ux^e| = n + 1$. Now $|Hu| = n$, for if its length were $m < n$, it would have a representative $v$ of length $m$, and then $vx^e$ would be a representative of $Hz$ of length $< n + 1$, a contradiction. By induction, there exists $a = \tau(Hu)$ with every initial segment also a representative; define $\tau(Hz) = ax^e$. •

**Lemma C-1.135.** *Let $F$ be a free group with basis $X$, let $H$ be a subgroup of $F$, and let $\tau$ be a Schreier transversal of $H$ in $F$. Then $H$ is generated by*

$$Y = \{\sigma(a, x) = \tau(Ha)\, x\, \tau(Hax)^{-1}\}.$$

**Proof.** This follows from the remark just after the proof of Proposition C-1.133. •

The generating set $Y$ arose by shrinking the larger generating set of all triple products $a_j x_k^e a_\ell^{-1}$ in two ways: first, by eliminating all those with $e = -1$; second, by eliminating some trivial triple products. It is possible that some surviving triple products in $Y$ are also trivial; that is, some $\sigma(a, x) = \tau(Ha)\, x\, \tau(Hax)^{-1} = 1$.

**Theorem C-1.136 (Schreier–Nielsen).** *If $H$ is a subgroup of a free group $F(X)$, then $H$ is free with basis the nontrivial elements of*

$$S = \{\sigma(a, x) = \tau(Ha)\, x\, \tau(Hax)^{-1} \text{ which are distinct from } 1\},$$

*where $\tau$ is a Schreier transversal of $H$.*

**Proof.** We already know, by Lemma C-1.135, that $S$ generates $H$. We now examine the elements of $S$.

Step 1: The elements $\sigma(a_j, x_k) = a_j x_k a_\ell^{-1}$ are reduced words on $X$.

The words $a_j$ and $a_\ell$ are reduced, as are all $\tau(Ha)$, so that if $\sigma(a_j, x_k)$ is not reduced, the middle $x_k$ can be canceled. If $x_k$ is canceled on the left, then $a_j = ux_k^{-1}$, where $u$ is an initial segment of $a_j$; that is, $u = \tau(Hu)$. But $u = a_j x_k$, and thus $u = \tau(Ha_j x_k)$; that is, $u = a_\ell$, and $a_j x_k a_\ell^{-1} = ux_k^{-1} x_k a_\ell^{-1} = 1$, a contradiction. If $x_k$ is canceled on the right, apply the same argument to $(a_j x_k a_\ell^{-1})^{-1} = a_\ell x_k a_j^{-1}$.

Step 2: Let $g$ be any reduced word on $S$, say, $g = \sigma_1^{e_1} \cdots \sigma_m^{e_m}$, where $\sigma_q = a_{j_q} x_{k_q} a_{\ell_q}$, $e_q = \pm 1$, and $m \geq 1$. If $g$ is rewritten as a word in $X$, then no $x_{k_q}^{e_q}$ is canceled.

Simplifying notation to eliminate double subscripts, write

$$\sigma_q \sigma_{q+1}^e = (a_k x_k a_\ell^{-1})(a_r x_s a_t^{-1})^e.$$

Suppose that $x_k$ is canceled from $\sigma_{q+1}^e$ on its right. Since each $\sigma$ is reduced as a word on $X$, by Step 1, the only cancellation possible is at the interface; $a_\ell^{-1}$

must first be canceled completely, after which $x_k$ is canceled. There are only three possibilities:

(i) $x_k$ is canceled by $x_s$. In this case, we must have $e = -1$ and

$$\sigma_q \sigma_{q+1}^{e} = \sigma_q \sigma_{q+1}^{-1} = (a_j x_k a_\ell^{-1})(a_t x_s^{-1} a_r^{-1}).$$

Thus, $a_t = a_\ell$ and $x_s = x_k$. Now the definition of $\sigma(a_j, x_k)$ says that $a_\ell = \tau(Ha_j x_k)$; hence, $\sigma_{q+1} = \tau(Ha_r)x_k\tau(Ha_r x_k)^{-1}$. It follows that $\tau(Ha_r x_k) = a_\ell = \tau(Ha_j x_k)$, so that $Ha_r x_k = Ha_j x_k$ and $Ha_r = Ha_j$. Therefore, $\tau(Ha_r) = \tau(Ha_j) = a_j$, and

$$\sigma_{q+1} = \sigma(a_r, x_s) = \tau(Ha_r)x_s a_t^{-1} = \tau(Ha_j)x_k a_\ell^{-1} = \sigma_q,$$

which contradicts the word $g$ on $S$ being reduced.

(ii) $x_k^{-1}$ occurs in $a_r$ (remember that $a_r$ is a word on $X$); we take the first such occurrence. In this case, we must have $e = +1$ and $a_r$ starts with $a_\ell x_k^{-1}$; that is, $a_\ell x_k^{-1}$ is an initial segment of $a_r = \tau(Ha_r)$. Since $\tau$ is a Schreier transversal, $a_\ell x_k^{-1} = \tau(Ha_\ell x_k^{-1})$. Looking at $\sigma(a_j, x_k)^{-1} = a_\ell x_k^{-1} a_j^{-1}$, we see that $a_j = \tau(Ha_\ell x_k^{-1}) = a_\ell x_k^{-1}$, and thus $\sigma(a_j, x_k) = 1$, a contradiction.

(iii) $x_k$ is canceled by the first $x_k^{-1}$ occurring in $a_t$. In this case, we must have $e = -1$. Inverting $g$, we get a word in which $\sigma(a_r, x_s)^{-1}\sigma(a_j, x_k)^{-1}$ occurs, and the $x_s^{-1}$ in $\sigma(a_r, x_s)$ is canceled in the same manner as in the first case for $x_k$ above and, as there, this implies that $\sigma(a_r, x_s) = 1$.

In the same way, we deal with the possibility that $x_k$ is canceled from $\sigma_{q-1}^{e'}$ on its left.

It follows from Step 2 that every nontrivial reduced word on $S$ is not equal to 1, and so Corollary C-1.121 shows that $H$ is free with basis $S$. •

Here is a nice application of the Nielsen–Schreier Theorem.

**Corollary C-1.137.** *Let $F$ be a free group, and let $u, v \in F$. Then $u$ and $v$ commute if and only if there is $z \in F$ with $u, v \in \langle z \rangle$.*

**Proof.** Sufficiency is obvious; if both $u, v \in \langle z \rangle$, then they lie in an abelian subgroup, and hence they commute.

Conversely, the Nielsen–Schreier Theorem says that the subgroup $\langle u, v \rangle$ is free. On the other hand, the condition that $u$ and $v$ commute says that $\langle u, v \rangle$ is abelian. But an abelian free group is cyclic, by Exercise C-1.92 on page 91; therefore, $\langle u, v \rangle \cong \langle z \rangle$ for some $z \in G$. •

The next result shows, in contrast to abelian groups, that a subgroup of a finitely generated group need not be finitely generated.

**Corollary C-1.138.** *If $F$ is a free group of* rank 2*, then its commutator subgroup $F'$ is a free group of infinite rank.*

**Proof.** Let $\{x, y\}$ be a basis of $F$. Since $F/F'$ is free abelian with basis $\{xF', yF'\}$, by Lemma C-1.123, every coset $F'b$ has a unique representative of the form $x^m y^n$,

where $m, n \in \mathbb{Z}$; it follows that the transversal choosing $\tau(F'b) = x^m y^n$ is a Schreier transversal, for every subword of $x^m y^n$ is a word of the same form. If $n > 0$, then $\tau(F'y^n) = y^n$, but $\tau(F'y^n x) = xy^n \neq y^n x$. Therefore, there are infinitely many elements $\sigma(y^n, x) = \tau(F'y^n)x\tau(F'y^n x)^{-1} \neq 1$, and so the result follows from the Nielsen–Schreier Theorem. •

Recall that Proposition C-1.133 says, if $H$ is a subgroup of finite index of a finitely generated group, that $H$ is also finitely generated. In particular, if $H$ is a subgroup of a finitely generated free group, then $H$ is a finitely generated free group, hence has finite rank. We now compute $\operatorname{rank}(H)$.

**Corollary C-1.139.** *If $F$ is a free group of finite rank $n$ and $H$ is a subgroup of finite index $j$, then* $\operatorname{rank}(H) = jn - j + 1$.

**Remark.** In geometric proofs of the Nielsen–Schreier Theorem, one sees that there is a space $K$, with Euler–Poincaré characteristic $\chi(K)$, such that $\operatorname{rank}(F) = -\chi(K) + 1$ and $\operatorname{rank}(H) = j(-\chi(K_H) + 1)$, where $K_H$ is a *covering space* constructed from $K$ and $H$. ◄

**Proof.** Let $X = \{x_1, \ldots, x_n\}$ be a basis of $F$ and let $\tau$ be a Schreier transversal of $H$ in $F$. By Theorem C-1.136, a basis of $H$ consists of all those elements $\sigma(a, x)$ not equal to 1. There are $j$ choices for $Ha$ and $n$ choices for $x$, and so there are at most $jn$ elements in a basis of $H$, and $\operatorname{rank}(H) \leq jn$.

Call an ordered pair $(Ha, x)$ ***trivial*** if $\sigma(a, x) = 1$; that is, if $\tau(Ha)x = \tau(Hax)$. We will show that there is a bijection $\psi$ between the family of cosets $\{Ha : a \notin H\}$ and the trivial ordered pairs, so that there are $j - 1$ trivial ordered pairs. It will then follow that $\operatorname{rank}(H) = jn - (j - 1) = jn - j + 1$.

Let $\tau(Ha) = a$; since $Ha \neq H$, we have $a = ux^e$, where $e = \pm 1$ and $u = \tau(Hu)$ ($u = 1$ is possible). Define $\psi(Ha)$ as follows, where $a = ux^e$:

$$\psi(Ha) = \psi(Sux^e) = \begin{cases} (Hu, x) & \text{if } e = +1, \\ (Hux^{-1}, x) & \text{if } e = -1. \end{cases}$$

Note that $\psi(Hux^e)$ is a trivial ordered pair: if $e = +1$, then $(Hu, x)$ is trivial, for $\tau(Hu)x\tau(Hux)^{-1} = ux(ux)^{-1} = 1$; if $e = -1$, then $(Hux^{-1}, x)$ is trivial, for $(ux^{-1})x = u$ and $\tau(Hux^{-1})x\tau(Hux^{-1}x)^{-1} = (ux^{-1})xu^{-1} = 1$.

To see that $\psi$ is injective, suppose that $\psi(Ha) = \psi(Hb)$, where $a = ux^e$ and $b = vy^\eta$; we assume that $x, y$ lie in the given basis of $F$ and that $e = \pm 1$ and $\eta = \pm 1$. There are four possibilities, depending on the signs of $e$ and $\eta$. If $e = +1 = \eta$, then $(Hu, x) = (Hv, y)$; hence, $x = y$ and $Hu = Hv$; that is, $u = v$, for $\tau$ is a Schreier transversal and $u, v \in \operatorname{im}\tau$. Thus, $a = ux = vy = b$ and $Ha = Hb$. Similar calculations show that $Ha = Hb$ when $e = -1 = \eta$ and when $e, \eta$ have opposite sign.

To see that $\psi$ is surjective, take a trivial ordered pair $(Hw, x)$, and so $\tau(Hw)x = wx = \tau(Hwx)$. Now $w = ux^e$, where $u \in \operatorname{im}\tau$ and $e = \pm 1$. If $e = +1$, then $w$ does not end with $x^{-1}$, and $\psi(Hwx) = (Hw, x)$. If $e = -1$, then $w$ does end with $x^{-1}$, and so $\psi(Hu) = (Hux^{-1}, x) = (Hw, x)$. •

**Corollary C-1.140.** *There exist nonisomorphic finitely presented groups $G$ and $H$ each of which is isomorphic to a subgroup of the other.*

**Proof.** If $G$ is a free group of rank 2 and $H$ is a free group of rank 3, then $G \ncong H$. Both $G$ and $H$ are finitely generated free groups and, hence, are finitely presented. Clearly, $G$ is isomorphic to a subgroup of $H$. On the other hand, the commutator subgroup $G'$ is free of infinite rank, and so $G'$, hence $G$, contains a free subgroup of rank 3; that is, $H$ is isomorphic to a subgroup of $G$. •

## Exercises

**C-1.102.** Prove that if $F$ is free of finite rank $n \geq 2$, then its commutator subgroup $F'$ is free of infinite rank.

**C-1.103.** Let $G$ be a finite group that is not cyclic. If $G \cong F/N$, where $F$ is a free group of finite rank, prove that $\operatorname{rank}(N) > \operatorname{rank}(F)$.

**C-1.104.** Prove that if $G$ is a finite group generated by two elements $a, b$ having order 2, then $G \cong D_{2n}$ for some $n \geq 2$.

∗ **C-1.105.** Let $Y$ and $S$ be groups, and let $\varphi\colon Y \to S$ and $\theta\colon S \to Y$ be homomorphisms with $\varphi\theta = 1_S$.

(i) If $\rho\colon Y \to Y$ is defined by $\rho = \theta\varphi$, prove that $\rho\rho = \rho$ and $\rho(a) = a$ for every $a \in \operatorname{im}\theta$. (The homomorphism $\rho$ is called a ***retraction***.)

(ii) If $K$ is the normal subgroup of $Y$ generated by all $y^{-1}\rho(y)$ for $y \in Y$, prove that $K = \ker\varphi$.
**Hint.** Note that $\ker\varphi = \ker\rho$ because $\theta$ is an injection. Use the equation $y = \rho(y)(\rho(y)^{-1})y$ for all $y \in Y$.

## C-1.8. The Baer–Levi Proof

*Covering spaces* arise in studying various mathematical topics such as Riemann surfaces, Lie groups, and fundamental groups, for example. Highlights of this theory are presented here, but we will merely sketch proofs or just state theorems (a more complete account can be found in Rotman [**191**], Chapter 10, which discusses the topological setting, or in Rotman [**188**], pp. 366–406, and Rotman [**190**], which discuss the simpler combinatorial setting). Baer and Levi gave an elegant proof of the Schreier–Nielsen Theorem using covering spaces, which we will give here. We shall also use covering spaces to prove a theorem of Kurosh which describes the subgroups of free products.

### ▪ The Categories Simp and $\mathrm{Simp}_*$

Perhaps the simplest topological spaces are those which can be *triangulated*, that is, spaces that are homeomorphic to spaces which can be constructed from gluing together points, line segments, triangles, pyramids, etc.

**Definition.** A ***complex*** $K$ (often called an ***abstract simplicial complex***) is either the empty set $\varnothing$ or a family of nonempty finite subsets, called ***simplexes***, of a set $\mathrm{Vert}(K)$, called ***vertices***, such that

(i) if $v \in \mathrm{Vert}(K)$, then $\{v\}$ is a simplex;

(ii) if $s$ is a simplex, then so is every nonempty subset of $s$.

The complex $K = \varnothing$ has no simplexes.

A simplex $s = \{v_0, \ldots, v_q\}$ with $q+1$ different vertices is called a ***$q$-simplex***, and we say that $s$ has ***dimension*** $q$, denoted by $\dim(s) = q$. We say that a complex $K$ is an ***$n$-complex***, denoted by $\dim(K) = n$, if $n$ is the largest dimension of a simplex in $K$ (if there is no simplex of largest dimension, then $\dim(K) = \infty$).

A 0-simplex is a 1-point set $\{v\} \subseteq \mathrm{Vert}(K)$ (which we usually identify with the vertex $v \in \mathrm{Vert}(K)$); a 1-simplex $\{v_0, v_1\}$ can be viewed as a line segment with endpoints $v_0, v_1$; a 2-simplex $\{v_0, v_1, v_2\}$ can be viewed as a (two-dimensional) triangle with vertices $v_0, v_1, v_2$; a 3-simplex $\{v_0, v_1, v_2, v_3\}$ can be viewed as a solid tetrahedron with vertices $v_0, v_1, v_2, v_3$, and so forth. For our discussion here, 2-complexes will suffice. Geometrically, a complex is a space obtained by assembling various simplexes together nicely.

Here are more definitions (fortunately, most are quite intuitive).

**Definition.** A complex $L$ is a ***subcomplex*** of a complex $K$ if $L = \varnothing$ or $\mathrm{Vert}(L) \subseteq \mathrm{Vert}(K)$ and every simplex in $L$ is also a simplex in $K$.

If $L_1$ and $L_2$ are subcomplexes of a complex $K$, then both

$$L_1 \cup L_2 = \{\text{all simplexes } s \text{ in } L_1\} \cup \{\text{all simplexes } s' \text{ in } L_2\}$$

and

$$L_1 \cap L_2 = \{\text{all simplexes } s \text{ in both } L_1 \text{ and } L_2\}$$

are subcomplexes of $K$.

Given a complex $K$, define the ***$q$-skeleton*** $K^q$, for each $q \geq 0$, by

$$K^q = \{\text{all simplexes } s \text{ in } K : \dim(s) \leq q\}.$$

Note that $\mathrm{Vert}(K) = K^0$ and $K^0 \subseteq K^1 \subseteq K^2 \subseteq \cdots$.

For example, if $s = \{v_0, v_1, v_2\}$ is a 2-simplex, its 1-skeleton $s^1$ is a ***circle***, the 1-complex having 1-simplexes $\{v_0, v_1\}$, $\{v_0, v_2\}$, and $\{v_1, v_2\}$ (so $s^1$ is the perimeter of the triangle with vertices $\{v_0, v_1, v_2\}$).

A 1-complex can be viewed as a graph, where vertices $v_0, v_1$ are *adjacent* if $\{v_0, v_1\}$ is a $q$-simplex for $q \leq 1$. In particular, the 1-skeleton $K^1$ of a complex $K$ can be regarded as a graph. Thus, graph-theoretical definitions in Section C-1.1 can be used here.

There is also a notion of *quotient complex*.

**Definition.** Let $K$ be a complex and let $\equiv$ be an equivalence relation on $\mathrm{Vert}(K)$. The ***quotient complex*** $K/\equiv$ is the complex with

$$\mathrm{Vert}(K/\equiv) = \{\text{all equivalence classes } [v] \text{ of } \equiv\}$$

and simplexes $\{[v_{i_0}], \ldots, [v_{i_q}]\}$ if there are vertices $v_{i_j} \in [v_{i_j}]$ with $\{v_{i_0}, \ldots, v_{i_q}\}$ a simplex in $K$.

The reader may check that $K/\equiv$ is a complex.

We now define morphisms of complexes.

**Definition.** If $K$ and $L$ are complexes, then a ***simplicial map*** $\varphi\colon K \to L$ is a function $\varphi\colon \mathrm{Vert}(K) \to \mathrm{Vert}(L)$ that takes simplexes to simplexes; that is, if $s = \{v_0, \ldots, v_q\}$ is a simplex in $K$, then $\varphi s = \{\varphi v_0, \ldots, \varphi v_q\}$ is a simplex in $L$.

A (simplicial) map $\varphi\colon K \to L$ is an ***isomorphism*** if there is a simplicial map $\psi\colon L \to K$ with $\psi\varphi = 1_K$ and $\varphi\psi = 1_L$. A simplicial map $\varphi\colon K \to L$ is ***surjective*** if, for each simplex $s'$ in $L$, there is a simplex $s$ in $K$ with $\varphi s = s'$.

If $\varphi\colon K \to L$ is a simplicial map, then the simplex $\varphi s = \{\varphi v_0, \ldots, \varphi v_q\}$ in $L$ may have repeated vertices; thus, $\dim(s) \geq \dim(\varphi s)$, and strict inequality is possible. For example, the constant function $\varphi\colon \mathrm{Vert}(K) \to \mathrm{Vert}(L)$, given by $\varphi v = w$ for all $v \in \mathrm{Vert}(K)$, where $w \in \mathrm{Vert}(L)$, is a simplicial map.

The inclusion $i\colon L \to K$ of a subcomplex is a simplicial map. Another example arises from an equivalence relation $\equiv$ on $\mathrm{Vert}(K)$, where $K$ is a complex. The ***natural map*** $\nu\colon K \to K/\equiv$, defined by $\nu\colon v \mapsto [v]$, is a simplicial map, where $[v]$ is the equivalence class of $v$.

It is easy to check that the composite of simplicial maps is again a simplicial map and that

$$\mathbf{Simp},$$

consisting of all complexes and simplicial maps, is a category.

A related category is a variation of **Simp**. If $K$ is a complex, a choice of vertex $w \in \mathrm{Vert}(K)$ is called a ***basepoint***, and an ordered pair $(K, w)$ is called a ***pointed complex***. If $(K, w)$ and $(L, w')$ are pointed complexes, then a ***pointed map*** $\varphi\colon (K, w) \to (L, w')$ is a simplicial map $\varphi\colon K \to L$ with $\varphi w = w'$.[23] It is easy to check that any composite of pointed maps is a pointed map, and we define

$$\mathbf{Simp}_*$$

to be the category of pointed complexes.

## ▪ Fundamental Group

We are now going to define *paths* in complexes. This will enable us to define *fundamental groups* of complexes (actually, of pointed complexes) which gives a functor $\pi_1\colon \mathbf{Simp}_* \to \mathbf{Groups}$.

[23] We have defined pointed topological spaces and pointed maps on page 87.

Earlier, we remarked that the 1-skeleton $K^1$ of a complex $K$ can be viewed as a graph; we now make it into a directed graph.

**Definition.** Let $K$ be a complex, and let $u, v \in \mathrm{Vert}(K)$. If $\{u, v\}$ is a $q$-simplex in $K$, where $q \leq 1$, then the ordered pair $e = (u, v)$ is called an ***edge*** in $K$; we write $u = o(e)$ and $v = t(e)$ (***origin*** and ***terminus***); the edge $e^{-1} = (v, u)$ has $v = o(e^{-1})$ and $u = t(e^{-1})$.[24] A ***path*** $\alpha$ in $K$ of ***length*** $n$ from $u$ to $v$ is a sequence of $n$ edges

$$\alpha = e_1 \cdots e_n = (u, v_1)(v_1, v_2) \cdots (v_{n-1}, v),$$

where $t(e_i) = o(e_{i+1})$ for all $i$. We call $u$ the ***origin*** of $\alpha$, denoting it by $o(\alpha)$, and $v$ the ***terminus*** of $\alpha$, denoting it by $t(\alpha)$. A path $\alpha$ is ***closed at*** $w$ if $o(\alpha) = w = t(\alpha)$.

Simplicial maps take paths into paths.

**Definition.** Let $K$ and $L$ be complexes, and let $\alpha = (u, v_1)(v_1, v_2) \cdots (v_{n-1}, v)$ be a path in $K$ from $u$ to $v$. If $\varphi\colon K \to L$ is a simplicial map, then

$$\varphi\alpha = (\varphi u, \varphi v_1)(\varphi v_1, \varphi v_2) \cdots (\varphi v_{n-1}, \varphi v).$$

It is clear that $\varphi\alpha$ is a path in $L$ from $\varphi u$ to $\varphi v$.

**Definition.** A complex $K$ is ***connected*** if, for every pair of vertices $u, v \in \mathrm{Vert}(K)$, there is a path in $K$ from $u$ to $v$.

**Proposition C-1.141.** *Let $K$ and $L$ be complexes, and let $\varphi\colon K \to L$ be a surjective simplicial map. If $K$ is connected, then $L$ is connected.*

**Proof.** Let $u', v' \in \mathrm{Vert}(L)$; since $\varphi$ is surjective, there are $u, v \in \mathrm{Vert}(K)$ with $\varphi u = u'$ and $\varphi v = v'$. As $K$ is connected, there is a path $\alpha$ in $K$ from $u$ to $v$, and $\varphi\alpha$ is a path in $L$ from $u'$ to $v'$. •

Here is a combinatorial version of *homotopy*.

**Definition.** Define two types of ***elementary moves*** that can be applied to a path $\alpha$ in a complex $K$.

(i) Replace a pair of adjacent edges $(u, v)(v, w)$ in $\alpha$ by the single edge $(u, w)$ if $\{u, v, w\}$ is a simplex in $K$;

(ii) the inverse operation: replace an edge $(u, w)$ in $\alpha$ by $(u, v)(v, w)$ if $\{u, v, w\}$ is a simplex in $K$.

Paths $\alpha$ and $\beta$ in a complex $K$ are ***homotopic***, denoted by $\alpha \simeq \beta$, if $\beta$ can be obtained from $\alpha$ by finitely many elementary moves.

If $\alpha \simeq \beta$, then $o(\alpha) = o(\beta)$ and $t(\alpha) = t(\beta)$. In particular, if $\alpha$ is a closed path at $w$ and $\alpha \simeq \beta$, then $\beta$ is a closed path at $w$.

It is clear that homotopy is an equivalence relation on the family of all paths in $K$.

[24] We think of $(u, v)$ as an edge from $u$ to $v$, while the edge $(v, u)$ is viewed as going backwards, from $v$ to $u$.

**Definition.** If $\alpha$ is a path in a complex $K$, then its equivalence class with respect to homotopy is called its ***path class***; it is denoted by

$$[\alpha].$$

Since homotopic paths have the same origin and the same terminus, we can define the origin and terminus of path classes: if $\alpha$ is a path in a complex $K$, then

$$o([\alpha]) = o(\alpha) \quad \text{and} \quad t([\alpha]) = t(\alpha).$$

Paths in a complex can be multiplied if the first ends where the second begins.

**Definition.** If $\alpha = e_1 \cdots e_n$ and $\beta = d_1 \cdots d_m$ are paths in a complex $K$ with $t(\alpha) = o(\beta)$, where the $e_i$ and $d_j$ are edges, then their ***product*** is

$$\alpha\beta = e_1 \cdots e_n d_1 \cdots d_m,$$

which is a path from $o(\alpha)$ to $t(\beta)$. In particular, if both $\alpha$ and $\beta$ are closed paths at $w$, then $\alpha\beta$ is defined and it is also a closed path at $w$.

Observe that multiplication of paths is associative.

Homotopy is compatible with multiplication of paths.

**Lemma C-1.142.** *Let $\alpha$, $\alpha'$, $\beta$, and $\beta'$ be paths in a complex $K$ with $t(\alpha) = o(\beta)$. If $\alpha \simeq \alpha'$ and $\beta \simeq \beta'$, then $\alpha\beta \simeq \alpha'\beta'$.*

**Proof.** First, $t(\alpha) = t(\alpha')$ and $o(\beta) = o(\beta')$ implies that the product $\alpha'\beta'$ is defined. Second, it is easy to see that $\alpha\beta \simeq \alpha'\beta$ and $\alpha'\beta \simeq \alpha'\beta'$. Since homotopy is a transitive relation, we have $\alpha\beta \simeq \alpha'\beta'$. •

**Corollary C-1.143.** *Let $(K, w)$ be a pointed complex. Then*

$$[\alpha][\beta] = [\alpha\beta]$$

*is a well-defined binary operation on*

$$\pi_1(K, w) = \{[\alpha] : \alpha \text{ is a closed path in } K \text{ at } w\}.$$

**Proof.** If $\alpha$ and $\beta$ are closed paths in $K$ at $w$, then so is their product $\alpha\beta$. That $[\alpha][\beta] = [\alpha\beta]$ follows from Lemma C-1.142. •

Here is the important definition.

**Definition.** If $(K, w)$ is a pointed complex, then $\pi_1(K, w)$ is called its ***fundamental group*** (or its ***edgepath group***).

Just because we call $\pi_1(K, w)$ a group doesn't automatically make it one.

**Theorem C-1.144.** *The fundamental group $\pi_1(K, w)$ of a pointed complex $(K, w)$ is a group with binary operation the multiplication of path classes in* Corollary C-1.143.

**Proof.** Recall Exercise A-4.27 on page 138 in Part 1 which gives a short list of axioms defining a group. Let $H$ be a semigroup containing an element $e$ such that $ex = x$ for all $x \in H$. If, for every $x \in H$, there is $x' \in H$ with $x'x = e$, then $H$ is a group.

That multiplication of path classes is associative follows from associativity of multiplication of paths:

$$\begin{aligned}([\alpha][\beta])[\gamma] &= [\alpha\beta][\gamma] = [(\alpha\beta)\gamma] \\ &= [\alpha(\beta\gamma)] = [\alpha][\beta\gamma] = [\alpha]([\beta][\gamma]).\end{aligned}$$

Now $(w, w)$ is a closed path in $K$ at $w$; it is called the ***trivial path***, and its path class $\varepsilon = [(w, w)]$ is the identity: $[\varepsilon][\alpha] = [\alpha]$.

Define the ***inverse*** of an edge $e = (u, v)$ to be $e^{-1} = (v, u)$, and define the ***inverse*** of a closed path $\alpha = e_1 \cdots e_n$ to be the path $\alpha^{-1} = e_n^{-1} \cdots e_1^{-1}$. Then $[e_n^{-1} \cdots e_1^{-1}][e_1 \cdots e_n] = \varepsilon$. •

Exercise C-1.107 on page 109 says that if $w$ and $w'$ are vertices in a connected complex $K$, then $\pi_1(K, w) \cong \pi_1(K, w')$.

**Remark.** Every complex $K$ has a ***geometric realization***: a topological space $|K|$ that can be triangulated according to the simplexes in $K$ (Rotman [**191**], p. 142), and the fundamental group $\pi_1(|K|, w)$, defined in Example B-4.15 in Part 1, is isomorphic to $\pi_1(K, w)$ (see Rotman [**191**], Theorem 7.36). ◀

**Theorem C-1.145.** *The fundamental group defines a* (*covariant*) *functor*

$$\pi_1 \colon \mathbf{Simp}_* \to \mathbf{Groups}.$$

**Proof.** If $\varphi \colon (K, w) \to (L, w')$ is a pointed simplicial map, define $\varphi_* \colon \pi_1(K, w) \to \pi_1(L, w')$ by

$$\varphi_* \colon [\alpha] \mapsto [\varphi\alpha]. \quad \bullet$$

Here are several general theorems that compute fundamental groups.

**Theorem C-1.146.** *Given a group $G$, there exists a connected pointed 2-complex $(K, w)$ with $G \cong \pi_1(K, w)$. Moreover, $G$ is finitely presented if and only if there is such a $(K, w)$ with* $\mathrm{Vert}(K)$ *finite.*

**Proof.** See Rotman [**188**], Theorem 11.64 and Corollary 11.65. •

Tietze's Theorem below gives a presentation of $\pi_1(K, w)$ when $K$ is connected. We need the notion of a *tree* to state this theorem; recall the definition from Section C-1.1.

**Definition.** A path $\alpha = e_1 \cdots e_n$ is ***reduced*** if either $\alpha$ is trivial, i.e., $\alpha = (v, v)$, or if no $e_i$ is trivial and no edge $e_j = (u, v)$ is adjacent to its inverse $(v, u)$. A ***circuit*** is a reduced closed path. A ***tree*** is a connected graph $T$ having no circuits.

**Lemma C-1.147.** *Every closed path $\alpha$ in a complex $K$ is homotopic either to a trivial path or a circuit.*

**Proof.** If $\alpha$ contains a subpath $(u,v)(v,u)$, then $\alpha \simeq \alpha'$, where $\alpha'$ is the path obtained from $\alpha$ by replacing $(u,v)(v,u)$ by the trivial edge $(u,u)$. If $\alpha'$ is trivial, we are done; if $\alpha'$ is not trivial, then $\alpha' \simeq \alpha''$, where $\alpha''$ is obtained from $\alpha'$ by removing $(u,u)$. These elementary moves can be iterated, each time reducing the length of the path, and the final path reached is either reduced or trivial. But a reduced closed path is a circuit. •

**Definition.** A pointed complex $(K,w)$ is ***simply connected*** if $\pi_1(K,w) = \{1\}$.

**Proposition C-1.148.** *Every tree $T$ is simply connected.*

**Proof.** By Lemma C-1.147, every closed path in $T$ is homotopic to a trivial path or a circuit. But a tree has no circuits. •

**Definition.** If $K$ is a connected complex, then a subcomplex $T$ of $K$ is a ***maximal tree*** if $T$ is a tree and there does not exist a subcomplex $T'$, which is a tree, such that $T \subsetneq T' \subseteq K$.

**Proposition C-1.149.** *Let $K$ be a connected complex.*

(i) *$K$ contains a maximal tree.*

(ii) *A tree $T \subseteq K$ is a maximal tree if and only if* $\mathrm{Vert}(T) = \mathrm{Vert}(K)$.

**Proof.**

(i) A routine application of Zorn's Lemma.

(ii) Suppose that $T$ is a tree in $K$ and there is $v \in \mathrm{Vert}(K)$ which is not a vertex of $T$. Choose a vertex $v_0 \in \mathrm{Vert}(T)$; as $K$ is connected, there is a path $e_1 \cdots e_n$ in $K$ from $v_0$ to $v$. Since $v_0 \in \mathrm{Vert}(T)$ and $v \notin \mathrm{Vert}(T)$, there is some $e_i = (v_i, v_{i+1})$ with $v_i \in \mathrm{Vert}(T)$ and $v_{i+1} \notin \mathrm{Vert}(T)$. Consider the subcomplex $T'$ of $K$ obtained from $T$ by adjoining $v_{i+1}$ and the simplex $\{v_i, v_{i+1}\}$. Clearly, $T'$ is connected; moreover, any circuit in $T'$ must involve the new vertex $v_{i+1}$. Now there are only two nontrivial edges in $T'$ involving $v_{i+1}$, namely, $e = (v_i, v_{i+1})$ and its inverse $(v_{i+1}, v_i)$. Thus, any closed path in $T'$ involving $v_{i+1}$ (not as its origin) is not reduced, while any circuit at $v_{i+1}$ yields a circuit in $T$ at $v_{i+1}$. Thus, $T'$ is a tree properly containing $T$, so that the tree $T$ is not maximal. The proof of the converse is similar to that just given. •

A connected complex may have many maximal trees.

**Theorem C-1.150 (Tietze).** *If $T$ is a maximal tree in a connected pointed complex $(K,w)$, then a presentation of $\pi_1(K,w)$ is*

$$(X \mid R_1 \cup R_2),$$

*where $X$ is the set of all edges $(u,v)$ in $K$, $R_1 = \{(u,v) \in T\}$, and $R_2 = \{(u,v)(v,y) = (u,y) : \{u,v,y\}$ is a $q$-simplex in $K$ for $q \leq 2\}$.*

**Proof.** See Rotman [**188**], Theorem 11.31. •

**Corollary C-1.151.** *Let $(K, w)$ be a pointed complex. If $\dim(K) \leq 1$, then $\pi_1(K, w)$ is a free group.*

**Proof.** $R_2 = \varnothing$. •

The next result says that if $(K, w)$ is a pointed complex containing pointed subcomplexes $(L_1, w)$ and $(L_2, w)$ (same basepoints) whose union and intersection are connected, then a presentation of $\pi_1(K, w)$ can be given in terms of presentations of $\pi_1(L_1, w)$ and $\pi_1(L_2, w)$.

**Theorem C-1.152 (van Kampen).** *Let $(K, w)$ be a connected pointed complex, and let $(L_1, w)$ and $(L_2, w)$ be connected pointed subcomplexes. If $K = L_1 \cup L_2$ and $L_1 \cap L_2$ are connected, then $\pi_1(K, w)$ is the pushout of the diagram*

$$\begin{array}{ccc} \pi_1(L_1 \cap L_2, w) & \xrightarrow{j_{1*}} & \pi_1(L_1, w) \\ {\scriptstyle j_{2*}}\downarrow & & \\ \pi_1(L_2, w) & & \end{array}$$

*where $j_i \colon L_i \to K$ is the inclusion for $i = 1, 2$. Moreover, a presentation of $\pi_1(K, w)$ can be given in terms of presentations of $\pi_1(L_1, w)$ and $\pi_1(L_2, w)$.*

**Proof.** See Rotman [**188**], p. 396. •

The pushout is called a ***free product with amalgamated subgroup*** in the special case in which the homomorphisms $j_{1*}$ and $j_{2*}$ are injections.

## Exercises

∗ **C-1.106.** Let $K$ be a complex. Define a relation on $\mathrm{Vert}(K)$ by $u \equiv v$ if there is a path in $K$ from $u$ to $v$.

(i) Prove that $\equiv$ is an equivalence relation on $\mathrm{Vert}(K)$. The equivalence classes are called the ***components*** of $K$.

(ii) Prove that every component of $K$ is a connected subcomplex.

(iii) If $(K, w)$ is a pointed complex and $L$ is the component of $K$ containing $w$, prove that $\pi_1(K, w) \cong \pi_1(L, w)$.

∗ **C-1.107.** If $K$ is a connected complex and $w, v \in \mathrm{Vert}(K)$, prove that

$$\pi_1(K, w) \cong \pi_1(K, v).$$

**Hint.** If $\gamma$ is a path in $K$ from $w$ to $v$, prove that $[\alpha] \mapsto [\gamma^{-1}][\alpha][\gamma]$ is an isomorphism.

**C-1.108.** Does the First Isomorphism Theorem hold for complexes; that is, if $\varphi \colon K \to L$ is a surjective simplicial map of complexes, is there an equivalence relation $\equiv$ on $\mathrm{Vert}(K)$ with $L \cong K/\!\equiv$?

**C-1.109.** Prove that a complex $K$ is connected if and only if its 1-skeleton $K^1$ is connected.

**C-1.110.** Let $I_n$ be the 1-complex with $\mathrm{Vert}(I_n) = \{t_0, \ldots, t_n\}$ and simplexes $\{t_i, t_{i+1}\}$ for all $i \geq 0$. Prove that any path of length $n$ in a complex $K$ is a simplicial map $I_n \to K$.

**C-1.111.** If $T$ is a tree and $u, v \in \mathrm{Vert}(T)$, prove that there is a unique path in $T$ from $u$ to $v$.

**C-1.112.** Prove that every simplex is simply connected.

∗ **C-1.113.** Let $K$ be a connected complex, and let $L$ be a subcomplex that is a disjoint union of trees. Prove that there exists a maximal tree in $K$ containing $L$.

**Hint.** Construct a suitable equivalence relation $\equiv$ on $\mathrm{Vert}(K)$, and use the quotient complex $K/\!\equiv$.

**C-1.114.** Use Tietze's Theorem to show that $\pi_1(K, w) \cong \mathbb{Z}$, where $K$ is the 1-skeleton of a 2-simplex. (This isomorphism is needed to prove that ***winding numbers*** are well-defined.)

∗ **C-1.115.** Let $F$ be a free group with basis $X = \{x, y\}$, and let $\Gamma$ be the Cayley graph of $F$ with respect to the generating set $X$. Prove that $\Gamma$ is a tree.

∗ **C-1.116.** (i) Use Tietze's Theorem to prove that if $K$ is a connected 1-complex, then $\pi_1(K, w)$ is a free group. Moreover, if $T$ is a maximal tree in $K$, then

$$\mathrm{rank}(\pi_1(K, w)) = |\{\text{1-simplexes not in } T\}|.$$

(ii) Let $B_n$ be the 1-complex with $\mathrm{Vert}(B_n) = \{w\} \cup \{u_i, v_i : 1 \leq i \leq n\}$ and 1-simplexes $\{\{u_i, v_i\}, \{u_i, w\}, \{v_i, w\} : 1 \leq i \leq n\}$ ($B_n$ is called the ***bouquet of*** $n$ ***circles***). Prove that $\pi_1(B_n, w)$ is a free group of rank $n$.

(iii) If $K$ is a finite complex, define its ***Euler–Poincaré characteristic*** $\chi(K)$ by

$$\chi(K) = \sum_{q=0}^{\dim(K)} (-1)^q \sigma_q(K),$$

where $\sigma_q(K)$ is the number of $q$-simplexes in $K$.

Show that $\sigma_1(B_n) = 3n$ and $\sigma_0(B_n) = 2n + 1$. Conclude that $\chi(B_n) = -n + 1$.

(iv) Prove that $\pi_1(B_n, w)$ is free of rank $-\chi(B_n) + 1$.

## ▪ Covering Complexes

The usual definition of a covering space of a topological space $X$ is an ordered pair $(E, p)$, where $E$ and $X$ are connected topological spaces and $p\colon E \to X$ is continuous, such that

(i) $p$ is a surjection;

(ii) each $x \in X$ has an open neighborhood $V_x$ such that $p^{-1}(V_x)$ is a disjoint union of subspaces $\widetilde{V}_i \subseteq E$ with $p|\widetilde{V}_i\colon \widetilde{V}_i \to V_x$ a homeomorphism for all $i$.

For example, $\exp\colon \mathbb{R} \to S^1$, defined by $\exp\colon t \mapsto e^{2\pi i t}$, is a covering space. If $z = e^{2\pi i x} \in S^1$, define $V_z = \{e^{2\pi i t} : 0 < t < 1\}$; the subspaces $\widetilde{V}_n$ are the open intervals $\widetilde{V}_n = (n, n+1)$ for all $n$.

We give the version for complexes after defining the inverse image of a simplicial map.

**Definition.** Let $p\colon K \to L$ be a simplicial map of complexes. If $s$ is a simplex in $L$, then its ***inverse image*** is

$$p^{-1}(s) = \{\text{all simplexes } s' \text{ in } K : p(s') = s\}.$$

The inverse image $p^{-1}(s)$ is a subcomplex of $K$.

**Definition.** A ***covering complex*** of a connected complex $K$ is an ordered pair $(\widetilde{K}, p)$, where $\widetilde{K}$ is a connected complex and $p\colon \widetilde{K} \to K$ is a simplicial map, such that

(i) $p$ is a surjection;

(ii) for each simplex $s$ in $K$, the inverse image $p^{-1}(s)$ is a disjoint union $\bigcup_{i\in I} \widetilde{s}_i \subseteq \widetilde{K}$ of simplexes with $p|\widetilde{s}_i\colon \widetilde{s}_i \to s$ an isomorphism for all $i$.

All the ingredients have names: $p$ is the ***projection***; $\widetilde{K}$ is a ***covering***; $K$ is the ***base***, the $\widetilde{s}_i$ are called the ***sheets*** over $s$, and $p^{-1}(v)$ is called the ***fiber*** over $v$, where $v \in \mathrm{Vert}(K)$. Instead of saying that an ordered pair $(\widetilde{K}, p)$ is a covering complex, we usually write $p\colon \widetilde{K} \to K$ is a covering complex.

Since, for every simplex $s$ in $K$, each sheet $\widetilde{s}_i \cong s$, it follows that

$$\dim(\widetilde{K}) = \dim(K). \tag{1}$$

It is true (Theorem C-1.159 below) that if $(K, w)$ is a connected complex and $H \subseteq \pi_1(K, w)$ is any subgroup, then there exists a covering complex $p\colon \widetilde{K}_H \to K$ with $\pi_1(\widetilde{K}, \widetilde{w}) \cong H$. The analog of this result for a topological space $X$ requires more hypotheses; $X$ must be connected, locally path connected, and semilocally 1-connected (Rotman [**191**], Theorem 10.36). These hypotheses necessarily hold for the geometric realization $|K|$ of a connected complex $K$.

**Theorem C-1.153 (Unique Path Lifting).** *Let $p\colon \widetilde{K} \to K$ be a covering complex, let $w$ be a basepoint in $K$, and let $\widetilde{w} \in p^{-1}(w)$.*

(i) *Given a path $\alpha$ in $K$ with origin $w$, there is a unique path $\widetilde{\alpha}$ in $\widetilde{K}$ with origin $\widetilde{w}$ and $p\widetilde{\alpha} = \alpha$. We call $\widetilde{\alpha}$ the **lifting of** $\alpha$ **at** $\widetilde{w}$.*

(ii) *If $\alpha$ and $\beta$ are homotopic paths in $K$ with origin $w$, then their liftings $\widetilde{\alpha}, \widetilde{\beta}$ with origin $\widetilde{w}$ are homotopic and $t(\widetilde{\alpha}) = t(\widetilde{\beta})$.*

**Proof.**

(i) The proof is by induction on the length $n$ of $\alpha$. If $n = 1$, then $\alpha = (w, v)$, where $s = \{w, v\}$ is a simplex; we may assume that $w \neq v$; that is, $s$ is a 1-simplex. If $\widetilde{s}$ is the sheet over $s$ containing $\widetilde{w}$, then $\widetilde{v} \in \widetilde{s}$ (because $\widetilde{s} \cong s$), and so $\widetilde{\alpha} = (\widetilde{w}, \widetilde{v})$ is a lifting of $\alpha$. The inductive step $n > 1$ is routine.

To prove uniqueness, suppose again that $\alpha = (w, v)$ has length 1. If $\alpha$ has another lifting $(\widetilde{w}, \widetilde{u})$, then $\widetilde{s'} = \{\widetilde{w}, \widetilde{u}\}$ is a 1-simplex. Both $\widetilde{s}$ and $\widetilde{s'}$ are sheets over $s$, but they are not disjoint, a contradiction. For the inductive

step, let $\alpha = (w, v)\beta$, where $\beta$ is a path of length $n-1$ beginning at $v$. By the base step, there is a unique $(\widetilde{w}, \widetilde{v})$ lifting $(w, v)$; by the inductive hypothesis, there is a unique lifting $\widetilde{\beta}$ of $\beta$ beginning at $\widetilde{v}$, and it is easy to see that the lifting $\widetilde{\alpha} = (\widetilde{w}, \widetilde{v})\widetilde{\beta}$ is unique.

(ii) It suffices to prove that if $(\widetilde{w}, \widetilde{v})(\widetilde{v}, \widetilde{x})$ is a lifting of $(u, v)(v, x)$, where $\{u, v, x\}$ is a simplex in $K$, then $\{\widetilde{u}, \widetilde{v}, \widetilde{x}\}$ is a simplex in $\widetilde{K}$. Let $\widetilde{s} = \{\widetilde{v}, \widetilde{u'}, \widetilde{x'}\}$ be the sheet over $s$ containing $\widetilde{v}$, and let $\widetilde{t} = \{\widetilde{u}, \widetilde{v''}, \widetilde{x''}\}$ be the sheet over $s$ containing $\widetilde{u}$, where $p\widetilde{v} = p\widetilde{v''} = v$ and $p\widetilde{u} = p\widetilde{u''} = u$. Now $(\widetilde{u}, \widetilde{v})$ and $(\widetilde{u}, \widetilde{v''})$ are liftings of $(u, v)$ beginning at $\widetilde{u}$, so that uniqueness of lifting gives $\widetilde{v} = \widetilde{v''}$. Thus, the sheets $\widetilde{s}$ and $\widetilde{t}$ are not disjoint; that is, $\widetilde{s} = \widetilde{t}$, and so $\widetilde{u'} = \widetilde{u}$ and $\widetilde{x'} = \widetilde{x''}$. A similar argument comparing $\widetilde{s}$ with the sheet over $s$ containing $\widetilde{x}$ shows that $\widetilde{x} = \widetilde{x'}$. •

**Theorem C-1.154.** *Let $p\colon \widetilde{K} \to K$ be a covering complex, let $w$ be a basepoint in $K$, and let $\widetilde{w} \in p^{-1}(w)$. Then $p_*\colon \pi_1(\widetilde{K}, \widetilde{w}) \to \pi_1(K, w)$ is an injection.*

**Proof.** Assume that $[A], [B] \in \pi_1(\widetilde{K}, \widetilde{w})$ and $p_*[A] = p_*[B]$; that is, $[pA] = [pB]$. Writing $\alpha = pA$, we have $A$ is a lifting of $\alpha$ beginning at $\widetilde{w}$, and uniqueness of lifting gives $A = \widetilde{\alpha}$. Similarly, writing $\beta = pB$ gives $B = \widetilde{\beta}$. But $[\alpha] = [pA] = [pB] = [\beta]$, so that Theorem C-1.153 gives $A \simeq B$ and $[A] = [B]$. •

What happens to the subgroup $p_*\pi_1(\widetilde{K}, \widetilde{w})$ of $\pi_1(K, w)$ if we change the basepoint $\widetilde{w}$?

**Theorem C-1.155.** *Let $p\colon \widetilde{K} \to K$ be a covering complex, let $w$ be a basepoint in $K$, and let $\widetilde{w} \in p^{-1}(w)$. If $\widetilde{u} \in p^{-1}(w)$, then $p_*\pi_1(\widetilde{K}, \widetilde{w})$ and $p_*\pi_1(\widetilde{K}, \widetilde{u})$ are conjugate subgroups of $\pi_1(K, w)$.*

*Conversely, if $H$ is a subgroup of $\pi_1(K, w)$ which is conjugate to $p_*\pi_1(\widetilde{K}, \widetilde{w})$, then there is $\widetilde{u}$ in the fiber over $w$ with $H = p_*\pi_1(\widetilde{K}, \widetilde{u})$.*

**Proof.** Since $\widetilde{K}$ is connected, there is a path $B$ from $\widetilde{w}$ to $\widetilde{u}$. Then $\beta = pB$ is a closed path at $w$, $[\beta] \in \pi_1(K, w)$, and Exercise C-1.107 on page 109 gives

$$[B^{-1}]\pi_1(\widetilde{K}, \widetilde{w})[B] = \pi_1(\widetilde{K}, \widetilde{u}).$$

Hence,

$$[\beta^{-1}]p_*\pi_1(\widetilde{K}, \widetilde{w})[\beta] = \pi_1(\widetilde{K}, \widetilde{u}).$$

Conversely, assume that $H = [\alpha^{-1}]p_*\pi_1(\widetilde{K}, \widetilde{w})[\alpha]$. If $\widetilde{\alpha}$ is the lifting of $\alpha$ at $\widetilde{w}$ and if $t(\widetilde{\alpha}) = \widetilde{u}$, then $p\widetilde{u} = w$. By Exercise C-1.107,

$$[\widetilde{\alpha}^{-1}]\pi_1(\widetilde{K}, \widetilde{w})[\widetilde{\alpha}] = \pi_1(\widetilde{K}, \widetilde{u}).$$

Hence, $p_*\pi_1(\widetilde{K}, \widetilde{u}) = p_*[\widetilde{\alpha}^{-1}]\pi_1(\widetilde{K}, \widetilde{w})[\widetilde{\alpha}] = H$. •

If $p\colon \widetilde{K} \to K$ is a covering complex and $w$ is a basepoint in $K$, then the fundamental group $\pi_1(K, w)$ acts on the fiber $p^{-1}(w)$.

**Proposition C-1.156.** *If $K$ is a connected complex with basepoint $w$ and $p\colon \widetilde{K} \to K$ is a covering complex, then the fiber $X = p^{-1}(w)$ is a right $\pi_1(K,w)$-set. The action is given by*

$$\widetilde{x}[\alpha] = t(\widetilde{\alpha}),$$

*where $\widetilde{x} \in X$, $[\alpha] \in \pi_1(K,w)$, and $\widetilde{\alpha}$ is the lifting of $\alpha$ at $\widetilde{x}$.*

**Proof.** Theorem C-1.153 shows that the given definition does not depend on the choice of the representative $\alpha$ in the path class $[\alpha]$. We now verify the axioms for a right $\pi_1(K,w)$-set.

The identity in $\pi_1(K,w)$ is $[\varepsilon] = [(w,w)]$; its lifting at $\widetilde{x}$ is obviously $(\widetilde{x},\widetilde{x})$, and its terminus is also $\widetilde{x}$. Thus, $\widetilde{x}[\varepsilon] = \widetilde{x}[(w,w)] = t(\widetilde{(w,w)}) = t((\widetilde{x},\widetilde{x})) = \widetilde{x}$.

Let $[\alpha],[\beta] \in \pi_1(K,w)$, let $\widetilde{\alpha}$ be the lifting at $\widetilde{x}$, and let $\widetilde{y} = t(\widetilde{\alpha})$. If $\widetilde{\beta}$ is the lifting of $\beta$ at $\widetilde{y}$, then $\widetilde{\alpha}\widetilde{\beta}$ is a lifting of $\alpha\beta$ at $\widetilde{x}$. By uniqueness of lifting, $\widetilde{\alpha}\widetilde{\beta}$ is *the* lifting of $\alpha\beta$ at $\widetilde{x}$, and so $\widetilde{x}[\alpha\beta] = t(\widetilde{\alpha\beta}) = t(\widetilde{\alpha}\widetilde{\beta}) = t(\widetilde{\beta}) = \widetilde{y}$. On the other hand, $(\widetilde{x}[\alpha])[\beta] = (t(\widetilde{\alpha}))[\beta] = \widetilde{y}[\beta] = t(\widetilde{\beta}) = \widetilde{y}$. •

**Theorem C-1.157.** *If $p\colon \widetilde{K} \to K$ is a covering complex and $w$ is a basepoint in $K$, then the fiber $X = p^{-1}(w)$ is a transitive right $\pi_1(K,w)$-set and the stabilizer of a point $\widetilde{w}$ is $p_*\pi_1(\widetilde{K},\widetilde{w})$.*

**Proof.** To see that $X$ is transitive, let $\widetilde{x},\widetilde{y} \in p^{-1}(w)$. Since $\widetilde{K}$ is connected, there is a path $A$ in $\widetilde{K}$ from $\widetilde{x}$ to $\widetilde{y}$. If $\alpha = pA$, then $[\alpha] \in \pi_1(K,w)$ and $[\alpha]\widetilde{x} = t(A)$.

The stabilizer of a point $\widetilde{x} \in X$ consists of all $[\alpha] \in \pi_1(K,w)$ for which $t(\widetilde{\alpha}) = \widetilde{x}$. But $t(\widetilde{\alpha}) = \widetilde{x}$ if and only if $[\widetilde{\alpha}] \in \pi_1(\widetilde{K},\widetilde{x})$ if and only if $[\alpha] \in p_*\pi_1(\widetilde{K},\widetilde{x})$. •

**Corollary C-1.158.** *Let $p\colon \widetilde{K} \to K$ be a covering complex.*

(i) *If $w$ is a basepoint in $K$ and $\widetilde{w} \in p^{-1}(w)$, then the index*

$$[\pi_1(K,w) : p_*\pi_1(\widetilde{K},\widetilde{w})] = |p^{-1}(w)|.$$

(ii) *If $w$ and $u$ are basepoints in $K$, then $|p^{-1}(w)| = |p^{-1}(u)|$.*

**Proof.**

(i) The number of elements in a transitive set is the index of the stabilizer of a point.

(ii) If $\widetilde{u} \in p^{-1}(u)$, then there is a path $B$ in $\widetilde{K}$ from $\widetilde{w}$ to $\widetilde{u}$; denote $pB$ by $\beta$. Define homomorphisms $\Phi\colon \pi_1(\widetilde{K},\widetilde{w}) \to \pi_1(\widetilde{K},\widetilde{u})$ by $[A] \mapsto [B^{-1}AB]$, and define $\varphi\colon \pi_1(K,w) \to \pi_1(K,u)$ by $[\alpha] \mapsto [\beta^{-1}\alpha\beta]$. It is easy to check that the following diagram commutes:

$$\begin{array}{ccc} \pi_1(\widetilde{K},\widetilde{w}) & \xrightarrow{\Phi} & \pi_1(\widetilde{K},\widetilde{u}) \\ {\scriptstyle p_*}\downarrow & & \downarrow{\scriptstyle p_*} \\ \pi_1(K,w) & \xrightarrow[\varphi]{} & \pi_1(K,u). \end{array}$$

Since $\Phi$ and $\varphi$ are isomorphisms, it follows that the index of $\operatorname{im} p_*$ on the left is equal to the index of $\operatorname{im} p_*$ on the right. •

We are now going to construct coverings of connected complexes; in fact, we construct simply connected covering complexes.

**Definition.** Let $(K, w)$ be a pointed complex with $K$ connected and let $H$ be a subgroup of $\pi_1(K, w)$. Define a relation $\sim_H$ on the family of all paths in $K$ with origin $w$ by

$$\alpha \sim_H \beta \quad \text{if } t(\alpha) = t(\beta) \text{ and } [\alpha\beta^{-1}] \in H.$$

It is routine to check that $\sim_H$ is an equivalence relation. Denote the equivalence class of such a path $\alpha$ by

$$\operatorname{cls} \alpha,$$

and denote the family of all such classes by $K_H$:

$$K_H = \{\operatorname{cls} \alpha : \alpha \text{ is a path in } K \text{ with } o(\alpha) = w\}.$$

The set $K_H$ can be made into a complex.

**Definition.** Let $s$ be a simplex in $K$. If $\alpha$ is a path in $K$ with $o(\alpha) = w$ and $t(\alpha) \in s$, then a ***continuation of*** $\alpha$ ***in*** $s$ is a path $\beta = \alpha\alpha'$ with $\alpha'$ wholly in $s$; that is, $t(\alpha) = o(\alpha')$ and every edge in $\alpha'$ joins two vertices in $s$. For each simplex $s$ in $K$ and $\operatorname{cls} \alpha \in K_H$, define

$$[s, \operatorname{cls} \alpha] = \{\operatorname{cls} \beta : \beta \text{ is a continuation of } \alpha \text{ in } s\},$$

and define ***simplexes*** in $K_H$ to be the nonempty subsets of all $[s, \operatorname{cls} \alpha]$.

If $\varepsilon = (w, w)$ is the trivial path, then $\varepsilon$ is a path in $K$ with $o(\varepsilon) = w$. Define

$$\widetilde{w} = \operatorname{cls} \varepsilon$$

to be a basepoint in $K_H$.

**Theorem C-1.159.** *If $K$ is a connected complex with basepoint $w$ and $H$ is a subgroup of $\pi_1(K, w)$, then there exist a covering complex $p\colon \widetilde{K}_H \to K$ and $\widetilde{w} \in p^{-1}(w)$ with $p_*\pi_1(\widetilde{K}_H, \widetilde{w}) = H$.*

**Proof.** Define a function $p\colon K_H \to K$ by $\operatorname{cls} \alpha \mapsto t(\alpha)$; note that $p\widetilde{w} = w$. The proof is a series of verifications: $K_H$ is connected; $p$ is a surjective simplicial map; $(K_H, p)$ is a covering complex; $p_*\pi_1(K_H, \widetilde{w}) = H$. For details of a proof, see Rotman [**188**], Theorem 11.43. •

**Corollary C-1.160.** *Let $(K, w)$ be a connected pointed complex, and let $H \subseteq \pi_1(K, w)$. If $p\colon K_H \to K$ is the covering complex in* Theorem C-1.159*, then $\pi_1(K, w)$ acts on $K_H$.*

**Proof.** If $[\gamma] \in \pi_1(K, w)$ and $\operatorname{cls} \alpha \in K_H$, define

$$[\gamma]\operatorname{cls} \alpha = \operatorname{cls} (\gamma\alpha).$$

The reader should check that this is a well-defined (left) action. •

The theory of covering spaces for topological spaces was well known when the Baer–Levi proof was given .

**Theorem C-1.161 (Schreier–Nielsen Redux).** *Every subgroup $H$ of a free group $F$ is itself free.*

**Proof (Baer–Levi).** Exercise C-1.116 on page 110 says that there is a bouquet of circles $K$ with $F \cong \pi_1(K, w)$, and we may assume that $H \subseteq \pi_1(K, w)$. If $p\colon K_H \to K$ is the corresponding covering complex, then $H = p_*\pi_1(K_H, \widetilde{w})$, by Theorem C-1.154. Now $\dim(K_H) = \dim(K)$, by Eq. (1) on page 111. But $\dim(K) = 1$, so that $\dim(K_H) = 1$, and Corollary C-1.151 says that $H$ is a free group. •

Here is another proof of Corollary C-1.139.

**Proposition C-1.162.** *If $F$ is a free group of finite rank $n$ and $H$ is a subgroup of finite index $j$, then* $\operatorname{rank}(H) = jn - j + 1$.

**Proof.** If $K = B_n$ is a bouquet of $n$ circles, Exercise C-1.116 on page 110 says that $\sigma_1(B_n) = 3n$ and $\sigma_0(B_n) = 2n + 1$, where $\sigma_q(K)$ is the number of $q$-simplexes in $K$; moreover, $n = \operatorname{rank}(F) = \operatorname{rank}(\pi_1(K, w)) = -\chi(H) + 1$. If $p\colon K_H \to K$ is the covering complex from $H$, then Exercise C-1.119 gives $\sigma_1(K_H) = 3jn$ and $\sigma_0(K_H) = 2jn + j$; hence, $\chi(K_H) = -jn + j$. Since $H$ is free, we have $\operatorname{rank}(H) = \operatorname{rank}(\pi_1(K_h, \widetilde{w})) = -\chi(K_H) + 1 = jn - j + 1$. •

## Exercises

∗ **C-1.117.** Let $p\colon \widetilde{K} \to K$ be a covering complex. If $L$ is a connected subcomplex of $K$ and $\widetilde{L_1}$ is a component of $p^{-1}(L)$, prove that $p|\widetilde{L_1}\colon \widetilde{L_1} \to L$ is a covering complex.

∗ **C-1.118.** Let a tree $T$ be a subcomplex of a connected complex $K$.

(i) If $p\colon \widetilde{K} \to K$ is a covering complex, prove that $p^{-1}(T)$ is a disjoint union of trees.

(ii) Use Exercise C-1.113 on page 110 to prove that there exists a maximal tree in $\widetilde{K}$ containing $p^{-1}(T)$.

∗ **C-1.119.** Let $p\colon \widetilde{K} \to K$ be a covering complex. If the fiber $p^{-1}(v)$ over a vertex $v$ is finite, say, $j = |p^{-1}(v)|$, prove that there are exactly $j$ sheets in $\widetilde{K}$ over every simplex in $K$. Conclude that

$$\chi(\widetilde{K}) = j\chi(K).$$

### ▪ Co-Galois Theory

An analog of Galois theory emerges when we try to classify all the covering complexes of a connected complex $K$. We begin by comparing two covering complexes of $K$.

**Proposition C-1.163.** *Let $p\colon (\widetilde{K_1}, \widetilde{w_1}) \to (K, w)$ and $q\colon (\widetilde{K_2}, \widetilde{w_2}) \to (K, w)$ be covering complexes, where $(K, w)$ is a connected pointed complex and $p\widetilde{w_1} = w = q\widetilde{w_2}$.*

(i) *If*

$$q_*\pi_1(\widetilde{K_2}, \widetilde{w_2}) \subseteq p_*\pi_1(\widetilde{K_1}, \widetilde{w_1}),$$

*then there exists a unique simplical map $h\colon \widetilde{K_2} \to \widetilde{K_1}$ making the following diagram commute:*

$$\begin{array}{ccc} (\widetilde{K_2}, \widetilde{w_2}) & \overset{h}{\longrightarrow} & (\widetilde{K_1}, \widetilde{w_1}) \\ & {\scriptstyle q}\searrow & \downarrow {\scriptstyle p} \\ & & (K, w). \end{array}$$

(ii) *$h\colon (\widetilde{K_2}, \widetilde{w_2}) \to (\widetilde{K_1}, \widetilde{w_1})$ is a covering complex.*

(iii) *If $q_*\pi_1(\widetilde{K_2}, \widetilde{w_2}) = p_*\pi_1(\widetilde{K_1}, \widetilde{w_1})$, then $h$ is an isomorphism.*

**Proof.** Rotman [**190**], Theorem 3.3 and Corollary 3.4. •

In the diagram above, let us call the covering complex $\widetilde{K_1}$ an ***intermediate*** covering complex between $\widetilde{K_2}$ and $K$. A covering complex $\widetilde{U}$ of a connected complex $K$ is a *universal covering complex* if every covering space of $K$ is intermediate between $\widetilde{U}$ and $K$.

**Definition.** A ***universal covering complex*** of a connected complex $K$ is a covering complex $q\colon \widetilde{U} \to K$ such that, for every covering complex $p\colon \widetilde{K} \to K$, there exists a unique simplicial map $h\colon \widetilde{U} \to K$ making the following diagram commute:

$$\begin{array}{ccc} \widetilde{U} & \overset{h}{\longrightarrow} & \widetilde{K} \\ & {\scriptstyle q}\searrow & \downarrow {\scriptstyle p} \\ & & K. \end{array}$$

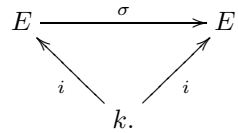
As usual, a solution to a universal mapping problem is unique (via a unique isomorphism) if it exists.

**Theorem C-1.164.** *Every connected complex $K$ has a universal covering complex. Moreover, a covering complex $q\colon \widetilde{K} \to K$ is universal if and only if $\widetilde{K}$ is simply connected.*

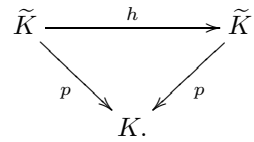
**Proof.** Rotman [**190**], Theorem 3.6. Existence is proved by taking $\widetilde{K} = K_H$, where $H = \{1\}$. •

We note that Theorem C-1.164 may not be true if we replace complexes by topological spaces; a simply connected covering space is universal only if it is semilocally 1-connected.

Recall that if $E/k$ is an extension field, then $\mathrm{Gal}(E/k)$ consists of all automorphisms $\sigma\colon E \to E$ fixing $k$ pointwise. If $i\colon k \to E$ is the inclusion, this says that the following diagram commutes:




**Definition.** Let $p\colon \widetilde{K} \to K$ be a covering complex. A ***covering map*** (or ***deck transformation***) is a simplicial isomorphism making the following diagram commute:




Define

$$\mathrm{Cov}(\widetilde{K}/K) = \{\text{all covering maps } \widetilde{K} \to \widetilde{K}\}.$$

Note that $\mathrm{Cov}(\widetilde{K}/K)$ is a group under composition.

Galois theory classifies intermediate fields of $E/k$ if $E/k$ is a Galois extension.

**Definition.** A covering complex $p\colon \widetilde{K} \to K$ is ***regular*** if there are basepoints $w \in \mathrm{Vert}(K)$ and $\widetilde{w} \in \mathrm{Vert}(\widetilde{K})$ with $p\widetilde{w} = w$ such that $p_*\pi_1(\widetilde{K}, \widetilde{w})$ is a normal subgroup of $\pi_1(K, w)$.

**Theorem C-1.165.** *If $p\colon \widetilde{K} \to K$ is a regular covering complex, then*

$$\mathrm{Cov}(\widetilde{K}/K) \cong \pi_1(K, w)/p_*\pi_1(\widetilde{K}, \widetilde{w}).$$

*In particular, if $\widetilde{K}$ is a universal covering complex, then*

$$\mathrm{Cov}(\widetilde{K}/K) \cong \pi_1(K, w).$$

**Proof.** Rotman [**190**], Theorem 3.9. •

Note that $\mathrm{Cov}(\widetilde{K}/K) \cong \pi_1(K, w)/K)$ describes the fundamental group $\pi_1(\widetilde{K}, \widetilde{w})$ without mentioning basepoints.

Finally, here is an analog of the Fundamental Theorem of Galois Theory.

**Theorem C-1.166.** *If $p\colon \widetilde{K} \to K$ is a regular covering complex, then there is a bijection $\mathcal{S} \to \mathcal{C}$, where $\mathcal{S}$ is the family of intermediate subgroups $G$ with*

$$p_*\pi_1(\widetilde{K}, \widetilde{w}) \subseteq G \subseteq \pi_1(K, w)$$

*and $\mathcal{C}$ is the family of all intermediate covering complexes between $\widetilde{K}$ and $K$.*

**Proof.** See Rotman [**190**], Theorem 3.15, for a more precise statement and its proof. •

## C-1.9. Free Products and the Kurosh Theorem

Given a family $(A_i)_{i\in I}$ of groups, their direct product $P = \prod_{i\in I} A_i$ is the group whose elements are the $I$-tuples $(a_i)$ in the cartesian product (so $a_i \in A_i$) and whose operation is pointwise:

$$(a_i)(a_i') = (a_i a_i').$$

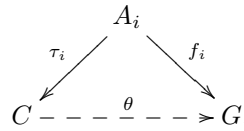
For each $i \in I$, the group $A_i$ is imbedded in the product $P$ as all $I$-tuples $\lambda_i(a_i)$ having $a_i$ in the $i$th coordinate and 1 in every other coordinate. Note that if $i \neq j$, then $\lambda_i(a_i)$ and $\lambda_j(a_j)$ commute. In contrast, the *free product* will be a group $C$ having injections $\tau_i\colon A_i \to C$ in which, for $i \neq j$, the elements $\tau_i(a_i)$ and $\tau_j(a_j)$ do not commute.

Exercise C-1.120 on page 123 says that the direct product $\prod_{i\in I} A_i$ is the categorical product in **Groups**.

**Definition.** Given a family $(A_i)_{i\in I}$ of groups, their ***free product***

$$C = *_{i\in I} A_i$$

is their coproduct in **Groups**; that is, for each $i \in I$, there is an ***injection*** $\tau_i\colon A_i \to C$ such that, for every group $G$ and every family of homomorphisms $(f_i\colon A_i \to G)_{i\in I}$, there exists a unique homomorphism $\theta\colon C \to G$ making the following diagram commute for every $i$:




If the index set $I$ is finite, we usually denote the free product by

$$A_1 * \cdots * A_n.$$

As usual, if free products do exist, then they are unique to isomorphism. Furthermore, the injections $\tau_i\colon A_i \to C$ in the definition of coproduct are actually injections (i.e., one-to-one) here: if $G = A_i$ and $f_i = 1_{A_i}$, then $\theta\tau_i = 1_{A_i}$.

We are obliged to prove that free products exist. The proof is similar to the proof of existence of free groups, and so we give only the highlights.

**Definition.** If $A$ is a group, denote the subset of all its nontrivial elements by $A^{\#}$:

$$A^{\#} = A - \{1\}.$$

Assume that the underlying sets of the $A_i$ are pairwise disjoint (or, if you are fussy, consider their *disjoint union* as defined in Part 1). Define the ***alphabet*** to be $\bigcup_{i\in I} A_i^{\#}$; define a ***word*** to be either the ***empty word*** 1 (which we assume does not lie in the alphabet) or an $n$-tuple $w = (a_1, \ldots, a_n)$ for $n \geq 1$, where each coordinate $a_j$ lies in the alphabet: $a_j \in A_{i_j}^{\#}$ for some $i_j \in I$.

A word $w$ is ***reduced*** if either $w = 1$ (the empty word) or $w = (a_1, \dots, a_n)$, where adjacent coordinates belong to different $A_i^{\#}$'s. We define

$$\mathcal{W}^*$$

to be the set of all words, and we define

$$C = \{\text{all reduced words in } \mathcal{W}^*\}.$$

If $w = (a_1, \dots, a_n)$ and $w' = (b_1, \dots, b_m)$, define ***juxtaposition*** by

$$w \odot w' = (a_1, \dots, a_n) \odot (b_1, \dots, b_m) = (a_1, \dots, a_n, b_1, \dots, b_m).$$

As for free groups, the product of reduced words need not be reduced, and

$$\odot \colon C \times C \to \mathcal{W}^*.$$

**Theorem C-1.167.** *The free product $C = *_{i\in I} A_i$ of a family $(A_i)_{i\in I}$ of groups exists.*

**Proof.** Note that the spelling of every (reduced) word in $\mathcal{W}^*$ is unique.

If a word $w \in \mathcal{W}^*$ is not reduced, then $w \neq 1$ and $w = (a_1, \dots, a_n)$, where two adjacent coordinates $a_j$ and $a_{j+1}$ lie in the same $A_i^{\#}$. There are two types of ***elementary reduction*** $w \to w_1$: if $a_j a_{j+1} \neq 1$, then

$$w_1 = (a_1, \dots, a_{j-1}, a_j a_{j+1}, a_{j+2}, \dots, a_n)$$

has length $n - 1$; if $a_j a_{j+1} = 1$, then

$$w_1 = (a_1, \dots, a_{j-1}, , a_{j+2}, \dots, a_n)$$

has length $n - 2$. A ***reduction*** is a finite sequence of elementary reductions

$$w \to w_1 \to \cdots \to w_r,$$

where $w_r$ is reduced. As in Corollary C-1.118, the function

$$\text{red} \colon \mathcal{W}^* \to C,$$

given by $w \mapsto w_r$, is well-defined, as is the function $C \times C \to C$, given by

$$ww' = \text{red}(w \odot w').$$

That $C$ is a group which is the coproduct of $(A_i)_{i\in I}$ is proved by adapting the proof of Theorem C-1.119. In particular, if $G$ is a group, $(f_i \colon A_i \to G)_{i\in I}$ is a family of homomorphisms, and $w = (a_1, \dots, a_n)$, where $a_j \in A_{i_j}^{\#}$, then

$$\theta \colon (a_1, \dots, a_n) \mapsto f_{i_1}(a_1) \cdots f_{i_n}(a_n). \quad \bullet$$

**Corollary C-1.168.** *Pushouts exist in the category* **Groups**.

**Proof.** The proof that pushouts exist in ${}_R\mathbf{Mod}$, Proposition B-4.13 in Part 1, constructed the pushout as a quotient of a direct sum. But the direct sum of modules is their coproduct, and that proof can be adapted here to show the pushout is a quotient of a free product. $\bullet$

**Corollary C-1.169.** *The free group $F$ with basis $X$ is a free product of copies of $\mathbb{Z}$.*

**Proof.** For each $x \in X$, define the infinite cyclic group $\mathbb{Z}_x = \langle x \rangle \subseteq F$. Since $F$ is free with basis $X$, if $G$ is a group and $f\colon X \to G$ is a function, there exists a unique homomorpism $\theta\colon F \to G$ with $\theta(x) = f(x)$ for all $x \in X$. Now define injections $\tau_x\colon \mathbb{Z}_x \to G$ by $\tau_x = \theta|\mathbb{Z}_x$, and we see that $F \cong *_{x\in X}\, \mathbb{Z}_x$. •

**Corollary C-1.170 (Normal Form).** *Every $g \in *_{i\in I}\, A_i$ with $g \neq 1$ has a unique factorization*

$$g = a_1 \cdots a_n,$$

*where each factor $a_k \neq 1$ lies in some $A_i$ and adjacent factors lie in distinct $A_i$'s.*

**Proof.** The spelling of every (reduced) word in $\mathcal{W}^*$ is unique. •

**Theorem C-1.171.** *Let $(A_i)_{i\in I}$ be a family of groups. If $(X_i \mid \Delta_i)$ is a presentation of $A_i$, then a presentation of $*_{i\in I}\, A_i$ is*

$$*_{i\in I}\, A_i = \left( \bigcup_{i\in I} X_i \mid \bigcup_{i\in I} \Delta_i \right).$$

**Proof.** By Exercise C-1.122 on page 123 below, if $(F_i)_{i\in I}$ is a family of free groups with $F_i$ having basis $X_i$, then $F = *_{i\in I}\, F_i$ is a free group with basis $\bigcup_{i\in I} X_i$. Let $\{\tau_i\colon A_i \to *_{i\in I}\, A_i\}$ be the injections. If $N_i$ is the normal subgroup of $F_i$ generated by the relations $\Delta_i$ and if $\nu_i\colon F_i \to A_i = F_i/N_i$ is the natural map (so $\ker \nu_i = N_i$), then the map $\varphi\colon F \to *_{i\in I}\, A_i$ extending all composites $F_i \to A_i \to *_{i\in I}\, A_i$ has kernel the normal subgroup generated by all $N_i$, which is the normal subgroup generated by $\bigcup_{i\in I} \Delta_i$. •

Another proof of the existence of free products can now be given by showing that the group with the presentation given in Theorem C-1.171 is a free product.

We only state the next result.

**Theorem C-1.172 (Grushko).** *If $F$ is a free group and $\varphi\colon F \to *_{i\in I}\, A_i$ is a surjective homomorphism, then $F = *_{i\in I}\, F_i$, where $\varphi(F_i) \subseteq A_i$ for all $i$.*

An algebraic proof of this theorem is in Kurosh [**130**], pp. 57–70, but a proof via covering spaces due to Stallings [**211**] can be found in Massey [**149**], pp. 225–233. The review of Stallings's proof in *Mathematical Reviews* says, "The terrifying cancellation arguments and inductions within inductions within inductions, to which one was subjected in the past, can now be entirely avoided."

Here is one consequence of Grushko's Theorem. If $A$ is a finitely generated group, let $\mu(A)$ denote the smallest number of elements that can generate $A$. Your first guess is that

$$\mu(A * B) = \mu(A) + \mu(B).$$

This follows from Grushko's Theorem, and it is very difficult to prove directly.

We are now going to prove, using covering complexes, that every subgroup of a free product is itself a free product. The reader can better appreciate this proof after looking at the terrifying proof of the Kurosh Theorem in Kurosh [**130**], pp. 17–26.

**Lemma C-1.173.** *Let $(K, w)$ be a pointed connected complex having connected subcomplexes $(K_i)_{i\in I}$ such that $K = \bigcup_{i\in I} K_i$. If there is a tree $T$ in $K$ with $T = K_i \cap K_j$ for all $i \neq j$, then*

$$\pi_i(K, w) \cong *_{i\in I}\, \pi_1(K_i, w_i)$$

*for vertices $w_i$ in $K_i$.*

**Proof.** For each $i$, choose a maximal tree $T_i$ in $K_i$ containing $T$. By Exercise C-1.113 on page 110, $T^* = \bigcup_{i\in I} T_i$ is a tree in $K$; by Proposition C-1.149, $T^*$ is a maximal tree because it contains every vertex of $K$.

By Tietze's Theorem, $\pi_1(K_i, w_i)$ has a presentation $(E_i \mid R_1^i \cup R_2^i)$, where $E_i$ is the set of all edges in $K_i$, $R_1^i$ consists of relations $(u, v) = 1$ if $(u, v) \in T_i$, and $R_2^i$ consists of relations $(u, v)(v, x) = (u, x)$ if $\{u, v, x\}$ is a simplex in $K_i$. Another application of Tietze's Theorem gives a presentation

$$\pi_1(K, w) = \left( \bigcup_{i\in I} E_i \mid \bigcup_{i\in I} (R_1^i \cup R_2^i) \right).$$

It now follows from Theorem C-1.171 that $\pi_1(K, w) \cong *_{i\in I}\, \pi_1(K_i, w_i)$. •

**Theorem C-1.174 (Kurosh).** *If $H$ is a subgroup of the free product $*_{i\in I}\, A_i$, then*

$$H = F * (*_{\lambda\in\Lambda} H_\lambda),$$

*where $F$ is a free group, $\Lambda$ is some (possibly empty) index set, and each $H_\lambda$ is a conjugate of a subgroup of some $A_i$.*

**Proof.** Choose connected pointed complexes $(K_i, w_i)$ with $A_i \cong \pi_1(K_i, w_i)$, and let $X$ be the disjoint union $\bigcup_{i\in I} \mathrm{Vert}(K_i)$. Define a new pointed complex $(K, w)$ as follows: $\mathrm{Vert}(K) = X \cup \{w\}$ for some $w \notin X$; $\{a, b, c\}$ is a 2-simplex in $K$ if and only if it a simplex in some $K_i$; $\{a, b\}$ is a 1-simplex in $K$ if it is a 1-simplex in some $K_i$ or it has the form $\{w, w_i\}$ for $i \in I$. If $T$ is the tree in $K$ consisting of all the vertices $\{w, w_i\}$, then Lemma C-1.173 gives

$$\pi_1(K, w) \cong *_{i\in I}\, \pi_1(K_i \cup T, w_i).$$

But Tietze's Theorem gives $\pi_1(K_i \cup T, w_i) \cong \pi_1(K_i, w_i) \cong A_i$ for all $i$; hence

$$\pi_1(K, w) \cong *_{i\in I}\, A_i.$$

Identify $\pi_1(K, w)$ with $*_{i\in I}\, A_i$, let $p\colon K_H \to K$ be the covering complex corresponding to the subgroup $H$, and choose $\widetilde{w} \in p^{-1}(w)$ with $p_*\pi_1(K_H, \widetilde{w}) = H$.

Now $p^{-1}(K_i)$ may not be connected, but it is the disjoint union of its components $\widetilde{K}_{ij}$:[25]

$$p^{-1}(K_i) = \bigcup_{ij} \widetilde{K}_{ij}.$$

Choose a maximal tree $\widetilde{T}_{ij}$ in $\widetilde{K}_{ij}$, and define a 1-complex in $K_H$ by

$$\widetilde{L} = p^{-1}(T) \cup \bigcup_{ij} \widetilde{T}_{ij}.$$

Finally, let $\widetilde{T}$ be a maximal tree in $\widetilde{L}$ containing $\bigcup_{ij} \widetilde{T}_{ij}$, which exists by Exercise C-1.118 on page 115. Observe that $\widetilde{T} \cap \widetilde{K}_{ij} = \widetilde{T}_{ij}$ lest we violate the maximality of $\widetilde{T}_{ij}$ in $\widetilde{K}_{ij}$.

For all $i, j$, consider the subcomplexes $\widetilde{K}_{ij} \cup \widetilde{T}$ of $K_H$. Clearly, $K_H$ is the union of $\widetilde{L}$ and all of these, while the intersection of any pair of them is the tree $\widetilde{T}$. By Lemma C-1.173,

$$\pi_1(K_H, \widetilde{w}) \cong \pi_1(\widetilde{L}, \widetilde{w}) * \big( *_{ij}\ \pi_1(\widetilde{K}_{ij} \cup \widetilde{T}, \widetilde{w}_{ij}) \big),$$

where $\widetilde{w}_{ij} \in p^{-1}(w) \cap \widetilde{K}_{ij}$. Now $\pi_1(\widetilde{L}, \widetilde{w})$ is free, because $\dim(\widetilde{L}) = 1$. Since $\widetilde{T}$ is a maximal tree in $\widetilde{K}_{ij} \cup \widetilde{T}$, Tietze's Theorem gives $\pi_1(\widetilde{K}_{ij} \cup \widetilde{T}, \widetilde{w}_{ij}) \cong \pi_1(\widetilde{K}_{ij}, \widetilde{w}_{ij})$ for all $i, j$. By Exercise C-1.117 on page 115, $p|\widetilde{K}_{ij} \colon \widetilde{K}_{ij} \to K_i$ is a covering complex, and so $p_*\pi_1(\widetilde{K}_{ij}, \widetilde{w}_{ij})$ is isomorphic, via $(p|\widetilde{K}_{ij})_*$, to a subgroup of $\pi_1(K_i, w)$. Finally, Theorem C-1.155 shows that this subgroup is a conjugate of a subgroup of $\pi_1(K_i, w)$. •

**Corollary C-1.175 (Schreier–Nielsen Again).** *Every subgroup of a free group is itself free.*

**Proof.** This is a special case of the Kurosh Theorem, for a free group is a free product of infinite cyclic groups. •

**Corollary C-1.176.** *Let a group $G$ be a free product $*_{i \in I} A_i$.*

(i) *If each $A_i$ is torsion, then every torsion-free subgroup of $G$ is a free group.*

(ii) *Every finite subgroup of $G$ is conjugate to a subgroup of some $A_i$. In particular, every element of finite order in $G$ is conjugate to an element of finite order is some $A_i$.*

**Proof.**

(i) By the Kurosh Theorem, every nontrivial subgroup $H$ of $G$ is the free product of a free group and conjugates of subgroups of the $A_i$. But the latter groups all have elements of finite order, and so none of them occurs as a subgroup of $H$; that is, $H$ is free.

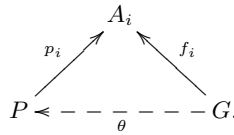
(ii) Every nontrivial free product contains elements of infinite order. •

[25] The index $j$ depends on $i$, but more accurate notation would only make things look more complicated without enlightening anyone.

This last corollary shows that the Sylow Theorems can fail for infinite groups. If $A$ and $B$ are any two finite $p$-groups, then each is a maximal $p$-subgroup of $G = A * B$. For example, if $|A| = p$ and $|B| = p^2$, then both $A$ and $B$ are Sylow $p$-subgroups of $G$; obviously, $A$ and $B$ are not isomorphic, let alone conjugate.

## Exercises

∗ **C-1.120.** Given a family $(A_i)_{i\in I}$ of groups, their ***direct product*** is the cartesian product $P = \prod_{i\in I} A_i$ with operation $(a_i)(a_i') = (a_i a_i')$. For each $j \in I$, define the ***projection*** $p_j\colon P \to A_j$ by $p_j\colon (a_i) \mapsto a_j$. Prove that $P$ is the categorical product in **Groups**; that is, for every group $G$ and every family of homomorphisms $(f_i\colon G \to A_i)_{i\in I}$, there exists a unique homomorphism $\theta\colon G \to P$ making the following diagram commute for every $i$:




**C-1.121.** Prove that pullbacks exist in **Groups**.

∗ **C-1.122.** Let $(F_I)_{i\in I}$ be a family of free groups, where $F_i$ has basis $X_i$. Prove that $*_{i\in I} F_i$ is a free group with basis $\bigcup_{i\in I} X_i$.

∗ **C-1.123.** (i) If $A$ and $B$ are nontrivial groups, prove that $A * B$ contains elements of infinite order.

(ii) Prove that $A * B$ is an infinite centerless group; that is, $Z(A * B) = \{1\}$.

(iii) Prove that every group can be imbedded in a centerless group. (In Exercise C-3.3 on page 235, this result will be used to prove that there are no injectives in the category **Groups** other than $\{1\}$.)

**C-1.124.** If $A$ and $B$ are groups, prove that $(A * B)/N \cong B$, where $N$ is the normal subgroup generated by $A$.

**C-1.125.** Prove that the ***infinite dihedral group*** $D_\infty$, defined by

$$D_\infty = \mathbb{Z}_2 * \mathbb{Z}_2,$$

has a presentation $(s, t \mid t^2 = 1, tst^{-1} = s^{-1})$.

**C-1.126.** Recall the ***modular group*** $M$ defined in Exercise A-4.88 on page 173 in Part 1: let $A, B \in \mathrm{GL}(2, \mathbb{Q})$, where

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix},$$

and define $M = E/\{\pm I\}$, where $E$ is the subgroup generated by $A$ and $B$.

(i) Prove that $M \cong \mathrm{SL}(2, \mathbb{Z})/\{\pm I\} = \mathrm{PSL}(2, \mathbb{Z})$.

(ii) Prove that $\mathrm{PSL}(2, \mathbb{Z}) \cong LF(\mathbb{Z})$, the group of all linear fractional transformations

$$z \mapsto \frac{az + b}{cz + d},$$

where $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$.

(iii) Prove that $\mathrm{PSL}(2,\mathbb{Z})$ has a presentation $(a, b \mid a^2 = 1 = b^3)$.
**Hint.** Use Exercise A-4.30 on page 138 in Part 1.

(iv) Prove that $\mathrm{PSL}(2,\mathbb{Z}) \cong \mathbb{Z}_2 * \mathbb{Z}_3$.

(v) Prove that $\mathrm{PSL}(2,\mathbb{Z})$ has a normal subgroup of index 6.

**C-1.127.** Prove that no group $G$ is both a direct product and a free product; that is, there do not exist nontrivial groups $A, B, C, D$ such that $G = A * B$ and $G = C \times D$.

**Hint.** (**P. M. Neumann**) If $G = A * B$ and $a \in A$ and $b \in B$ are nontrivial, then the centralizer $C_G(ab) \cong Z$. If $G = C \times D$, choose nontrivial elements $c \in C$ and $d \in D$ with $ab = cd$, and show that $C_G(ab) = C_G(cd)$ is a direct product.

**C-1.128.** Use Grushko's Theorem to prove that if $A$ and $B$ are finitely generated groups, then

$$\mu(A * B) = \mu(A) + \mu(B),$$

where $\mu(A)$ denotes the smallest number of elements that can generate $A$.

## C-1.10. Epilog

Here is a bit of history. Groups were invented by Galois around 1830; in modern terminology, they were subgroups of the symmetric group $S_X$, where $X$ is the set of roots of a polynomial. Permutations and products of them, but not *groups* of them, had been studied earlier by Lagrange in 1770, Ruffini in 1799, Cauchy in 1815, and Abel in 1824. For this reason, finite groups were the focus of those first studying groups. Cayley published axioms for (not necessarily finite) groups in the 1850s. The first group theory book [**117**], written by Jordan in 1870, discussed abstract groups, but in the context of Galois theory (then called theory of equations). When infinite groups were studied at that time, by Riemann and Lie, for example, they were usually topological groups. Free groups were introduced by von Dyck, in the 1880s, to treat generators and relations. Representation theory began toward the end of the century, with work of Frobenius, Burnside, and Schur.

In the twentieth century, there continued to be two streams of research in groups: finite groups and infinite groups, with topological groups still being the most popular type of infinite group. Finite groups were the most studied groups; the crowning achievement was the classification of all finite simple groups in the 1990s and early 2000s. But infinite discrete nonabelian groups were also studied. The connection with logic and computing, using presentations of groups, was called ***combinatorial group theory***; it investigates properties of groups following from constraints on their presentations. For example, can a finite group have a presentation with the same number of generators as relations? One of the most remarkable results is the unsolvability of the word problem. A group $G$ has a ***solvable word problem*** if it has a presentation $G = (X \mid R)$ for which there exists an algorithm (i.e., a Turing machine) which determines whether an arbitrary word $w$ on $X$ is equal to the identity element in $G$ (if $X$ and $R$ are finite, it can be proved that this property is independent of the choice of presentation). In the late 1950s, Novikov

and Boone, independently, proved that there exists a finitely presented group $G$ that does not have a solvable word problem (Rotman [**188**], p. 431). Other problems involve finding presentations for known groups, as we have done for $\mathbf{Q}_n$ and $D_{2n}$; such questions are treated in Coxeter–Moser [**46**] and Johnson [**116**].

Another problem is whether a group defined by a presentation is finite or infinite. For example, ***Burnside's problem*** asks whether a finitely generated group $G$ of ***finite exponent*** $m$, that is, $x^m = 1$ for all $x \in G$, must be finite (Burnside had proved that if such a group $G$ happens to be a subgroup of $\mathrm{GL}(n, \mathbb{C})$ for some $n$, then $G$ is finite (Robinson [**181**], p. 221)). However, the answer is negative in general; such a group can be infinite. This was first proved for $m$ odd and large, in 1968, by Novikov and Adian, in a long and complicated paper [**2**]. Using a geometric technique involving *van Kampen diagrams*, Ol′shanskii gave a much shorter and simpler proof in 1982 (Ol′shanskii [**171**]). Finally, Ivanov [**109**] completed the solution by showing that the presented group can be infinite when $m$ is even and large. It is an open question whether a *finitely presented* group of finite exponent must be finite.

The interaction between presentations and algorithms is both theoretical and practical. A theorem of G. Higman (Rotman [**188**], p. 451) states that a finitely generated group $G$ can be imbedded as a subgroup of a finitely presented group $H$ if and only if $G$ is *recursively presented*: there is a presentation of $G$ whose relations can be given by an algorithm. On the practical side, many efficient algorithms solving group-theoretic problems have been implemented (Sims [**207**]). The first such algorithm was ***coset enumeration*** (see Lyndon–Schupp [**142**], pp. 163–167) which computes the order of a group $G$ defined by a presentation, provided that $|G|$ is finite (unfortunately, there can be no algorithm to determine, in advance, whether $G$ is finite (Rotman [**188**], p. 469)).

Two major areas of interest today arose from work of Gromov in the 1980s analyzing Cayley graphs: one investigates the influence of ***growth functions*** [**83**]; the other is called ***geometric group theory***.

If $G = \langle x_1, \ldots, x_d \rangle$ is a finitely generated group, each element $g \in G$ has an expression $g = x_{i_1}^{e_1} \cdots x_{i_m}^{e_m}$, where $e_{i_j} = \pm 1$ for all $j$. This expression may not be unique, and the number of terms in a shortest one is called the ***length*** of $g$. Define $a_G(n)$ to be the number of elements in $G$ of length $n$, and define $s_G(n)$ to be the number of elements in $G$ of length $\leq n$. The functions $a_G$ and $s_G$ are called *growth functions* of $G$. In general, these functions are not recursive, hence are very difficult (if not impossible) to compute; however, their asymptotic behavior can be studied. In fact, if $G$ is a finitely presented group, then $G$ has a solvable word problem if and only if $a_G$ is recursive ([**148**], p. 9). Another result is that the generating function $\sum_n a_G(n)x^n$ is a rational function if and only if $G$ has a normal abelian subgroup of finite index. In 1968, Milnor and Wolf proved that if a finitely generated group $G$ is *virtually nilpotent*, that is, $G$ has a normal nilpotent subgroup of finite index, then $a_G$ has polynomial growth: $a_G(n) \leq Cn^k$ for some constant $C$ and positive integer $k$. Using the solution to Hilbert's fifth problem, Gromov proved the converse: $G$ is virtually nilpotent if and only if $a_G$ has polynomial growth.

*Geometric group theory* studies connections between algebraic properties of discrete groups and geometric properties of spaces admitting nice actions of such groups. It owes much to the study of ***hyperbolic groups*** introduced by Gromov [**84**]. A hyperbolic group is a finitely presented group having a Cayley graph which, when viewed as a metric space, resembles hyperbolic space. Examples of hyperbolic groups are finite groups, free groups, groups having an infinite cyclic group of finite index, fundamental groups of negatively curved surfaces. It is known that a subgroup of a hyperbolic group need not be hyperbolic; for example, since hyperbolic groups are finitely generated, the commutator subgroup of a free group of rank 2 is not hyperbolic. In 1986, Culler and Vogtmann introduced *outer space*, the boundary of a compactification of a Cayley graph (analogous to Teichmuller spaces in complex variables). They obtained information about automorphism groups of free groups by proving that $\mathrm{Out}(F_n) = \mathrm{Aut}(F_n)/\mathrm{Inn}(F_n)$ acts nicely on outer space; for example, $\mathrm{Out}(F_n)$ has a subgroup of finite index whose *cohomological dimension* is $2n-3$ (see Bridson–Haefliger [**27**], Collins–Grigorchuk–Kurchanov–Zieschang [**43**], and Stillwell [**213**]). These results show that geometric group theory is intimately related to the study of the ***mapping class group*** $\mathcal{M}(S)$, where $S$ is an orientable surface. This rich subject touches on many areas of mathematics, including algebraic geometry, algebraic topology, group theory, and Riemann surfaces (see [**64**]).

# Representation Theory

An abstract group is a "cloud", a capital letter $G$. In contrast, there are familiar concrete groups, such as the symmetric group $S_X$ of all permutations of a set $X$. The idea underlying representation theory is that comparing abstract groups with familiar groups via homomorphisms can yield useful information.[1] For example, Chapter C-1 began by discussing $G$-sets: groups can act on sets. This leads to interesting results such as Cauchy's Theorem: if the order of a finite group $G$ is divisible by a prime $p$, then $G$ has an element of order $p$.

In Theorem C-1.5, we saw that an action of a group $G$ on a set $X$ is merely another way of viewing homomorphisms $G \to S_X$. In this chapter, we will focus on finite groups acting on finite-dimensional vector spaces, that is, homomorphisms $G \to \mathrm{GL}(V)$, and we will see, in Proposition C-2.13, that when a group $G$ acts on a vector space $V$ over a field $k$, then $V$ can be viewed as a $kG$-module. Such representations will yield numerical data that will enable us to prove important theorems of Burnside and of Frobenius.

## C-2.1. Artinian and Noetherian

We begin by investigating noncommutative rings with the goal of understanding the group algebras $kG$. Let us recall some basic facts. There are three types of ideals in a noncommutative ring $R$: left, right, and two-sided. In $\mathrm{Mat}_2(\mathbb{R})$, for example, the equation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} r & 0 \\ s & 0 \end{bmatrix} = \begin{bmatrix} * & 0 \\ * & 0 \end{bmatrix}$$

shows that the "first columns" (that is, the matrices that are 0 off the first column), form a left ideal (the "second columns" also form a left ideal); neither of these left

[1] There are representation theories for other algebraic systems; for example, Lie groups, associative algebras, Lie algebras (see Fulton–Harris [**75**] or Humphreys [**101**]).

ideals is a right ideal. The equation

$$\begin{bmatrix} r & s \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} * & * \\ 0 & 0 \end{bmatrix}$$

shows that the "first rows" (that is, the matrices that are 0 off the first row) form a right ideal (the "second rows" also form a right ideal); neither of these right ideals is a left ideal. The only two-sided ideals are $\{0\}$ and $\mathrm{Mat}_2(\mathbb{R})$.

If $I$ is a left ideal in $R$, then the quotient $R/I$ is a left $R$-module, and if $J$ is a right ideal in $R$, then the quotient $R/J$ is a right $R$-module. If $I$ is a two-sided ideal in $R$, then $R/I$ is a ring.

In Part 1, we considered chain conditions on rings.

**Definition.** A ring $R$ is ***left noetherian*** if it has **ACC** on left ideals: every ascending chain of left ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ stops; that is, there is some $t \geq 1$ with $I_t = I_{t+1} = I_{t+2} = \cdots$.

A ring $R$ is ***left artinian*** if it has **DCC**: every descending chain of left ideals $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$ stops; that is, there is some $t \geq 1$ with $I_t = I_{t+1} = I_{t+2} = \cdots$.

The Hilbert Basis Theorem says that $k[x_1, \ldots, x_n]$ is noetherian if $k$ is a field. There are examples of left noetherian rings that are not right noetherian and of left artinian rings that are not right artinian (Exercises B-1.28 and B-1.30 on page 288 in Part 1). If $k$ is a field, then every finite-dimensional $k$-algebra $A$ is both left and right artinian, for if $\dim_k(A) = n$, then there are at most $n$ strict inclusions in any descending chain of left ideals or of right ideals. In particular, if $G$ is a finite group, then $kG$ is finite-dimensional, and so it is left and right artinian. Similarly, finite-dimensional algebras over a field are left and right noetherian. We conclude that $kG$ has both chain conditions (on the left and on the right) when $k$ is a field and $G$ is a finite group.

Here is Proposition B-1.10 in Part 1.

**Proposition C-2.1.** *The following conditions on a ring $R$ are equivalent.*

(i) *$R$ is left noetherian; that is, $R$ has* ACC *on left ideals.*

(ii) *$R$ satisfies the **left maximum condition**: every nonempty family $\mathcal{F}$ of left ideals in $R$ has a maximal element; that is, there is some $M \in \mathcal{F}$ for which there is no $I \in \mathcal{F}$ with $M \subsetneq I$.*

(iii) *Every left ideal is finitely generated.*

Here is Proposition B-1.18 in Part 1.

**Proposition C-2.2.** *The following conditions are equivalent for a ring $R$.*

(i) *$R$ is left artinian; that is, $R$ has* DCC *on left ideals.*

(ii) *$R$ satisfies the **left minimum condition**: every nonempty family $\mathcal{F}$ of left ideals in $R$ has a minimal element; that is, there is some $M \in \mathcal{F}$ for which there is no $I \in \mathcal{F}$ with $M \supsetneq I$.*

**Proposition C-2.3.** *A left $R$-module $M$ over a ring $R$ has a composition series if and only if $M$ has both chain conditions on submodules.*

**Proof.** Proposition B-1.41 in Part 1. •

**Definition.** A left ideal $L$ in a ring $R$ is a ***minimal left ideal*** if $L \neq (0)$ and there is no left ideal $J$ with $(0) \subsetneq J \subsetneq L$.

A ring need not contain minimal left ideals; for example, $\mathbb{Z}$ has no minimal ideals. Every nonzero ideal $I$ in $\mathbb{Z}$ has the form $I = (n)$ for some nonzero integer $n$, and $I = (n) \supsetneq (2n) \neq (0)$; hence, $I$ is not minimal. However, minimal left ideals do exist in left artinian rings.

Recall that a left $R$-module $L$ (in particular, a left ideal) is ***simple*** if $L \neq \{0\}$ and its only submodules are $\{0\}$ and $L$ itself.

**Proposition C-2.4.**

(i) *Every minimal left ideal $L$ in a ring $R$ is a simple left $R$-module.*

(ii) *If $R$ is left artinian, then every nonzero left ideal $I$ contains a minimal left ideal.*

**Proof.**

(i) A submodule $S$ of $L$ is a left ideal of $R$. If $\{0\} \subsetneq S \subsetneq L$, then $S$ would contradict the minimality of $L$.

(ii) If $\mathcal{F}$ is the family of all nonzero left ideals contained in $I$, then $\mathcal{F} \neq \varnothing$ because $I$ is nonzero. By Proposition B-1.18 in Part 1, $\mathcal{F}$ has a minimal element, and any such element is a minimal left ideal. •

Let $R = \mathrm{Mat}_n(k)$, where $k$ is a division ring. For any $\ell$ between 1 and $n$, let $\mathrm{COL}(\ell)$ denote the $\ell$th columns; that is,

$$\mathrm{COL}(\ell) = \big\{[a_{ij}] \in \mathrm{Mat}_n(k) : a_{ij} = 0 \text{ for all } j \neq \ell\big\}.$$

If $e_1, \ldots, e_n$ is the standard basis of $k^n$ and we identify $R = \mathrm{Mat}_n(k)$ with $\mathrm{End}_k(k^n)$, then $\mathrm{COL}(\ell)$ is identified with

$$\mathrm{COL}(\ell) = \{T\colon k^n \to k^n : T(e_j) = 0 \text{ for all } j \neq \ell\}.$$

**Proposition C-2.5.** *If $k$ is a division ring and $1 \leq \ell \leq n$, then $\mathrm{COL}(\ell)$ is a minimal left ideal in $\mathrm{Mat}_n(k)$.*

**Proof.** If $I \subseteq \mathrm{COL}(\ell)$ is a nonzero left ideal, we must show that $\mathrm{COL}(\ell) = I$; that is, $\mathrm{COL}(\ell) \subseteq I$. We first choose a nonzero $T' \in I \subseteq \mathrm{End}_k(k^n)$. Now $T'(e_\ell) = u \neq 0$; otherwise, $T'$ would kill every basis element (since $T' \in \mathrm{COL}(\ell)$ implies $T'(e_i) = 0$ for $i \neq \ell$) and $T' = 0$, a contradiction.

Take $T \in \mathrm{COL}(\ell)$, and let $T(u) = w$. Since $u \neq 0$ (for $u = T'(e_\ell)$), there is $S \in \mathrm{End}_k(k^n)$ with $S(u) = w$. Observe that

$$ST'(e_i) = \begin{cases} 0 & \text{if } i \neq \ell, \\ S(u) = w & \text{if } i = \ell. \end{cases}$$

Hence, $T = ST'$, because they agree on a basis, and since $I$ is a left ideal, $T \in I$. Therefore, $\mathrm{COL}(\ell) = I$, and $\mathrm{COL}(\ell)$ is a minimal left ideal. •

## C-2.2. Jacobson Radical

The Jacobson radical of a ring $R$ is the analog of the Frattini subgroup in group theory; it is a two-sided ideal whose behavior has an impact on $R$. For example, the class of *semisimple rings*, which contains many group algebras, will be characterized in terms of the Jacobson radical and chain conditions.

**Definition.** If $R$ is a ring, then its ***Jacobson radical*** $J(R)$ is defined to be the intersection of all the maximal left ideals in $R$. A ring $R$ is called ***Jacobson semisimple*** if $J(R) = (0)$.

Strictly speaking, we should call $J(R)$ the left Jacobson radical, for we can obviously define another Jacobson radical: the intersection of all the maximal *right* ideals. However, Proposition C-2.12 below shows that these ideals coincide.

**Example C-2.6.**

(i) The ring $\mathbb{Z}$ is Jacobson semisimple. The maximal ideals in $\mathbb{Z}$ are the nonzero prime ideals $(p)$, and $J(\mathbb{Z}) = \bigcap_{p \text{ prime}} (p) = (0)$.

(ii) Recall that a ring $R$ is (left) ***local*** if it has a unique maximal left ideal $\mathfrak{m}$. If $R$ is a local ring, then $J(R) = \mathfrak{m}$. For example, $R = \{a/b \in \mathbb{Q} : b \text{ is odd}\}$ is such a ring; its unique maximal ideal is

$$\mathfrak{m} = (2) = \{2a/b : b \text{ is odd}\}.$$

(iii) In Proposition C-2.4, we saw that if $R = \mathrm{Mat}_n(k)$, where $k$ is a division ring, then $\mathrm{COL}(\ell)$ is a minimal left ideal, where $1 \leq \ell \leq n$ and

$$\mathrm{COL}(\ell) = \big\{[a_{ij}] \in \mathrm{Mat}_n(k) : a_{ij} = 0 \text{ for all } j \neq \ell\big\}.$$

Let us use these minimal left ideals to construct some *maximal* left ideals in $R$. Define

$$\mathrm{COL}^*(\ell) = \bigoplus_{j \neq \ell} \mathrm{COL}(j);$$

$\mathrm{COL}^*(\ell)$ is a left ideal with $R/\mathrm{COL}^*(\ell) \cong \mathrm{COL}(\ell)$ as left $R$-modules. Since $\mathrm{COL}(\ell)$ is a minimal left ideal, it is a simple left $R$-module, and the Correspondence Theorem shows that $\mathrm{COL}^*(\ell)$ is a maximal left ideal. Therefore,

$$J(R) \subseteq \bigcap_{\ell} \mathrm{COL}^*(\ell) = (0),$$

and so $\mathrm{Mat}_n(k)$ is Jacobson semisimple. ◀

**Proposition C-2.7.** *The following conditions are equivalent for an element $x$ in a ring $R$:*

(i) $x \in J(R)$;

(ii) *for every $r \in R$, the element $1 - rx$ has a left inverse; that is, there is $u \in R$ with $u(1 - rx) = 1$;*

(iii) $xM = (0)$ *for every simple left $R$-module $M$;*

(iv) $x(R/I) = (0)$ *for every maximal left ideal $I$.*

**Proof.**

(i) $\Rightarrow$ (ii). If there is $r \in R$ with $1 - rx$ not having a left inverse, then $R(1 - rx)$ is a proper left ideal, for it does not contain 1. By Exercise B-2.1 on page 318 in Part 1, there is a maximal left ideal $I$ with $1 - rx \in R(1 - rx) \subseteq I$. Now $rx \in J(R) \subseteq I$, because $J(R)$ is a left ideal, and so $1 = (1 - rx) + rx \in I$, a contradiction.

(ii) $\Rightarrow$ (iii). Suppose there is a simple module $M$ for which $xM \neq (0)$; hence, there is $m \in M$ with $xm \neq 0$ (of course, $m \neq 0$). It follows that the submodule $Rxm \neq (0)$, for it contains $1xm$. Since $M$ is simple, it has only one nonzero submodule, namely, $M$ itself, and so $Rxm = M$. Therefore, there is $r \in R$ with $rxm = m$; that is, $(1 - rx)m = 0$. By hypothesis, $1 - rx$ has a left inverse, say, $u(1 - rx) = 1$. Hence, $0 = u(1 - rx)m = m$, a contradiction.

(iii) $\Rightarrow$ (iv). By the Correspondence Theorem, a left $R$-module $M$ is simple if and only if $M \cong R/I$, where $I$ is a maximal left ideal.

(iv) $\Rightarrow$ (i). If $x(R/I) = (0)$, then $x(1 + I) = x + I = I$; that is, $x \in I$. Therefore, if $x(R/I) = (0)$ for every maximal left ideal $I$, then $x \in \bigcap_I I = J(R)$. •

Notice that condition (ii) in Proposition C-2.7 can be restated: $x \in J(R)$ if and only if $1 - z$ has a left inverse for every $z \in Rx$, where $Rx$ is the principal left ideal generated by $x$.

The following result is most frequently used in commutative algebra.

**Corollary C-2.8 (Nakayama's Lemma).** *If $A$ is a finitely generated left $R$-module and $JA = A$, where $J = J(R)$ is the Jacobson radical, then $A = \{0\}$.*

*In particular, let $R$ be a commutative local ring with unique maximal ideal $\mathfrak{m}$. If $A$ is a finitely generated $R$-module with $\mathfrak{m}A = A$, then $A = \{0\}$.*

**Proof.** Let $a_1, \ldots, a_n$ be a generating set of $A$ that is minimal in the sense that no proper subset generates $A$. Since $JA = A$, we have $a_1 = \sum_{i=1}^{n} r_i a_i$, where $r_i \in J$. It follows that

$$(1 - r_1)a_1 = \sum_{i=2}^{n} r_i a_i.$$

Since $r_1 \in J$, Proposition C-2.7 says that $1 - r_1$ has a left inverse, say, $u$, and so $a_1 = \sum_{i=2}^{n} u r_i a_i$. This is a contradiction, for now $A$ can be generated by the proper subset $\{a_2, \ldots, a_n\}$. The second statement follows at once because $J(R) = \mathfrak{m}$ when $R$ is a local ring with maximal ideal $\mathfrak{m}$. •

**Remark.** The hypothesis in Nakayama's Lemma that the module $A$ be finitely generated is necessary. For example, it is easy to check that $R = \{a/b \in \mathbb{Q} : b \text{ is odd}\}$ is a local ring with maximal ideal $\mathfrak{m} = (2)$, while $\mathbb{Q}$ is an $R$-module with $\mathfrak{m}\mathbb{Q} = 2\mathbb{Q} = \mathbb{Q}$. ◄

**Remark.** There are other characterizations of $J(R)$. One such will be given in Proposition C-2.12, in terms of elements having two-sided inverses. Another characterization is in terms of *left quasi-regular* elements: those $x \in R$ for which there exist $y \in R$ with $x + y - yx = 0$. A left ideal is called ***left quasi-regular*** if each of

its elements is left quasi-regular. It can be proved that $J(R)$ is the unique maximal left quasi-regular ideal in $R$ (Lam [**133**], pp. 67–68). ◀

Recall that an *element* $a$ in a ring $R$ is *nilpotent* if $a^m = 0$ for some $m \geq 1$.

**Definition.** A left ideal $I$ in a ring $R$ is ***nilpotent*** if there is some integer $m \geq 1$ with $I^m = (0)$.

The left ideal $I^m$ is the set of all sums of products of the form $a_1 \cdots a_m$, where $a_j \in I$ for all $j$; that is,

$$I^m = \left\{ \sum_i a_{i1} \cdots a_{im} : a_{ij} \in I \right\}.$$

It follows that if $I$ is nilpotent, then every element $a \in I$ is nilpotent; that is, $a^m = 0$ for some $m$. On the other hand, if $a \in R$ is a nilpotent element, it does not follow that $Ra$, the left ideal generated by $a$, is a nilpotent ideal. For example, let $R = \mathrm{Mat}_2(k)$, for some commutative ring $k$, and let $a = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. Now $a^2 = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, but $Ra$ contains $e = \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, which is ***idempotent***: $e \neq 0$ and $e^2 = e$. Hence, $e^m = e \neq 0$ for all $m$, and so $(Re)^m \neq (0)$.

**Corollary C-2.9.** *If $R$ is a ring, then $I \subseteq J(R)$ for every nilpotent left ideal $I$ in $R$.*

**Proof.** Let $I^n = (0)$, and let $x \in I$. For every $r \in R$, we have $rx \in I$, and so $(rx)^n = 0$. The equation

$$(1 + rx + (rx)^2 + \cdots + (rx)^{n-1})(1 - rx) = 1$$

shows that $1 - rx$ is left invertible, and so $x \in J(R)$, by Proposition C-2.7. •

**Proposition C-2.10.** *If $R$ is a left artinian ring, then $J(R)$ is a nilpotent ideal.*

**Proof.** Denote $J(R)$ by $J$ in this proof. The descending chain of left ideals

$$J \supseteq J^2 \supseteq J^3 \supseteq \cdots$$

stops, because $R$ is left artinian; say, $J^m = J^{m+1} = \cdots$. Define $I = J^m$; it follows that $I^2 = I$. We will assume that $I \neq (0)$ and reach a contradiction.

Let $\mathcal{F}$ be the family of all nonzero left ideals $B$ with $IB \neq (0)$; note that $\mathcal{F} \neq \varnothing$ because $I \in \mathcal{F}$. By Proposition C-2.2, there is a minimal element $B_0 \in \mathcal{F}$. Choose $b \in B_0$ with $Ib \neq (0)$. Now

$$I(Ib) = I^2 b = Ib \neq (0),$$

so that $Ib \subseteq B_0 \in \mathcal{F}$, and minimality gives $B_0 = Ib$. Since $b \in B_0$, there is $x \in I \subseteq J = J(R)$ with $b = xb$. Hence, $0 = (1 - x)b$. But $1 - x$ has a left inverse, say, $u$, by Proposition C-2.7, so that $0 = u(1 - x)b = b$, a contradiction. •

The Jacobson radical is obviously a left ideal, being an intersection of left ideals, but it turns out to be a right ideal as well; that is, $J(R)$ is a two-sided ideal. We begin by giving another general source of two-sided ideals other than kernels of ring maps.

**Definition.** If $R$ is a ring and $M$ is a left $R$-module, the ***annihilator*** of $M$ is

$$\operatorname{ann}(M) = \{a \in R : am = 0 \text{ for all } m \in M\}.$$

Let us show that $\operatorname{ann}(M)$ is a two-sided ideal in $R$. Now $\operatorname{ann}(M)$ is a left ideal, for if $am = 0$, then $(ra)m = r(am) = 0$. To see that $\operatorname{ann}(M)$ is a right ideal, let $a \in \operatorname{ann}(M)$, $r \in R$, and $m \in M$. Since $M$ is a left $R$-module, we have $rm \in M$; since $a$ annihilates every element of $M$, we have $a(rm) = 0$. Finally, associativity gives $(ar)m = 0$ for all $m$, and so $ar \in \operatorname{ann}(M)$.

**Corollary C-2.11.**

(i) $J(R) = \bigcap_{\substack{I = \text{maximal} \\ \text{left ideal}}} \operatorname{ann}(R/I)$, *and so* $J(R)$ *is a two-sided ideal in* $R$.

(ii) $R/J(R)$ *is a Jacobson semisimple ring.*

**Proof.**

(i) If $x \in J(R)$, then $xM = \{0\}$ for every simple left $R$-module $M$, by Proposition C-2.7(iii). But $M \cong R/I$ for some maximal left ideal $I$; that is, $x \in \operatorname{ann}(R/I)$. Thus, $x \in \bigcap_{\substack{I = \text{maximal} \\ \text{left ideal}}} \operatorname{ann}(R/I)$.

For the reverse inclusion, if $x \in \bigcap_{\substack{I = \text{maximal} \\ \text{left ideal}}} \operatorname{ann}(R/I)$, then $xM = \{0\}$ for every left $R$-module $M$ of the form $M \cong R/I$ for some maximal left ideal $I$. But every simple left $R$-module has this form. Therefore, $x \in J(R)$.

(ii) First, $R/J(R)$ is a ring, because $J(R)$ is a two-sided ideal, and Exercise C-2.7 on page 135 says that if $I$ is any two-sided ideal in $R$ contained in $J(R)$, then $J(R/I) = J(R)/I$. The result follows if $I = J(R)$. •

We now show that the Jacobson radical could have been defined using right ideals instead of left ideals.

**Definition.** A ***unit*** in a ring $R$ is an element $u \in R$ having a two-sided inverse; that is, there is $v \in R$ with

$$uv = 1 = vu.$$

**Proposition C-2.12.**

(i) *If* $R$ *is a ring, then*

$$J(R) = \{x \in R : 1 + rxs \text{ is a unit in } R \text{ for all } r, s \in R\}.$$

(ii) *If* $R$ *is a ring and* $J'(R)$ *is the intersection of all the maximal right ideals of* $R$, *then* $J'(R) = J(R)$.

**Proof.**

(i) Let $W$ be the set of all $x \in R$ such that $1 + rxs$ is a unit for all $r, s \in R$. If $x \in W$, then setting $s = -1$ gives $1 - rx$ a unit for all $r \in R$. Hence, $1 - rx$ has a left inverse, and so $x \in J(R)$, by Proposition C-2.7. Therefore, $W \subseteq J(R)$. For the reverse inclusion, let $x \in J(R)$. Since $J(R)$ is a two-sided ideal, by Corollary C-2.11, we have $xs \in J(R)$ for all $s \in R$. Proposition C-2.7 says

that $1 - rxs$ is left invertible for all $r \in R$; that is, there is $u \in R$ with $u(1 - rxs) = 1$. Thus, $u = 1 + urxs$. Now $(-ur)xs \in J(R)$, since $J(R)$ is a two-sided ideal, and so $u$ has a left inverse (Proposition C-2.7 once again). On the other hand, $u$ also has a right inverse, namely, $1 - rxs$. By Exercise B-1.19 on page 282 in Part 1, $u$ is a unit in $R$. Therefore, $1 - rxs$ is a unit in $R$ for all $r, s \in R$. Finally, replacing $r$ by $-r$, we have $1 + rxs$ a unit, and so $J(R) \subseteq W$.

(ii) The description of $J(R)$ in (i) is left-right symmetric. After proving right-sided versions of Proposition C-2.7 and Corollary C-2.11, we see that $J'(R)$ can also be described as in (i). We conclude that $J'(R) = J(R)$. •

## Exercises

∗ **C-2.1.** If $k$ is a field and $A$ is a finite-dimensional $k$-algebra, define

$$L = \{\lambda_a \in \mathrm{End}_k(A) : \lambda_a \colon x \mapsto ax\} \text{ and } R = \{\rho_a \in \mathrm{End}_k(A) : \rho_a \colon x \mapsto xa\}.$$

Prove that $L$ and $R$ are $k$-algebras and that there are $k$-algebra isomorphisms

$$L \cong A \text{ and } R \cong A^{\mathrm{op}},$$

where $A^{\mathrm{op}}$ is the opposite ring.

**Hint.** Show that the function $A \to L$, defined by $a \mapsto \lambda_a$, is an injective $k$-algebra map which is surjective because $A$ is finite-dimensional.

∗ **C-2.2.** Let $k$ be a division ring.

(i) Prove that the center $Z(k)$ is a field.

(ii) If $k^\times$ is the multiplicative group of nonzero elements of $k$, prove that $Z(k^\times) = Z(k)^\times$; that is, the center of the multiplicative group $k^\times$ consists of the nonzero elements of $Z(k)$.

∗ **C-2.3.** (i) Let $C$ be a subdivision ring of a division ring $k$. Prove that $k$ is a left vector space over $C$, and conclude that $[k : C] = \dim_C(k)$ is defined (if $k$ is an infinite-dimensional vector space over $C$, we merely say $\dim_C(k) = \infty$).

(ii) If $Z \subseteq C \subseteq D$ is a tower of division rings with $[k : C]$ and $[C : Z]$ finite, prove that $[k : Z]$ is finite and

$$[k : Z] = [k : C][C : Z].$$

**Hint.** If $u_1, \ldots, u_m$ is a basis of $k$ as a left vector space over $C$ and $c_1, \ldots, c_d$ is a basis of $C$ as a left vector space over $Z$, show that the set of all $c_i u_j$ (in this order) is a basis of $k$ over $Z$.

∗ **C-2.4.** (i) **(Peirce[2] Decomposition).** Prove that if $e$ is an idempotent in a ring $R$, then

$$R = Re \oplus R(1 - e).$$

(ii) Let $R$ be a ring having left ideals $I$ and $J$ such that $R = I \oplus J$. Prove that there are idempotents $e \in I$ and $f \in J$ with $1 = e + f$; moreover, $I = Ie$ and $J = Jf$.
**Hint.** Decompose $1 = e + f$, and show that $ef = 0 = fe$.

[2] This unusual spelling is not a misprint.

**C-2.5.** (i) If $R$ is a commutative ring with $J(R) = (0)$, prove that $R$ has no nilpotent elements.

(ii) Give an example of a commutative ring $R$ having no nilpotent elements and for which $J(R) \neq (0)$.

**C-2.6.** Let $k$ be a field and $R = \mathrm{Mat}_2(k)$. Prove that $a = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ is left quasi-regular, but that the principal left ideal $Ra$ is not a left quasi-regular ideal.

∗ **C-2.7.** Prove, for every ring $R$, that if $I$ is any two-sided ideal contained in $J(R)$, then $J(R/I) = J(R)/I$.

**C-2.8.** Prove that $R$ is Jacobson semisimple if and only if $R^{\mathrm{op}}$ is.

**C-2.9.** Let $I$ be a two-sided ideal in a ring $R$. Prove that if $I \subseteq J(R)$, then

$$J(R/I) = J(R)/I.$$

## C-2.3. Group Actions on Modules

We begin by showing the connection between group representations and group algebras.

**Definition.** Let $k$ be a commutative ring. A ***$k$-representation*** of a group $G$ is a homomorphism

$$\sigma\colon G \to \mathrm{Aut}(V),$$

where $V$ is a $k$-module.

The most interesting special case for us is when $k$ is a field (indeed, when $k = \mathbb{C}$), $V$ is a finite-dimensional vector space over $k$, and $\mathrm{Aut}(V) = \mathrm{GL}(V)$.

**Proposition C-2.13.** *Every $k$-representation $\sigma\colon G \to \mathrm{GL}(V)$ equips the $k$-module $V$ with the structure of a left $kG$-module: define scalar multiplication $kG \times V \to V$ by*

$$\Big(\sum_{g\in G} a_g g\Big)v = \sum_{g\in G} a_g \sigma_g(v),$$

*where $v \in V$ and $a_g \in k$. Denote this left $kG$-module by*

$$V^{\sigma}.$$

*Conversely, every left $kG$-module $V$ determines a $k$-representation $\sigma\colon G \to \mathrm{GL}(V)$.*

**Proof.** Given a homomorphism $\sigma\colon G \to \mathrm{GL}(V)$, denote $\sigma(g)\colon V \to V$ by $\sigma_g$, and define an action $kG \times V \to V$ as in the statement. A routine calculation shows that $V$, equipped with this scalar multiplication, is a left $kG$-module.

Conversely, assume that $V$ is a left $kG$-module. If $g \in G$, then $v \mapsto gv$ defines a linear transformation $T_g\colon V \to V$; moreover, $T_g$ is nonsingular, for its inverse is $T_{g^{-1}}$. It is easily checked that the function $\sigma\colon G \to \mathrm{GL}(V)$, given by $\sigma\colon g \mapsto T_g$, is a $k$-representation. •

**Remark.** If $\dim(V) = n$, then $\mathrm{GL}(V)$ contains an isomorphic copy of $S_n$. If $X = v_1, \ldots, v_n$ is a basis of $V$ and $\tau \in S_X$, then there is a unique nonsingular linear transformation $T\colon V \to V$ with $T(v_i) = v_{\tau(i)}$ for all $i$, and $\tau \mapsto T$ is an injective homomorphism $S_X \to \mathrm{GL}(V)$. Therefore, representations $G \to S_X$ can be viewed as $k$-representations; that is, $G$-sets are special cases of $kG$-modules. ◀

**Example C-2.14.**

(i) If $G$ is a finite group and $V$ is a vector space over a field $k$, then the ***trivial homomorphism*** $\sigma\colon G \to \mathrm{GL}(V)$ is defined by $\sigma(x) = 1_V$ for all $x \in G$. The corresponding $kG$-module $V^\sigma$ is called the ***trivial*** $kG$-module: if $v \in V$, then $xv = v$ for all $x \in G$. The trivial module $k$ (also called the ***principal*** $kG$***-module***) is denoted by

$$V_0(k).$$

(ii) Given a representation $\sigma\colon G \to \mathrm{GL}(V)$, where $V$ is an $n$-dimensional vector space, then a choice of basis of $V$ gives a ***matrix representation***, a family of $n \times n$ matrices with

$$\{\sigma(g) : g \in G\} = \{A(g) = [a_{ij}(g)] : g \in G\}. \quad ◀$$

If $\sigma, \tau\colon G \to \mathrm{GL}(V)$ are $k$-representations and $V^\sigma, V^\tau$ are the $kG$-modules determined by $\sigma, \tau$ in Proposition C-2.13, when is $V^\tau \cong V^\sigma$?

**Proposition C-2.15.** *Let $G$ be a group and let $\sigma, \tau\colon G \to \mathrm{GL}(V)$ be $k$-representations, where $k$ is a field. If $V^\sigma$ and $V^\tau$ are the corresponding $kG$-modules defined in* Proposition C-2.13*, then $V^\sigma \cong V^\tau$ as $kG$-modules if and only if there exists a nonsingular $k$-linear transformation $\varphi\colon V \to V$ with*

$$\varphi\tau(g) = \sigma(g)\varphi$$

*for every $g \in G$.*

**Remark.** We often say that $\varphi$ ***intertwines*** $\sigma$ and $\tau$. ◀

**Proof.** If $\varphi\colon V^\tau \to V^\sigma$ is a $kG$-isomorphism, then $\varphi\colon V \to V$ is an isomorphism of vector spaces with

$$\varphi\left(\sum a_g g v\right) = \left(\sum a_g g\right)\varphi(v)$$

for all $v \in V$ and all $g \in G$. But the definition of scalar multiplication in $V^\tau$ is $gv = \tau(g)(v)$, while the definition of scalar multiplication in $V^\sigma$ is $gv = \sigma(g)(v)$. Hence, for all $g \in G$ and $v \in V$, we have $\varphi(\tau(g)(v)) = \sigma(g)(\varphi(v))$. Therefore,

$$\varphi\tau(g) = \sigma(g)\varphi$$

for all $g \in G$.

Conversely, the hypothesis gives $\varphi\tau(g) = \sigma(g)\varphi$ for all $g \in G$, where $\varphi$ is a nonsingular $k$-linear transformation, and so $\varphi(\tau(g)v) = \sigma(g)\varphi(v)$ for all $g \in G$ and $v \in V$. It now follows easily that $\varphi$ is a $kG$-isomorphism; that is, $\varphi$ preserves scalar multiplication by $\sum_{g\in G} a_g g$. •

We restate the last proposition in terms of matrices.

**Corollary C-2.16.** *Let $G$ be a group and let $\sigma, \tau\colon G \to \mathrm{Mat}_n(k)$ be $k$-representations. Then $(k^n)^\sigma \cong (k^n)^\tau$ as $kG$-modules if and only if there is a nonsingular $n \times n$ matrix $P$ with*

$$P\tau(x)P^{-1} = \sigma(x)$$

*for every $x \in G$.*

**Proof.** Choose a basis of $V$, view $\sigma$ and $\tau$ as matrix representations, and let $P$ be the matrix of $\varphi$. •

## C-2.4. Semisimple Rings

In Chapter B-2 in Part 1, we introduced the class of *semisimple rings*, and we will now see that this class contains most group algebras $kG$.

Recall that a left $R$-module $M$ (over some ring $R$) is ***simple*** if $M \neq \{0\}$ and its only submodules are $\{0\}$ and $M$ itself; $M$ is ***semisimple*** if it is a direct sum of (possibly infinitely many) simple modules. In particular, a ring $R$ (considered a left module over itself) is ***simple*** if it is not the zero ring and its only left ideals are $\{0\}$ and $R$ itself. In light of Proposition C-2.4, which says that minimal left ideals are simple left $R$-modules, we restate the definition of left semisimple ring.

**Definition.** A *ring* $R$ is ***left semisimple***[3] if it is a direct sum of minimal left ideals.

Although every minimal left ideal is a simple left $R$-module, it is not obvious, conversely, that every simple left $R$-module is isomorphic to a minimal left ideal. This is, in fact, true, and it is proved in Theorem C-2.33 below.

The next result shows that left semisimple rings are direct sums of only finitely many summands.

**Proposition C-2.17.** *Let $R$ be a left semisimple ring.*

(i) *$R$ is a direct sum of finitely many minimal left ideals.*

(ii) *$R$ has both chain conditions on left ideals.*

**Proof.**

(i) This is Lemma B-2.31 in Part 1, but we repeat its short proof here. Since $R$ is left semisimple, it is a direct sum of minimal left ideals: $R = \bigoplus_i L_i$. Let $1 = \sum_i e_i$, where $e_i \in L_i$. If $r = \sum_i r_i \in \bigoplus_i L_i$, then $r = 1r$ and so $r_i = e_i r_i$. Hence, if $e_i = 0$, then $L_i = 0$. We conclude that there are only finitely many nonzero $L_i$; that is, $R = L_1 \oplus \cdots \oplus L_n$.

[3] We can define a ring to be ***right semisimple*** if it is a direct sum of minimal right ideals, but we shall see, in Corollary C-2.36, that a ring is a left semisimple ring if and only if it is right semisimple.

(ii) The series

$$R = L_1 \oplus \cdots \oplus L_n \supseteq L_2 \oplus \cdots \oplus L_n \supseteq \cdots \supseteq L_n \supseteq (0)$$

is a composition series, for the factor modules are $L_1, \ldots, L_n$, which are simple. It follows from Proposition C-2.3 that $R$ (as a left $R$-module over itself) has both chain conditions. •

**Corollary C-2.18.** *The direct product $R = R_1 \times \cdots \times R_m$ of left semisimple rings $R_1, \ldots, R_m$ is also a left semisimple ring.*

**Proof.** This is Corollary B-2.32 in Part 1. •

**Corollary C-2.19.**

(i) *If $R$ is a left semisimple ring, then every left $R$-module $M$ is a semisimple module.*

(ii) *If $I$ is a two-sided ideal in a left semisimple ring $R$, then the quotient ring $R/I$ is also a semisimple ring.*

**Proof.**

(i) There is a free left $R$-module $F$ and a surjective $R$-map $\varphi\colon F \to M$. Now $R$ is a semisimple module over itself (this is the definition of semisimple ring), and so $F$ is a semisimple module (for $F$ is a direct sum of copies of $R$). Thus, $M$ is a quotient of the semisimple module $F$, and so it is itself semisimple, by Corollary B-2.30 in Part 1.

(ii) First, $R/I$ is a ring, because $I$ is a two-sided ideal. The left $R$-module $R/I$ is semisimple, by (i), and so it is a direct sum $R/I \cong \bigoplus S_j$, where the $S_j$ are simple left $R$-modules. But each $S_j$ is also simple as a left $(R/I)$-module, for any $(R/I)$-submodule of $S_j$ is also an $R$-submodule of $S_j$. Therefore, $R/I$ is semisimple. •

It follows that a finite direct product of fields is a commutative semisimple ring (we will prove the converse later). For example, if $n = p_1 \cdots p_t$ is a squarefree integer, then $\mathbb{Z}_n \cong \mathbb{F}_{p_1} \times \cdots \times \mathbb{F}_{p_t}$ is a semisimple ring. Similarly, if $k$ is a field and $f(x) \in k[x]$ is a product of distinct irreducible polynomials, then $k[x]/(f)$ is a semisimple ring.

Left semisimple rings can be characterized in terms of the Jacobson radical.

**Theorem C-2.20.** *A ring $R$ is left semisimple if and only if it is left artinian and Jacobson semisimple; that is, $J(R) = (0)$.*

**Proof.** If $R$ is left semisimple, then there is a left ideal $I$ with $R = J(R) \oplus I$, by Proposition B-2.29 in Part 1. It follows from Exercise C-2.4 on page 134 that there are idempotents $e \in J(R)$ and $f \in I$ with $1 = e + f$. Since $e \in J(R)$, Proposition C-2.7 says that $f = 1 - e$ has a left inverse; there is $u \in R$ with $uf = 1$. But $f$ is an idempotent, so that $f = f^2$. Hence, $1 = uf = uf^2 = (uf)f = f$, so that $e = 1 - f = 0$. Since $J(R)e = J(R)$, by Exercise C-2.4 on page 134, we have $J(R) = (0)$. Finally, Proposition C-2.17(ii) shows that $R$ is left artinian.

Conversely, assume that $R$ is left artinian and $J(R) = (0)$. We show first that if $I$ is a minimal left ideal of $R$, then $I$ is a direct summand of $R$. Now $I \neq (0)$, and so $I \subsetneq J(R)$; therefore, there is a maximal left ideal $A$ not containing $I$. Since $I$ is minimal, it is simple, so that $I \cap A$ is either $I$ or $(0)$. But $I \cap A = I$ implies $I \subseteq A$, a contradiction, and so $I \cap A = (0)$. Maximality of $A$ gives $I + A = R$, and so $R = I \oplus A$.

Choose a minimal left ideal $I_1$, which exists because $R$ is left artinian. As we have just seen, $R = I_1 \oplus B_1$ for some left ideal $B_1$. Now $B_1$ contains a minimal left ideal, say, $I_2$, by Proposition C-2.4, and so there is a left ideal $B_2$ with $B_1 = I_2 \oplus B_2$. This construction can be iterated to produce a strictly decreasing chain of left ideals $B_1 \supsetneq B_2 \supsetneq \cdots \supsetneq B_r$ as long as $B_r \neq (0)$. If $B_r \neq (0)$ for all $r$, then DCC is violated. Therefore, $B_r = (0)$ for some $r$, so that $R = I_1 \oplus \cdots \oplus I_r$ and $R$ is semisimple. •

Note that the chain condition is needed. For example, $\mathbb{Z}$ is Jacobson semisimple, but $\mathbb{Z}$ is not a semisimple ring.

We can now prove the following remarkable result.

**Theorem C-2.21 (Hopkins–Levitzki).** *If a ring $R$ is left artinian, then it is left noetherian.*

**Proof.** It suffices to prove that $R$, regarded as a left module over itself, has a composition series, for then Proposition C-2.3 applies at once to show that $R$ has the ACC on left ideals (its submodules).

If $J = J(R)$ denotes the Jacobson radical, then $J^m = (0)$ for some $m \geq 1$, by Proposition C-2.10, and so there is a descending chain

$$R = J^0 \supseteq J \supseteq J^2 \supseteq J^3 \supseteq \cdots \supseteq J^m = (0).$$

Since each $J^q$ is an ideal in $R$, it has DCC, as does its quotient $J^q/J^{q+1}$. Now $R/J$ is a semisimple ring, by Theorem C-2.20 (it is left artinian, being a quotient of a left artinian ring, and Jacobson semisimple, by Corollary C-2.11(ii)). The factor module $J^q/J^{q+1}$ is an $(R/J)$-module; hence, by Corollary B-2.30 in Part 1, $J^q/J^{q+1}$ is a semisimple module, and so it can be decomposed into a direct sum of (perhaps infinitely many) simple $(R/J)$-modules. But there can be only finitely many summands, for every $(R/J)$-submodule of $J^q/J^{q+1}$ is necessarily an $R$-submodule, and $J^q/J^{q+1}$ has DCC on $R$-submodules. Hence, there are simple $(R/J)$-modules $S_i$ with

$$J^q/J^{q+1} = S_1 \oplus S_2 \oplus \cdots \oplus S_p.$$

Throwing away one simple summand at a time yields a series of $J^q/J^{q+1}$ whose $i$th factor module is

$$(S_i \oplus S_{i+1} \oplus \cdots \oplus S_p)/(S_{i+1} \oplus \cdots \oplus S_p) \cong S_i.$$

Now the simple $(R/J)$-module $S_i$ is also a simple $R$-module, by Corollary C-2.8, for it is an $R$-module annihilated by $J$, so that we have constructed a composition series for $J^q/J^{q+1}$ as a left $R$-module. Finally, refine the original series for $R$ in this way, for every $q$, to obtain a composition series for $R$. •

The converse of Theorem C-2.21 is false: $\mathbb{Z}$ is noetherian but not artinian.

The next result is fundamental.

**Theorem C-2.22 (Maschke's Theorem).** *If $G$ is a finite group and $k$ is a field whose characteristic does not divide $|G|$, then $kG$ is a left semisimple ring.*

**Remark.** The hypothesis holds if $k$ has characteristic 0. ◀

**Proof.** By Proposition B-2.29 in Part 1, it suffices to prove that every left ideal $I$ of $kG$ is a direct summand. Since $k$ is a field, $kG$ is a vector space over $k$ and $I$ is a subspace. By Corollary B-2.9 in Part 1, $I$ is a (vector space) direct summand: there is a subspace $V$ (which may not be a left ideal in $kG$) with $kG = I \oplus V$. There is a $k$-linear transformation

$$d\colon kG \to I$$

with $d(b) = b$ for all $b \in I$ and with $\ker d = V$ (each $u \in kG$ has a unique expression of the form $u = b + v$, where $b \in I$ and $v \in V$, and $d(u) = b$). Were $d$ a $kG$-map, not merely a $k$-map, then we would be done, by the criterion of Corollary B-2.15 in Part 1 ($I$ is a summand of $kG$ if and only if it is a retract: there is a $kG$-map $D\colon kG \to I$ with $D(u) = u$ for all $u \in I$). We now force $d$ to be a $kG$-map by an "averaging process".

Define $D : kG \to kG$ by

$$D(u) = \frac{1}{|G|} \sum_{x \in G} x d(x^{-1}u)$$

for all $u \in kG$. Note that $|G| \neq 0$ in $k$, by the hypothesis on the characteristic of $k$, and so $1/|G|$ is defined. It is obvious that $D$ is a $k$-map.

(i) $\operatorname{im} D \subseteq I$.
If $u \in kG$ and $x \in G$, then $d(x^{-1}u) \in I$ (because $\operatorname{im} d \subseteq I$), and $xd(x^{-1}u) \in I$ because $I$ is a left ideal. Therefore, $D(u) \in I$, for each term in the sum defining $D(u)$ lies in $I$.

(ii) If $b \in I$, then $D(b) = b$.
Since $b \in I$, so is $x^{-1}b$, and so $d(x^{-1}b) = x^{-1}b$. Hence, $xd(x^{-1}b) = xx^{-1}b = b$. Therefore, $\sum_{x \in G} xd(x^{-1}b) = |G|b$, and so $D(b) = b$.

(iii) $D$ is a $kG$-map.
It suffices to prove that $D(gu) = gD(u)$ for all $g \in G$ and all $u \in kG$:

$$gD(u) = \frac{1}{|G|} \sum_{x \in G} gxd(x^{-1}u) = \frac{1}{|G|} \sum_{x \in G} gxd(x^{-1}g^{-1}gu)$$
$$= \frac{1}{|G|} \sum_{y = gx \in G} yd(y^{-1}gu) = D(gu)$$

(as $x$ ranges over all of $G$, so does $y = gx$). •

The converse of Maschke's Theorem is true: if $G$ is a finite group and $k$ is a field whose characteristic $p$ divides $|G|$, then $kG$ is not left semisimple; a proof is outlined in Exercise C-2.12 on page 145.

Before analyzing left semisimple rings further, let us give several characterizations of them.

**Proposition C-2.23.** *The following conditions on a ring $R$ are equivalent.*

(i) *$R$ is left semisimple.*

(ii) *Every left $R$-module is a semisimple module.*

(iii) *Every left $R$-module is injective.*

(iv) *Every short exact sequence of left $R$-modules splits.*

(v) *Every left $R$-module is projective.*

**Proof.**

(i) $\Rightarrow$ (ii). This follows at once from Corollary B-2.30 in Part 1, which says that if $R$ is a semisimple ring, then every left $R$-module is a semisimple module.

(ii) $\Rightarrow$ (iii). Let $E$ be a left $R$-module; Proposition B-4.52 in Part 1 says that $E$ is injective if every exact sequence $0 \to E \to B \to C \to 0$ splits. By hypothesis, $B$ is a semisimple module, and so Proposition B-2.29 in Part 1 implies that the sequence splits; thus, $E$ is injective.

(iii) $\Rightarrow$ (iv). If $0 \to A \to B \to C \to 0$ is an exact sequence, then it must split because, as every module, $A$ is injective, by Proposition B-4.52 in Part 1.

(iv) $\Rightarrow$ (v). Given a module $M$, there is an exact sequence

$$0 \to F' \to F \to M \to 0,$$

where $F$ is free. This sequence splits, by hypothesis, and so $F \cong M \oplus F'$. Therefore, $M$ is a direct summand of a free module, and hence it is projective, by Theorem B-4.44 in Part 1.

(v) $\Rightarrow$ (i). If $I$ is a left ideal of $R$, then

$$0 \to I \to R \to R/I \to 0$$

is an exact sequence. By hypothesis, $R/I$ is projective, and so this sequence splits, by Proposition B-4.41 in Part 1; thus, $I$ is a direct summand of $R$. By Proposition B-2.29 in Part 1, $R$ is a semisimple left $R$-module; that is, $R$ is a left semisimple ring. •

Modules over semisimple rings are so nice that there is a notion of *global dimension* of a ring $R$ that measures how far removed $R$ is from being semisimple. We will discuss global dimension in the chapter on homological algebra.

In order to give more examples of left semisimple rings, we look at endomorphism rings of direct sums. Consider $\mathrm{Hom}_R(A, B)$, where both $A$ and $B$ are left $R$-modules that are finite direct sums: say, $A = \bigoplus_{i=1}^{n} A_i$ and $B = \bigoplus_{j=1}^{m} B_j$. Since a direct product of a finite family of modules is their direct sum, Theorem B-4.8 in Part 1 gives

$$\mathrm{Hom}_R(A, B) \cong \bigoplus_{i,j} \mathrm{Hom}_R(A_i, B_j).$$

More precisely, if $\alpha_i\colon A_i \to A$ is the $i$th injection and $p_j\colon B \to B_j$ is the $j$th projection, then each $f \in \mathrm{Hom}_R(A,B)$ gives maps

$$f_{ji} = p_j f \alpha_i \in \mathrm{Hom}_R(A_i, B_j).$$

Thus, $f$ defines a ***generalized*** $m\times n$ ***matrix*** $[f_{ji}]$ (we call $[f_{ji}]$ a *generalized* matrix because entries in different positions need not lie in the same algebraic system). The map $f \mapsto [f_{ji}]$ is an isomorphism $\mathrm{Hom}_R(A,B) \to \bigoplus_{ij} \mathrm{Hom}_R(A_i,B_j)$. Similarly, if $g\colon B \to C$, where $C = \bigoplus_{k=1}^{\ell} C_k$, then $g$ defines a generalized $\ell \times m$ matrix $[g_{kj}]$, where $g_{kj} = q_k g \beta_j\colon B_j \to C_k$, $\beta_j\colon B_j \to B$ are the injections, and $q_k\colon C \to C_k$ are the projections.

The composite $gf\colon A \to C$ defines a generalized $\ell \times n$ matrix, and we claim that it is given by matrix multiplication: $(gf)_{ki} = \sum_j g_{kj} f_{ji}$:

$$\sum_j g_{kj} f_{ji} = \sum_j q_k g \beta_j p_j f \alpha_i = q_k g \Big( \sum_j \beta_j p_j \Big) f \alpha_i = q_k g f \alpha_i = (gf)_{ki},$$

because $\sum_j \beta_j p_j = 1_B$.

By adding some hypotheses, we can pass from generalized matrices to honest matrices.

**Proposition C-2.24.** *Let $V = \bigoplus_{i=1}^{n} V_i$ be a left $R$-module. If there is a left $R$-module $L$ and, for each $i$, an isomorphism $\varphi_i\colon V_i \to L$, then there is a ring isomorphism*

$$\mathrm{End}_R(V) \cong \mathrm{Mat}_n(\mathrm{End}_R(L)).$$

**Proof.** Define $\theta\colon \mathrm{End}_R(V) \to \mathrm{Mat}_n(\mathrm{End}_R(L))$ by

$$\theta\colon f \mapsto [\varphi_j p_j f \alpha_i \varphi_i^{-1}],$$

where $\alpha_i\colon V_i \to V$ and $p_j\colon V \to V_j$ are injections and projections, respectively. That $\theta$ is an additive isomorphism is just the identity

$$\mathrm{Hom}\Big(\bigoplus_i V_i, \bigoplus_i V_i\Big) \cong \bigoplus_{i,j} \mathrm{Hom}(V_i, V_j),$$

which holds when the index sets are finite. In the paragraph above defining generalized matrices, the home of the $ij$ entry is $\mathrm{Hom}_R(V_i,V_j)$, whereas the present home of this entry is the isomorphic replica $\mathrm{Hom}_R(L,L) = \mathrm{End}_R(L)$.

We now show that $\theta$ preserves multiplication. If $g, f \in \mathrm{End}_R(V)$, then $\theta(gf) = [\varphi_j p_j g f \alpha_i \varphi_i^{-1}]$, while the matrix product is

$$\begin{aligned}
\theta(g)\theta(f) &= \Big[\sum_k (\varphi_j p_j g \alpha_k \varphi_k^{-1})(\varphi_k p_k f \alpha_i \varphi_i^{-1})\Big] \\
&= \Big[\sum_k \varphi_j p_j g \alpha_k p_k f \alpha_i \varphi_i^{-1}\Big] \\
&= \Big[\varphi_j p_j g \Big(\sum_k \alpha_k p_k\Big) f \alpha_i \varphi_i^{-1}\Big] \\
&= \Big[\varphi_j p_j g f \alpha_i \varphi_i^{-1}\Big]. \quad \bullet
\end{aligned}$$

**Corollary C-2.25.** *If $V$ is an $n$-dimensional left vector space over a division ring $k$, then there is an isomorphism of rings*

$$\mathrm{End}_k(V) \cong \mathrm{Mat}_n(k)^{\mathrm{op}}.$$

**Proof.** The isomorphism $\mathrm{End}_k(V) \cong \mathrm{Mat}_n(k^{\mathrm{op}})$ is the special case of Proposition C-2.24 for $V = V_1 \oplus \cdots \oplus V_n$, where each $V_i$ is one-dimensional, hence is isomorphic to $k$. Note that $\mathrm{End}_k(k) \cong k^{\mathrm{op}}$, by Proposition B-1.24 in Part 1. Now apply Proposition B-1.25 in Part 1, which says that $\mathrm{Mat}_n(k^{\mathrm{op}}) \cong \mathrm{Mat}_n(k)^{\mathrm{op}}$. •

The next result involves a direct sum decomposition at the opposite extreme of that in Proposition C-2.24.

**Corollary C-2.26.** *Let an $R$-module $M$ be a direct sum $M = B_1 \oplus \cdots \oplus B_r$ in which $\mathrm{Hom}_R(B_i, B_j) = \{0\}$ for all $i \neq j$. Then there is a ring isomorphism*

$$\mathrm{End}_R(M) \cong \mathrm{End}_R(B_1) \times \cdots \times \mathrm{End}_R(B_r).$$

**Proof.** If $f, g \in \mathrm{End}_R(M)$, let $[f_{ij}]$ and $[g_{ij}]$ be their generalized matrices. It suffices to show that $[g_{ij}][f_{ij}]$ is the diagonal matrix

$$\mathrm{diag}(g_{11}f_{11}, \ldots, g_{rr}f_{rr}).$$

But if $i \neq j$, then $g_{ik}f_{kj} \in \mathrm{Hom}_R(B_i, B_j) = 0$; hence, $(gf)_{ij} = \sum_k g_{ik}f_{kj} = 0$. •

We can now give more examples of semisimple rings. The Wedderburn–Artin Theorems (proved in the next section) will say that there are no others.

**Proposition C-2.27.**

(i) *If $k$ is a division ring and $V$ is a left vector space over $k$ with $\dim(V) = n$, then $\mathrm{End}_k(V) \cong \mathrm{Mat}_n(k^{\mathrm{op}})$ is a left semisimple ring.*

(ii) *If $k_1, \ldots, k_r$ are division rings, then*

$$\mathrm{Mat}_{n_1}(k_1) \times \cdots \times \mathrm{Mat}_{n_r}(k_r)$$

*is a left semisimple ring.*

**Proof.**

(i) By Proposition C-2.24, we have

$$\mathrm{End}_k(V) \cong \mathrm{Mat}_n(\mathrm{End}_k(k));$$

by Proposition B-1.24 in Part 1, $\mathrm{End}_k(k) \cong k^{\mathrm{op}}$. Therefore, $\mathrm{End}_k(V) \cong \mathrm{Mat}_n(k^{\mathrm{op}})$.

Let us now show that $\mathrm{End}_k(V)$ is left semisimple. If $v_1, \ldots, v_n$ is a basis of $V$, define

$$\mathrm{COL}(j) = \{T \in \mathrm{End}_k(V) : T(v_i) = 0 \text{ for all } i \neq j\}.$$

It is easy to see that $\mathrm{COL}(j)$ is a left ideal in $\mathrm{End}_k(V)$: if $S \in \mathrm{End}_k(V)$, then $S(Tv_i) = 0$ for all $i \neq j$. Recall Proposition C-2.5: if we look in $\mathrm{Mat}_n(k^{\mathrm{op}}) \cong \mathrm{End}_k(V)$, then $\mathrm{COL}(j)$ corresponds to all those linear transformations killing those $v_i$ with $i \neq j$. It is obvious that

$$\mathrm{Mat}_n(k^{\mathrm{op}}) = \mathrm{COL}(1) \oplus \cdots \oplus \mathrm{COL}(n).$$

Hence, $\mathrm{End}_k(V)$ is also such a direct sum. We saw, in Example C-2.6(iii), that each $\mathrm{COL}(\ell)$ is a minimal left ideal, and so $\mathrm{End}_k(V)$ is a left semisimple ring.

(ii) This follows at once from (i) and Corollary C-2.18, for if $k$ is a division ring, then so is $k^{\mathrm{op}}$, by Exercise B-1.38 on page 300 in Part 1. •

**Corollary C-2.28.** *If $V$ is an $n$-dimensional left vector space over a division ring $k$, then the minimal left ideals $\mathrm{COL}(\ell)$ in $\mathrm{End}_k(V)$, for $1 \le \ell \le n$, are all isomorphic.*

**Proof.** Let $v_1, \ldots, v_n$ be a basis of $V$. For each $\ell$, define $p_\ell \colon V \to V$ to be the linear transformation that interchanges $v_\ell$ and $v_1$ and that fixes all the other $v_j$. It is easy to see that $T \mapsto Tp_\ell$ is an isomorphism $\mathrm{COL}(1) \to \mathrm{COL}(\ell)$. •

There may be minimal left ideals other than $\mathrm{COL}(\ell)$ for some $\ell$. However, we will see (in Lemma C-2.40(ii)) that all the minimal left ideals in $\mathrm{End}_k(V)$ are isomorphic to one of these.

**Definition.** A ring $R$ is ***simple*** if it is not the zero ring and it has no proper nonzero two-sided ideals.

Our language is a little deceptive. It is true that left artinian simple rings are semisimple (Proposition C-2.38), but there are simple rings that are not semisimple.

**Proposition C-2.29.** *If $k$ is a division ring, then $R = \mathrm{Mat}_n(k)$ is a simple ring.*

**Proof.** A ***matrix unit*** $E_{pq}$ is the $n \times n$ matrix whose $p, q$ entry is 1 and all of whose other entries are 0. The matrix units form a basis for $\mathrm{Mat}_n(k)$ viewed as a left vector space over $k$, for each matrix $A = [a_{ij}]$ has a unique expression

$$A = \sum_{i,j} a_{ij} E_{ij}.$$

(Of course, this says that $\dim_k(\mathrm{Mat}_n(k)) = n^2$.) A routine calculation shows that matrix units multiply according to the following rule:

$$E_{ij}E_{k\ell} = \begin{cases} 0 & \text{if } j \neq k, \\ E_{i\ell} & \text{if } j = k. \end{cases}$$

Suppose that $N$ is a nonzero two-sided ideal in $\mathrm{Mat}_n(k)$. If $A$ is a nonzero matrix in $N$, it has a nonzero entry; say, $a_{ij} \neq 0$. Since $N$ is a two-sided ideal, $N$ contains $E_{pi}AE_{jq}$ for all $p, q$. But

$$E_{pi}AE_{jq} = E_{pi} \sum_{k,\ell} a_{k\ell} E_{k\ell} E_{jq} = E_{pi} \sum_k a_{kj} E_{kq} = \sum_k a_{kj} E_{pi} E_{kq} = a_{ij} E_{pq}.$$

Since $a_{ij} \neq 0$ and $k$ is a division ring, $a_{ij}^{-1} \in k$, and so $E_{pq} \in N$ for all $p, q$. But the collection of all $E_{pq}$ span the left vector space $\mathrm{Mat}_n(k)$ over $k$, and so $N = \mathrm{Mat}_n(k)$. •

## Exercises

**C-2.10.** Prove that a finitely generated left semisimple $R$-module $M$ over a ring $R$ is a direct sum of a finite number of simple left modules.

**C-2.11.** Let $A$ be an $n$-dimensional $k$-algebra over a field $k$. Prove that $A$ can be imbedded as a $k$-subalgebra of $\mathrm{Mat}_n(k)$.

**Hint.** If $a \in A$, define $L_a\colon A \to A$ by $L_a\colon x \mapsto ax$.

∗ **C-2.12.** Let $G$ be a finite group, and let $k$ be a commutative ring. Define $\varepsilon\colon kG \to k$ by

$$\varepsilon\Big(\sum_{g\in G} a_g g\Big) = \sum_{g\in G} a_g$$

(this map is called the ***augmentation***, and its kernel, denoted by $\mathcal{G}$, is called the ***augmentation ideal***).

(i) Prove that $\varepsilon$ is a $kG$-map and that $kG/\mathcal{G} \cong k$ as $k$-algebras. Conclude that $\mathcal{G}$ is a two-sided ideal in $kG$.

(ii) Prove that $kG/\mathcal{G} \cong V_0(k)$, where $V_0(k)$ is $k$ viewed as a trivial $kG$-module.
**Hint.** $\mathcal{G}$ is a two-sided ideal containing $xu - u = (x-1)u$.

(iii) Use (ii) to prove that if $kG = \mathcal{G} \oplus V$, then $V = \langle v\rangle$, where $v = \sum_{g\in G} g$.
**Hint.** Argue as in Example C-2.34.

(iv) Assume that $k$ is a field whose characteristic $p$ does divide $|G|$. Prove that $kG$ is not left semisimple.
**Hint.** First show that $\varepsilon(v) = 0$, and then show that the short exact sequence

$$0 \to \mathcal{G} \to kG \xrightarrow{\varepsilon} k \to 0$$

does not split. (If $G$ is a finite $p$-group and $k$ is a field of characteristic $p$, then the Jacobson radical $J(kG)$ is the augmentation ideal; see Lam [**133**], p. 131).

∗ **C-2.13.** Let $M$ be a left $R$-module over a semisimple ring $R$. Prove that $M$ is indecomposable if and only if $M$ is simple.

**C-2.14.** If $k$ is a division ring, prove that every two minimal left ideals in $\mathrm{Mat}_n(k)$ are isomorphic. (Compare Corollary C-2.28.)

∗ **C-2.15.** Let $k$ be a division ring and let $D \subseteq k$ be a subdivision ring.

(i) Prove that $k$ is a left vector space over $D$, and conclude that $\dim_D(k)$ is defined (of course, $\dim_D(k) = \infty$ is possible).

(ii) If $Z \subseteq D \subseteq k$ is a tower of division rings with $\dim_D(k)$ and $\dim_Z(D)$ finite, prove that $\dim_Z(k)$ is finite and

$$\dim_Z(k) = \dim_D(k)\dim_Z(D).$$

**Hint.** Adapt the proof of Proposition B-1.40 in Part 1.

**C-2.16.** Let $T\colon V \to V$ be a linear transformation, where $V$ is a vector space over a field $k$, and let $k[T]$ be defined by

$$k[T] = k[x]/(m(x)),$$

where $m(x)$ is the minimum polynomial of $T$.

(i) If $m(x) = \prod_p p(x)^{e_p}$, where the $p(x) \in k[x]$ are distinct irreducible polynomials and $e_p \geq 1$, prove that $k[T] \cong \prod_p k[x]/(p(x)^{e_p})$.

(ii) Prove that $k[T]$ is a semisimple ring if and only if $m(x)$ is a product of distinct linear factors. (In linear algebra, we show that this last condition is equivalent to $T$ being *diagonalizable*; that is, any matrix of $T$ (arising from some choice of basis of $T$) is similar to a diagonal matrix.)

**C-2.17.** If $\mathbb{H}$ is the division ring of real quaternions, prove that its multiplicative group $\mathbb{H}^\times$ has a finite subgroup that is not cyclic. Compare with Theorem A-3.59 in Part 1.

## C-2.5. Wedderburn–Artin Theorems

We are now going to prove the converse of Proposition C-2.27(ii): every left semisimple ring is isomorphic to a direct product of matrix rings over division rings. The first step shows how division rings arise.

**Theorem C-2.30 (Schur's Lemma).** *Let $M$ and $M'$ be simple left $R$-modules over a ring $R$.*

(i) *Every nonzero $R$-map $f\colon M \to M'$ is an isomorphism.*

(ii) $\mathrm{End}_R(M)$ *is a division ring. In particular, if $L$ is a minimal left ideal in a ring $R$, then* $\mathrm{End}_R(L)$ *is a division ring.*

**Proof.**

(i) Since $M$ is simple, it has only two submodules: $M$ itself and $\{0\}$. Now the submodule $\ker f \neq M$ because $f \neq 0$, so that $\ker f = \{0\}$ and $f$ is an injection. Similarly, the submodule $\operatorname{im} f \neq \{0\}$, so that $\operatorname{im} f = M'$ and $f$ is a surjection.

(ii) If $f\colon M \to M$ and $f \neq 0$, then $f$ is an isomorphism, by part (i), and hence it has an inverse $f^{-1} \in \mathrm{End}_R(M)$. Thus, $\mathrm{End}_R(M)$ is a division ring. •

Here is a surprising result.

**Theorem C-2.31 (Wedderburn).** *Every finite division ring $k$ is a field; that is, multiplication in $k$ is commutative.*

**Proof (Witt).**[4] If $Z$ denotes the center of $k$, then $Z$ is a finite field, and so it has $q$ elements (where $q$ is a power of some prime). It follows that $k$ is a vector space over $Z$, and so $|k| = q^n$ for some $n \geq 1$; that is, if we define

$$[k : Z] = \dim_Z(k),$$

then $[k : Z] = n$. The proof will be complete if we can show that $n > 1$ leads to a contradiction.

If $a \in k$, define $C(a) = \{u \in k : ua = au\}$. It is routine to check that $C(a)$ is a subdivision ring of $k$ that contains $Z$: if $u, v \in k$ commute with $a$, then so do $u + v, uv$, and $u^{-1}$ (when $u \neq 0$). Consequently, $|C(a)| = q^{d(a)}$ for some integer

[4] We shall give another proof of this in Theorem C-2.124.

$d(a)$; that is, $[C(a) : Z] = d(a)$. We do not know whether $C(a)$ is commutative, but Exercise C-2.15 on page 145 gives

$$[k : Z] = [k : C(a)][C(a) : Z],$$

where $[k : C(a)]$ denotes the dimension of $k$ as a left vector space over $C(a)$. That is, $n = [k : C(a)]d(a)$, and so $d(a)$ is a divisor of $n$.

Since $k$ is a division ring, its nonzero elements $k^\times$ form a multiplicative group of order $q^n - 1$. By Exercise C-2.2 on page 134, the center of the group $k^\times$ is $Z^\times$ and, if $a \in k^\times$, then its centralizer $C_{k^\times}(a) = C(a)^\times$. Hence, $|Z(k^\times)| = q - 1$ and $|C_{k^\times}(a)| = q^{d(a)} - 1$, where $d(a) \mid n$.

The class equation for $k^\times$ is

$$|k^\times| = |Z^\times| + \sum_i [k^\times : C_{k^\times}(a_i)],$$

where one $a_i$ is chosen from each noncentral conjugacy class. But

$$[k^\times : C_{k^\times}(a_i)] = |k^\times|/|C_{k^\times}(a_i)| = (q^n - 1)/(q^{d(a_i)} - 1),$$

so that the class equation becomes

$$q^n - 1 = q - 1 + \sum_i \frac{q^n - 1}{q^{d(a_i)} - 1}. \tag{1}$$

We have already noted that each $d(a_i)$ is a divisor of $n$, while the condition that $a_i$ is not central says that $d(a_i) < n$.

Recall that the $n$th cyclotomic polynomial is $\Phi_n(x) = \prod(x - \zeta)$, where $\zeta$ ranges over all the primitive $n$th roots of unity. In Corollary A-3.108 in Part 1, we proved that $\Phi_n(q)$ is a common divisor of $q^n - 1$ and $(q^n - 1)/(q^{d(a_i)} - 1)$ for all $i$, and so Eq. (1) gives

$$\Phi_n(q) \mid (q - 1).$$

If $n > 1$ and $\zeta$ is a primitive $n$th root of unity, then $\zeta \neq 1$, and hence $\zeta$ is a point on the unit circle to the left of the vertical line through $(1, 0)$. Since $q$ is a prime power, it is a point on the $x$-axis with $q \geq 2$, and so the distance $|q - \zeta| > q - 1$. Therefore,

$$|\Phi_n(q)| = \prod |q - \zeta| > q - 1,$$

and this contradicts $\Phi_n(q) \mid (q - 1)$. We conclude that $n = 1$; that is, $k = Z$, and so $k$ is commutative. •

The second step for the Wedderburn–Artin Theorems investigates minimal left ideals.

**Lemma C-2.32.** *If $L$ and $L'$ are minimal left ideals in a ring $R$, then each of the following statements implies the one below it:*

(i) $LL' \neq (0)$.

(ii) $\mathrm{Hom}_R(L, L') \neq \{0\}$ *and there exists* $b' \in L'$ *with* $L' = Lb'$.

(iii) $L \cong L'$ *as left $R$-modules.*

*If also* $L^2 \neq (0)$, *then* (iii) *implies* (i) *and the three statements are equivalent.*

**Proof.**

(i) $\Rightarrow$ (ii). If $LL' \neq (0)$, then there exists $b \in L$ and $b' \in L'$ with $bb' \neq 0$. Thus, the function $f\colon L \to L'$, defined by $x \mapsto xb'$, is a nonzero $R$-map, and so $\mathrm{Hom}_R(L, L') \neq \{0\}$. Moreover, $Lb' = L'$, for it is a nonzero submodule of the minimal left ideal $L'$.

(ii) $\Rightarrow$ (iii). If $\mathrm{Hom}_R(L, L') \neq \{0\}$, then there is a nonzero $f\colon L \to L'$, and $f$ is an isomorphism, by Schur's Lemma; that is, $L \cong L'$.

(iii) and $L^2 \neq (0) \Rightarrow$ (i). Assume now that $L^2 \neq (0)$, so there are $x, y \in L$ with $xy \neq 0$. If $g\colon L \to L'$ is an isomorphism, then $0 \neq g(xy) = xg(y) \in LL'$, and so $LL' \neq (0)$. •

Note that if $J(R) = (0)$, then $L^2 \neq (0)$; otherwise, $L$ is a nilpotent left ideal and Corollary C-2.9 gives $L \subseteq J(R) = (0)$, a contradiction.

**Theorem C-2.33.** *If $R = L_1 \oplus \cdots \oplus L_n$ is a left semisimple ring, where the $L_p$ are minimal left ideals, then every simple $R$-module $S$ is isomorphic to some $L_p$.*

**Proof.** Now $S \cong \mathrm{Hom}_R(R, S) \neq \{0\}$, by Corollary B-4.27 in Part 1. But, if $\mathrm{Hom}_R(L_p, S) = \{0\}$ for all $p$, then $\mathrm{Hom}_R(R, S) = \{0\}$ (for $R = L_1 \oplus \cdots \oplus L_n$). Thus, $\mathrm{Hom}_R(L_p, S) \neq \{0\}$ for some $p$. Since both $L_p$ and $S$ are simple, Theorem C-2.30(i) gives $L_p \cong S$. •

Here is a fancier proof of Theorem C-2.33.

**Theorem** C-2.33 **Again**. *If $R = L_1 \oplus \cdots \oplus L_n$ is a left semisimple ring, where the $L_p$ are minimal left ideals, then every simple $R$-module $S$ is isomorphic to some $L_p$.*

**Proof.** By Corollary B-1.33 in Part 1, there is a left ideal $I$ with $S \cong R/I$, and so there is a series

$$R \supseteq I \supseteq (0).$$

In Proposition C-2.17, we saw that

$$R = L_1 \oplus \cdots \oplus L_n \supseteq L_2 \oplus \cdots \oplus L_n \supseteq \cdots \supseteq L_n \supseteq (0)$$

is a composition series with factor modules $L_1, \ldots, L_n$. The Schreier Refinement Theorem (Theorem B-1.37 in Part 1) now says that these two series have equivalent refinements. Since a composition series admits only refinements that repeat a term, the factor module $S$ occurring in the refinement of the first series must be isomorphic to one of the factor modules in the second series; that is, $S \cong L_p$ for some $p$. •

**Example C-2.34.** The trivial $kG$-module $V_0(k)$ (Example C-2.14) is a simple $kG$-module (for it is one-dimensional, hence has no subspaces other than $\{0\}$ and itself). By Theorem C-2.33, $V_0(k)$ is isomorphic to some minimal left ideal $L$ of $kG$. We shall find $L$ by searching for elements $u = \sum_{g \in G} a_g g$ in $kG$ with $hu = u$ for all $h \in G$. For such elements $u$,

$$hu = \sum_{g \in G} a_g hg = \sum_{g \in G} a_g g = u.$$

Since the elements in $G$ form a basis for the vector space $kG$, we may equate coefficients, and so $a_g = a_{hg}$ for all $g \in G$; in particular, $a_1 = a_h$. As this holds for every $h \in G$, all the coefficients $a_g$ are equal. Therefore, if we define $\gamma \in kG$ by

$$\gamma = \sum_{g \in G} g,$$

then $u$ is a scalar multiple of $\gamma$. It follows that $L = \langle \gamma \rangle$ is a left ideal isomorphic to the trivial module $V_0(k)$; moreover, $\langle \gamma \rangle$ is the unique such left ideal. ◀

An abstract left semisimple ring $R$ is a direct sum of finitely many minimal left ideals: $R = \bigoplus_j L_j$, and we now know that $\mathrm{End}_R(L_j)$ is a division ring for every $j$. The next step is to find the direct summands of $R$ that will ultimately turn out to be matrix rings; they arise from a decomposition of $R$ into minimal left ideals by collecting isomorphic terms.

**Definition.** Let $R$ be a left semisimple ring, and let

$$R = L_1 \oplus \cdots \oplus L_n,$$

where the $L_p$ are minimal left ideals. Reindex the summands so that no two of the first $r$ ideals $L_1, \ldots, L_r$ are isomorphic, while every $L_p$ in the given decomposition is isomorphic to some $L_i$ for $1 \le i \le r$. The left ideals

$$B_i = \bigoplus_{L_p \cong L_i} L_p$$

are called the ***simple components*** of $R$ relative to the decomposition $R = \bigoplus_p L_p$.

We shall see, in Corollary C-2.41, that the simple components do not depend on the particular decomposition of $R$ as a direct sum of minimal left ideals.

We divide the usual version of the Wedderburn–Artin Theorem[5] classifying semisimple rings into two parts: an existence theorem and a uniqueness theorem.

**Theorem C-2.35 (Wedderburn–Artin I).** *A ring $R$ is left semisimple if and only if $R$ is isomorphic to a direct product of matrix rings over division rings.*

**Proof.** Sufficiency is Proposition C-2.27(ii).

For necessity, if $R$ is left semisimple, then it is the direct sum of its simple components:

$$R = B_1 \oplus \cdots \oplus B_r,$$

where each $B_i$ is a direct sum of isomorphic minimal left ideals. Proposition B-1.24 in Part 1 says that there is a ring isomorphism

$$R^{\mathrm{op}} \cong \mathrm{End}_R(R),$$

where $R$ is regarded as a left module over itself. Now $\mathrm{Hom}_R(B_i, B_j) = \{0\}$ for all $i \neq j$, by Lemma C-2.32, so that Corollary C-2.26 applies to give a ring isomorphism

$$R^{\mathrm{op}} \cong \mathrm{End}_R(R) \cong \mathrm{End}_R(B_1) \times \cdots \times \mathrm{End}_R(B_r).$$

[5] Wedderburn proved Theorems C-2.35 and C-2.43 for semisimple $k$-algebras, where $k$ is a field. E. Artin generalized these theorems to semisimple rings with left DCC. These theorems are why *artinian* rings are so called.

By Proposition C-2.24, there are isomorphisms of rings

$$\mathrm{End}_R(B_i) \cong \mathrm{Mat}_{n_i}(\mathrm{End}_R(L_i)),$$

because $B_i$ is a direct sum of isomorphic copies of $L_i$. By Schur's Lemma, $\mathrm{End}_R(L_i)$ is a division ring, say, $k_i$, and so

$$R^{\mathrm{op}} \cong \mathrm{Mat}_{n_1}(k_1) \times \cdots \times \mathrm{Mat}_{n_r}(k_r).$$

Hence,

$$R \cong \mathrm{Mat}_{n_1}(k_1)^{\mathrm{op}} \times \cdots \times \mathrm{Mat}_{n_r}(k_r)^{\mathrm{op}}.$$

Finally, Proposition B-1.25 in Part 1 gives

$$R \cong \mathrm{Mat}_{n_1}(k_1^{\mathrm{op}}) \times \cdots \times \mathrm{Mat}_{n_r}(k_r^{\mathrm{op}}).$$

This completes the proof, for $k_i^{\mathrm{op}}$ is also a division ring for all $i$, by Exercise B-1.38 on page 300 in Part 1. •

**Corollary C-2.36.** *A ring $R$ is left semisimple if and only if it is right semisimple.*

**Proof.** It is easy to see that a ring $R$ is right semisimple if and only if its opposite ring $R^{\mathrm{op}}$ is left semisimple. But we saw, in the middle of the proof of Theorem C-2.35, that

$$R^{\mathrm{op}} \cong \mathrm{Mat}_{n_1}(k_1) \times \cdots \times \mathrm{Mat}_{n_r}(k_r),$$

where $k_i = \mathrm{End}_R(L_i)$. •

As a consequence of this corollary, we say that a ring is ***semisimple*** without the adjectives left or right.

**Corollary C-2.37.** *A commutative ring $R$ is semisimple if and only if it is isomorphic to a direct product of finitely many fields.*

**Proof.** A field is a semisimple ring, and so a direct product of finitely many fields is also semisimple, by Corollary C-2.18. Conversely, if $R$ is semisimple, it is a direct product of matrix rings over division rings. Since $R$ is commutative, all the matrix rings must be of size $1 \times 1$ and all the division rings must be fields. •

Even though the name suggests it, it is not clear that simple rings are semisimple. Indeed, we must assume a chain condition: if $V$ is an infinite-dimensional vector space over a field $k$, then $R = \mathrm{End}_k(V)$ is a simple ring which is not semisimple (Lam [**133**], pp. 43–44).

**Proposition C-2.38.** *A simple left artinian ring $R$ is semisimple.*

**Proof (Janusz).** Since $R$ is left artinian, it contains a minimal left ideal, say, $L$; of course, $L$ is a simple left $R$-module. For each $a \in R$, the function $f_a\colon L \to R$, defined by $f_a(x) = xa$, is a map of left $R$-modules: if $r \in R$, then

$$f_a(rx) = (rx)a = r(xa) = r f_a(x).$$

Now $\mathrm{im}\, f_a = La$, while $L$ being a simple module forces $\ker f_a = L$ or $\ker f_a = (0)$. In the first case, we have $La = (0)$; in the second case, we have $L \cong La$. Thus, $La$ is either $(0)$ or a minimal left ideal.

Consider the sum $I = \left\langle \bigcup_{a\in R} La \right\rangle \subseteq R$. Plainly, $I$ is a left ideal; it is a right ideal as well, for if $b \in R$ and $La \subseteq I$, then $(La)b = L(ab) \subseteq I$. Since $R$ is a simple ring, the nonzero two-sided ideal $I$ must equal $R$. We claim that $R$ is a sum of only finitely many $La$'s. As any element of $R$, the unit 1 lies in some finite sum of $La$'s, say, $1 \in Le_1 + \cdots + Le_n$. If $b \in R$, then $b = b1 \in b(Le_1 + \cdots + Le_n) \subseteq Le_1 + \cdots + Le_n$ (because $Le_1 + \cdots + Le_n$ is a left ideal). Hence, $R = Le_1 + \cdots + Le_n$.

To prove that $R$ is semisimple, it remains to show that it is a *direct sum* of simple submodules. Choose $n$ minimal such that $R = Le_1 + \cdots + Le_n$; we claim that $R = Le_1 \oplus \cdots \oplus Le_n$. By Proposition C-2.17, it suffices to show, for all $i$, that

$$Le_i \ \cap \ \Big(\bigoplus_{j\neq i} Le_j\Big) = (0).$$

If this intersection is not (0), then simplicity of $Le_i$ says that $Le_i \cap (\bigoplus_{j\neq i} Le_j) = Le_i$; that is, $Le_i \subseteq \bigoplus_{j\neq i} Le_j$, and this contradicts the minimal choice of $n$. Therefore, $R$ is a semisimple ring. •

The following corollary follows at once from Proposition C-2.38 and Theorem C-2.35, the Wedderburn–Artin Theorem I.

**Corollary C-2.39.** *If $A$ is a simple left artinian ring, then $A \cong \mathrm{Mat}_n(k)$ for some $n \geq 1$ and some division ring $k$.*

The next lemma gives some interesting properties enjoyed by semisimple rings; it will be used to complete the Wedderburn–Artin Theorems by proving uniqueness of the constituent parts. In particular, it will say that the integer $n$ and the division ring $k$ in Corollary C-2.39 are uniquely determined by $R$.

**Lemma C-2.40.** *Let $R$ be a semisimple ring, and let*

$$R = L_1 \oplus \cdots \oplus L_n = B_1 \oplus \cdots \oplus B_r, \tag{2}$$

*where the $L_p$ are minimal left ideals indexed so that no two of the first $r$ ideals $L_1, \ldots, L_r$ are isomorphic, while every $L_p$ is isomorphic to one of these.*

*Let $B_i$ be the corresponding simple component of $R$; that is, $B_i$ is the direct sum of all the $L_p$ in* Eq. (2) *with $L_p \cong L_i$.*

(i) *Each $B_i$ is a ring that is also a two-sided ideal in $R$, and $B_iB_j = (0)$ if $j \neq i$.*

(ii) *If $L$ is any minimal left ideal in $R$, not necessarily some $L_p$ in* Eq. (2)*, then $L \cong L_i$ for some $i$ with $1 \leq i \leq r$, and $L \subseteq B_i$.*

(iii) *Every two-sided ideal $D$ in $R$ is a direct sum of simple components.*

(iv) *Each $B_i$ is a simple ring.*

**Proof.**

(i) Each $B_i$ is a left ideal. To see that it is also a right ideal, consider

$$B_iR = B_i(B_1 \oplus \cdots \oplus B_r) \subseteq B_iB_1 + \cdots + B_iB_r.$$

Recall, for each $i$ with $1 \leq i \leq r$, that $B_i$ is a direct sum of left ideals $L_p$ isomorphic to $L_i$. If $L \cong L_i$ and $L' \cong L_j$, then the contrapositive, "not (iii)"

$\Rightarrow$ "not (i)" in Lemma C-2.32, applies to give $LL' = (0)$ if $j \neq i$. Hence, if $j \neq i$,

$$B_iB_j = \Big(\bigoplus_{L\cong L_i} L\Big)\Big(\bigoplus_{L'\cong L_j} L'\Big) \subseteq \bigoplus LL' = (0).$$

Thus, $B_iB_1 + \cdots + B_iB_r \subseteq B_iB_i$. Since $B_i$ is a left ideal, $B_iB_i \subseteq RB_i \subseteq B_i$. Therefore, $B_iR \subseteq B_i$, so that $B_i$ is a right ideal and, hence, is a two-sided ideal.

In the last step, proving that $B_i$ is a right ideal, we saw that $B_iB_i \subseteq B_i$; that is, $B_i$ is closed under multiplication. Therefore, to prove that $B_i$ is a ring, it now suffices to prove that it contains a unit element. If 1 is the unit element in $R$, then $1 = e_1 + \cdots + e_r$, where $e_i \in B_i$ for all $i$. If $b_i \in B_i$, then

$$b_i = 1b_i = (e_1 + \cdots + e_r)b_i = e_ib_i,$$

for $B_jB_i = (0)$ whenever $j \neq i$. Similarly, the equation $b_i = b_i1$ gives $b_ie_i = b_i$, and so $e_i$ is a unit in $B_i$. Thus, $B_i$ is a ring.[6]

(ii) By Theorem C-2.33, a minimal left ideal $L$ is isomorphic to $L_i$ for some $i$. Now

$$L = RL = (B_1 \oplus \cdots \oplus B_r)L \subseteq B_1L + \cdots + B_rL.$$

If $j \neq i$, then $B_jL = (0)$, by Lemma C-2.32, so that

$$L \subseteq B_iL \subseteq B_i,$$

because $B_i$ is a right ideal.

(iii) A nonzero two-sided ideal $D$ in $R$ is a left ideal, and so it contains some minimal left ideal $L$, by Proposition C-2.4. Now $L \cong L_i$ for some $i$, by Theorem C-2.33; we claim that $B_i \subseteq D$. By Lemma C-2.32, if $L'$ is any minimal left ideal in $B_i$, then $L' = Lb'$ for some $b' \in L'$. Since $L \subseteq D$ and $D$ is a right ideal, we have $L' = Lb' \subseteq LL' \subseteq DR \subseteq D$. We have shown that $D$ contains every left ideal isomorphic to $L_i$; as $B_i$ is generated by such ideals, $B_i \subseteq D$. Write $R = B_I \oplus B_J$, where $B_I = \bigoplus_i B_i$ with $B_i \subseteq D$ and $B_J = \bigoplus_j B_j$ with $B_j \not\subseteq D$. By Corollary B-2.16 in Part 1 (which holds for modules over noncommutative rings), $D = B_I \oplus (D \cap B_J)$. But $D \cap B_J = (0)$; otherwise, it would contain a minimal left ideal $L \cong L_j$ for some $j \in J$ and, as above, this would force $B_j \subseteq D$. Therefore, $D = B_I$.

(iv) A left ideal in $B_i$ is also a left ideal in $R$: if $a \in R$, then $a = \sum_j a_j$, where $a_j \in B_j$; if $b_i \in B_i$, then

$$ab_i = (a_1 + \cdots + a_r)b_i = a_ib_i \in B_i,$$

because $B_jB_i = (0)$ for $j \neq i$. Similarly, a right ideal in $B_i$ is a right ideal in $R$, and so a two-sided ideal $D$ in $B_i$ is a two-sided ideal in $R$. By (iii), the only two-sided ideals in $R$ are direct sums of simple components, and so $D \subseteq B_i$ implies $D = (0)$ or $D = B_i$. Therefore, $B_i$ is a simple ring. •

[6] $B_i$ is not a subring of $R$ because its unit $e_i$ is not the unit 1 in $R$.

**Corollary C-2.41.** *If $R$ is a semisimple ring, then the simple component $B_i$ containing a minimal left ideal $L_i$ is the left ideal generated by all the minimal left ideals that are isomorphic to $L_i$, where $1 \leq i \leq r$. Therefore, the simple components $B_1, \ldots, B_r$ of a semisimple ring do not depend on a decomposition of $R$ as a direct sum of minimal left ideals.*

**Proof.** This follows from Lemma C-2.40(ii). •

**Corollary C-2.42.** *Let $A$ be a simple artinian ring.*

(i) *$A \cong \mathrm{Mat}_n(k)$ for some division ring $k$. If $L$ is a minimal left ideal in $A$, then every simple left $A$-module is isomorphic to $L$; moreover, $k^{\mathrm{op}} \cong \mathrm{End}_A(L)$.*

(ii) *Two finitely generated left $A$-modules $M$ and $N$ are isomorphic if and only if $\dim_k(M) = \dim_k(N)$.*

**Proof.**

(i) Since $A$ is a semisimple ring, by Proposition C-2.38, every left module $M$ is isomorphic to a direct sum of minimal left ideals. By Lemma C-2.40(ii), all minimal left ideals are isomorphic, say, to $L$.

We now prove that $k^{\mathrm{op}} \cong \mathrm{End}_A(L)$. We may assume that $A = \mathrm{Mat}_n(k)$ and that $L = \mathrm{COL}(1)$, the minimal left ideal consisting of all the $n \times n$ matrices whose last $n-1$ columns are 0 (Proposition C-2.27). Define $\varphi\colon k \to \mathrm{End}_A(L)$ as follows: if $d \in k$ and $\ell \in L$, then $\varphi_d\colon \ell \mapsto \ell d$. Note that $\varphi_d$ is an $A$-map: it is additive and, if $a \in A$ and $\ell \in L$, then $\varphi_d(a\ell) = (a\ell)d = a(\ell d) = a\varphi_d(\ell)$. Next, $\varphi$ is a ring antihomomorphism: $\varphi_1 = 1_L$, it is additive, and $\varphi_{dd'} = \varphi_{d'}\varphi_d$: if $\ell \in L$, then $\varphi_{d'}\varphi_d(\ell) = \varphi_d(\ell d') = \ell d'd = \varphi_{dd'}(\ell)$; that is, $\varphi$ is a ring homomorphism $k^{\mathrm{op}} \to \mathrm{End}_A(L)$. To see that $\varphi$ is injective, note that each $\ell \in L \subseteq \mathrm{Mat}_n(k)$ is a matrix with entries in $k$; hence, $\ell d = 0$ implies $\ell = 0$. Finally, we show that $\varphi$ is surjective. Let $f \in \mathrm{End}_A(L)$. Now $L = AE_{11}$, where $E_{11}$ is the matrix unit (every simple module is generated by any nonzero element in it). If $u_i \in k$, let $[u_1, \ldots, u_n]$ denote the $n \times n$ matrix in $L$ whose first column is $(u_1, \ldots, u_n)^{\top}$ and whose other entries are all 0. Write $f(E_{11}) = [d_1, \ldots, d_n]$. If $\ell \in L$, then $\ell$ has the form $[u_1, \ldots, u_n]$, and using only the definition of matrix multiplication, it is easy to see that $[u_1, \ldots, u_n] = [u_1, \ldots, u_n]E_{11}$. Since $f$ is an $A$-map,

$$\begin{aligned} f([u_1, \ldots, u_n]) &= f([u_1, \ldots, u_n]E_{11}) \\ &= [u_1, \ldots, u_n]f(E_{11}) \\ &= [u_1, \ldots, u_n][d_1, \ldots, d_n] \\ &= [u_1, \ldots, u_n]d_1 = \varphi_{d_1}([u_1, \ldots, u_n]). \end{aligned}$$

Therefore, $f = \varphi_{d_1} \in \mathrm{im}\,\varphi$, as desired.

(ii) All minimal left ideals in $A$ are isomorphic to $L$, and so $M$ is a direct sum of $\dim_k(M)/n$ copies of $L$. If $M \cong N$ as left $\mathrm{Mat}_n(k)$-modules, then $M \cong N$ as left $k$-modules, and so $\dim_k(M) = \dim_k(N)$. Conversely, if $\dim_k(M) = nd = \dim_k(N)$, then both $M$ and $N$ are direct sums of $d$ copies of $L$, and hence $M \cong N$ as left $A$-modules. •

The number $m$ of simple components of $R$ is an invariant, for it is the number of nonisomorphic simple left $R$-modules (even better, we will see, in Theorem C-2.49, that if $R = \mathbb{C}G$, then $m$ is the number of conjugacy classes in $G$). However, there is a much stronger uniqueness result.

**Theorem C-2.43 (Wedderburn–Artin II).** *Every semisimple ring $R$ is a direct product,*

$$R \cong \mathrm{Mat}_{n_1}(k_1) \times \cdots \times \mathrm{Mat}_{n_r}(k_r),$$

*where $n_i \geq 1$, $k_i$ is a division ring, and the numbers $r$ and $n_i$, as well as the division rings $k_i$, are uniquely determined by $R$.*

**Proof.** Let $R$ be a semisimple ring, and let $R = B_1 \oplus \cdots \oplus B_r$ be a decomposition into simple components arising from some decomposition of $R$ as a direct sum of minimal left ideals. Suppose that $R = B'_1 \times \cdots \times B'_t$, where each $B'_\ell$ is a two-sided ideal that is also a simple ring. By Lemma C-2.40, each two-sided ideal $B'_\ell$ is a direct sum of $B_i$'s. But $B'_\ell$ cannot have more than one summand $B_i$, lest the simple ring $B'_\ell$ contain a proper nonzero two-sided ideal. Therefore, $t = r$ and, after reindexing, $B'_i = B_i$ for all $i$.

Dropping subscripts, it remains to prove that if $B = \mathrm{Mat}_n(k) \cong \mathrm{Mat}_{n'}(k') = B'$, then $n = n'$ and $k \cong k'$. In Proposition C-2.27, we proved that $\mathrm{COL}(\ell)$, consisting of the matrices with $j$th columns 0 for all $j \neq \ell$, is a minimal left ideal in $B$, so that $\mathrm{COL}(\ell)$ is a simple $B$-module. Therefore,

$$(0) \subseteq \mathrm{COL}(1) \subseteq [\mathrm{COL}(1) \oplus \mathrm{COL}(2)] \subseteq \cdots \subseteq [\mathrm{COL}(1) \oplus \cdots \oplus \mathrm{COL}(n)] = B$$

is a composition series of $B$ as a module over itself. By the Jordan–Hölder Theorem (Theorem B-1.38 in Part 1), $n$ and the factor modules $\mathrm{COL}(\ell)$ are invariants of $B$. Now $\mathrm{COL}(\ell) \cong \mathrm{COL}(1)$ for all $\ell$, by Corollary C-2.42, and so it suffices to prove that $k$ can be recaptured from $\mathrm{COL}(1)$. But this has been done in Corollary C-2.42(i): $k \cong \mathrm{End}_B(\mathrm{COL}(1))^{\mathrm{op}}$. •

The description of the group algebra $kG$ simplifies when the field $k$ is algebraically closed. Here is the most useful version of Maschke's Theorem.

**Corollary C-2.44 (Molien).** *If $G$ is a finite group and $k$ is an algebraically closed field whose characteristic does not divide $|G|$, then*

$$kG \cong \mathrm{Mat}_{n_1}(k) \times \cdots \times \mathrm{Mat}_{n_r}(k).$$

**Proof.** By Maschke's Theorem, $kG$ is a semisimple ring, and its simple components are isomorphic to matrix rings of the form $\mathrm{Mat}_n(D)$, where $D$ arises as $\mathrm{End}_{kG}(L)^{\mathrm{op}}$ for some minimal left ideal $L$ in $kG$. Therefore, it suffices to show that $\mathrm{End}_{kG}(L)^{\mathrm{op}} = D = k$.

Now $\mathrm{End}_{kG}(L)^{\mathrm{op}} \subseteq \mathrm{End}_k(L)^{\mathrm{op}}$, which is finite-dimensional over $k$ because $L$ is; hence, $D = \mathrm{End}_{kG}(L)^{\mathrm{op}}$ is finite-dimensional over $k$. Each $f \in \mathrm{End}_{kG}(L)$ is a $kG$-map, hence is a $k$-map; that is, $f(au) = af(u)$ for all $a \in k$ and $u \in L$. Therefore, the map $\varphi_a \colon L \to L$, given by $u \mapsto au$, commutes with $f$; that is, $k$ (identified with all $\varphi_a$) is contained in $Z(D)$, the center of $D$. If $\delta \in D$, then $\delta$ commutes with every element in $k$, and so $k(\delta)$, the subdivision ring generated by $k$ and $\delta$, is a

(commutative) field. As $D$ is finite-dimensional over $k$, so is $k(\delta)$; that is, $k(\delta)$ is a finite extension of the field $k$, and so $\delta$ is algebraic over $k$, by Proposition A-3.84 in Part 1. But $k$ is algebraically closed, so that $\delta \in k$ and $D = k$. •

The next corollary makes explicit a detail from the Wedderburn–Artin Theorem II, using Molien's simplification.

**Corollary C-2.45.** *If $G$ is a finite group and $L_i$ is a minimal left ideal in $\mathbb{C}G$, then $\mathbb{C}G = B_1 \oplus \cdots \oplus B_r$, where $B_i$ is the ideal generated by all the minimal left ideals isomorphic to $L_i$. If $\dim_{\mathbb{C}}(L_i) = n_i$, then*

$$B_i = \mathrm{Mat}_{n_i}(\mathbb{C}).$$

**Example C-2.46.** There are nonisomorphic finite groups $G$ and $H$ having isomorphic complex group algebras. If $G$ is an abelian group of order $d$, then $\mathbb{C}G$ is a direct product of matrix rings over $\mathbb{C}$, because $\mathbb{C}$ is algebraically closed. But $G$ abelian implies $\mathbb{C}G$ commutative. Hence, $\mathbb{C}G$ is the direct product of $d$ copies of $\mathbb{C}$ (for $\mathrm{Mat}_n(\mathbb{C})$ is commutative only when $n = 1$). It follows that if $H$ is any abelian group of order $d$, then $\mathbb{C}G \cong \mathbb{C}H$. In particular, $\mathbb{Z}_4$ and $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ are nonisomorphic groups having isomorphic complex group algebras. It follows from this example that certain properties of a group $G$ get lost in the group algebra $\mathbb{C}G$. ◀

**Corollary C-2.47.** *If $G$ is a finite group and $k$ is an algebraically closed field whose characteristic does not divide $|G|$, then*

$$|G| = n_1^2 + n_2^2 + \cdots + n_r^2,$$

*where the $i$th simple component $B_i$ of $kG$ consists of $n_i \times n_i$ matrices. Moreover, we may assume that $n_1 = 1$.*[7]

**Remark.** Theorem C-2.97 says that all the $n_i$ are divisors of $|G|$. ◀

**Proof.** As vector spaces over $k$, both $kG$ and $\mathrm{Mat}_{n_1}(k) \times \cdots \times \mathrm{Mat}_{n_r}(k)$ have the same dimension, for they are isomorphic, by Corollary C-2.44. But $\dim(kG) = |G|$, and the dimension of the right side is $\sum_i \dim(\mathrm{Mat}_{n_i}(k)) = \sum_i n_i^2$.

Finally, Example C-2.34 shows that there is a unique minimal left ideal isomorphic to the trivial module $V_0(k)$; the corresponding simple component, say, $B_1$, is one-dimensional, and so $n_1 = 1$. •

The number $m$ of simple components in $\mathbb{C}G$ has a group-theoretic interpretation; we begin by finding the center of the group algebra.

**Definition.** Let $C_1, \ldots, C_r$ be the conjugacy classes in a finite group $G$. For each $C_j$, define the ***class sum*** to be the element $z_j \in \mathbb{C}G$ given by

$$z_j = \sum_{g \in C_j} g.$$

Here is a ring-theoretic interpretation of the number $r$ of conjugacy classes.

[7] By Example C-2.34, the group algebra $kG$ always has a unique minimal left ideal isomorphic to $V_0(k)$, even when $k$ is not algebraically closed.

**Lemma C-2.48.** *If $c$ is the number of conjugacy classes in a finite group $G$, then*

$$c = \dim_{\mathbb{C}}(Z(\mathbb{C}G)),$$

*where $Z(\mathbb{C}G)$ is the center of the group algebra. In fact, a basis of $Z(\mathbb{C}G)$ consists of all the class sums.*

**Proof.** If $z_j = \sum_{g \in C_j} g$ is a class sum, then we claim that $z_j \in Z(\mathbb{C}G)$. If $h \in G$, then $hz_jh^{-1} = z_j$, because conjugation by any element of $G$ merely permutes the elements in a conjugacy class. Note that if $j \neq \ell$, then $z_j$ and $z_\ell$ have no nonzero components in common, and so $z_1, \ldots, z_r$ is a linearly independent list. It remains to prove that the $z_j$ span the center.

Let $u = \sum_{g \in G} a_g g \in Z(\mathbb{C}G)$. If $h \in G$, then $huh^{-1} = u$, and so $a_{hgh^{-1}} = a_g$ for all $g \in G$. Thus, if $g$ and $g'$ lie in the same conjugacy class of $G$, then their coefficients in $u$ are the same. But this says that $u$ is a linear combination of the class sums $z_j$. •

**Theorem C-2.49.** *If $G$ is a finite group, then the number $r$ of simple components in $\mathbb{C}G$ is equal to the number $c$ of conjugacy classes in $G$.*

**Proof.** We have just seen, in Lemma C-2.48, that $c = \dim_{\mathbb{C}}(Z(\mathbb{C}G))$. On the other hand, $Z(\mathrm{Mat}_{n_i}(\mathbb{C}))$, the center of a matrix ring, is the subspace of all scalar matrices, so that $r = \dim_{\mathbb{C}}(Z(\mathbb{C}G))$, by Lemma C-2.48. •

We began this section by seeing that $k$-representations of a group $G$ correspond to $kG$-modules. Let us now return to representations.

**Definition.** A $k$-representation of a group $\sigma\colon G \to \mathrm{GL}(V)$ is ***irreducible*** if the corresponding $kG$-module $V^\sigma$ is simple.

For example, a one-dimensional (necessarily irreducible) $k$-representation is a group homomorphism $\lambda\colon G \to k^\times$, where $k^\times$ is the multiplicative group of nonzero elements of $k$. The trivial $kG$-module $V_0(k)$ corresponds to the representation $\lambda_g = 1$ for all $g \in G$.

**Theorem C-2.50.** *Every irreducible representation of a finite abelian group $G$ is linear.*

**Proof.** By the Wedderburn–Artin Theorem I, $\mathbb{C}G$ is a direct sum of matrix rings $\mathbb{C}G = L_1 \oplus \cdots \oplus L_t$, where each $L_j \cong \mathrm{Mat}_{n_j}(\mathbb{C})$. Moreover, Theorem C-2.33 says that every irreducible $\mathbb{C}G$-module is isomorphic to some $L_j$. It follows that $L_j$ is one-dimensional, for $\mathbb{C}G$ is commutative (because $G$ is abelian) and $L_j \cong \mathrm{Mat}_{n_j}(\mathbb{C})$ is noncommutative if $n_j > 1$. •

The next result is basic to the construction of the character table of a finite group.

**Theorem C-2.51.** *If $G$ is a finite group, then the number of its irreducible complex representations is equal to the number $r$ of its conjugacy classes.*

**Proof.** By Theorem C-2.33, every simple $\mathbb{C}G$-module is isomorphic to a minimal left ideal. Since the number of minimal left ideals is $r$ (the number of simple components of $\mathbb{C}G$), we see that $r$ is the number of irreducible $\mathbb{C}$-representations of $G$. But Theorem C-2.49 equates $r$ with the number $c$ of conjugacy classes in $G$. •

**Example C-2.52.**

(i) If $G = S_3$, then $\mathbb{C}G$ is six-dimensional. There are three simple components, for $S_3$ has three conjugacy classes (by Theorem A-4.7 in Part 1, the number of conjugacy classes in $S_n$ is equal to the number of different cycle structures) having dimensions 1, 1, and 4, respectively. (We could have seen this without Theorem C-2.49, for this is the only way to write 6 as a sum of squares aside from a sum of six 1's.) Therefore,

$$\mathbb{C}S_3 \cong \mathbb{C} \times \mathbb{C} \times \mathrm{Mat}_2(\mathbb{C}).$$

One of the one-dimensional irreducible representations is the trivial one; the other is sgn (= signum).

(ii) We now analyze $kG$ for $G = \mathbf{Q}$, the quaternion group of order 8. If $k = \mathbb{C}$, then Corollary C-2.44 gives

$$\mathbb{C}\mathbf{Q} \cong \mathrm{Mat}_{n_1}(\mathbb{C}) \times \cdots \times \mathrm{Mat}_{n_r}(\mathbb{C}),$$

while Corollary C-2.47 gives

$$|\mathbf{Q}| = 8 = 1 + n_2^2 + \cdots + n_r^2.$$

It follows that either all $n_i = 1$ or four $n_i = 1$ and one $n_i = 2$. The first case cannot occur, for it would imply that $\mathbb{C}\mathbf{Q}$ is a commutative ring, whereas the group $\mathbf{Q}$ of quaternions is not abelian. Therefore,

$$\mathbb{C}\mathbf{Q} \cong \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathrm{Mat}_2(\mathbb{C}).$$

We could also have used Theorem C-2.49, for $\mathbf{Q}$ has exactly five conjugacy classes, namely, $\{1\}$, $\{\overline{1}\}$, $\{i, \overline{i}\}$, $\{j, \overline{j}\}$, $\{k, \overline{k}\}$.

The group algebra $\mathbb{R}\mathbf{Q}$ is more complicated because $\mathbb{R}$ is not algebraically closed. Exercise B-1.14 on page 281 in Part 1 shows that $\mathbb{H}$ is a quotient of $\mathbb{R}\mathbf{Q}$, hence $\mathbb{H}$ is isomorphic to a direct summand of $\mathbb{R}\mathbf{Q}$ because $\mathbb{R}\mathbf{Q}$ is semisimple. It turns out that

$$\mathbb{R}\mathbf{Q} \cong \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{H}. \quad \blacktriangleleft$$

Here is an amusing application of the Wedderburn–Artin Theorems.

**Proposition C-2.53.** *Let $R$ be a ring whose group of units $U = U(R)$ is finite and of odd order. Then $U$ is abelian and there are positive integers $m_i$ with*

$$|U| = \prod_{i=1}^{t} (2^{m_i} - 1).$$

**Proof.** First, we note that $1 = -1$ in $R$, lest $-1$ be a unit of even order. Consider the group algebra $kU$, where $k = \mathbb{F}_2$. Since $k$ has characteristic 2 and $|U|$ is odd, Maschke's Theorem says that $kU$ is semisimple. There is a ring map $\varphi\colon kU \to R$ carrying every $k$-linear combination of elements of $U$ to "itself". Now $R' = \operatorname{im}\varphi$

is a finite subring of $R$ containing $U$ (for $kU$ is finite); since dropping to a subring cannot create any new units, we have $U = U(R')$. By Corollary C-2.19, the ring $R'$ is semisimple, so that Wedderburn–Artin Theorem I gives

$$R' \cong \prod_{i=1}^{t} \mathrm{Mat}_{n_i}(k_i),$$

where each $k_i$ is a division ring.

Now $k_i$ is finite, because $R'$ is finite, and so $k_i$ is a finite division ring. By the "other" theorem of Wedderburn, Theorem C-2.31, each $k_i$ is a field. But $-1 = 1$ in $R$ implies that $-1 = 1$ in $k_i$ (for $k_i \subseteq R'$), and so each field $k_i$ has characteristic 2; hence,

$$|k_i| = 2^{m_i}$$

for integers $m_i \geq 1$. All the matrix rings must be $1 \times 1$, for any matrix ring of larger size must contain an element of order 2, namely, $I + K$, where $K$ has entry 1 in the first position in the bottom row, and all other entries 0. For example,

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} = I.$$

Therefore, $R'$ is a direct product of finite fields of characteristic 2, and so $U = U(R')$ is an abelian group whose order is described in the statement. •

It follows, for example, that there is no ring having exactly five units.

The ***Jacobson–Chevalley Density Theorem***, an important generalization of the Wedderburn–Artin Theorems for certain nonartinian rings, was proved in the 1930s. Call a ring $R$ ***left primitive*** if there exists a faithful simple left $R$-module $S$; that is, $S$ is simple and, if $r \in R$ and $rS = \{0\}$, then $r = 0$. It can be proved that commutative primitive rings are fields, while left artinian left primitive rings are simple. Assume now that $R$ is a left primitive ring, that $S$ is a faithful simple left $R$-module, and that $k$ denotes the division ring $\mathrm{End}_R(S)$. The Density Theorem says that if $R$ is left artinian, then $R \cong \mathrm{Mat}_n(k)$, while if $R$ is not left artinian, then for every integer $n > 0$, there exists a subring $R_n$ of $R$ with $R_n \cong \mathrm{Mat}_n(k)$. We refer the reader to Lam [**133**], pp. 191–193.

The Wedderburn–Artin Theorems led to several areas of research, two of which are descriptions of division rings and of finite-dimensional algebras. Division rings will be considered later in the context of central simple algebras.

Let us discuss finite-dimensional algebras now. Thanks to the theorems of Maschke and Molien, the Wedderburn–Artin Theorems apply to ***ordinary*** representations of a finite group $G$, that is, to $kG$-modules, where $k$ is a field whose characteristic does not divide $|G|$, for $kG$ is semisimple in this case. However, ***modular*** representations, that is, $kG$-modules for which the characteristic of $k$ does divide $|G|$, arise naturally. For example, if $G$ is a finite solvable group, then a minimal normal subgroup $N$ is a vector space over $\mathbb{F}_p$ for some prime $p$ (Theorem C-1.54). Now $G$ acts on $N$ (by conjugation), and so $N$ is an $\mathbb{F}_pG$-module. Modular representations are used extensively in the classification of the finite simple groups.

In his study of modular representations, Brauer observed that the important modules $M$ are indecomposable rather than irreducible. Recall that a module $M$ is ***indecomposable*** if $M \neq \{0\}$ and there are no nonzero modules $A$ and $B$ with $M = A \oplus B$. In the ordinary case, a module is indecomposable if and only if it is irreducible, but this is no longer true in the modular case.

There is a uniqueness theorem for direct sums.

**Theorem** (**Krull–Schmidt**).[8] *Let $A$ be a left $R$-module over some ring $R$. If $A$ has both chain conditions on its submodules and*

$$A = H_1 \oplus \cdots \oplus H_s = K_1 \oplus \cdots \oplus K_t,$$

*where the $H_i, K_j$ are indecomposable modules, then $s = t$ and, after reindexing, $H_i \cong K_i$ for all $i$. Moreover, there is a **replacement property**: given any $r$ between 1 and $s$, the reindexing may be chosen so that*

$$G = H_1 \oplus \cdots \oplus H_r \oplus K_{r+1} \oplus \cdots \oplus K_s.$$

Given the Basis Theorem, the Krull–Schmidt Theorem implies the Fundamental Theorem of Finite Abelian Groups.

**Corollary** *If a finite abelian group $G$ is a direct sum*

$$G = H_1 \oplus \cdots \oplus H_s = K_1 \oplus \cdots \oplus K_t,$$

*where each $H_i$ and $K_j$ is a primary cyclic group, then $s = t$ and after reindexing, $H_i \cong K_i$ for all $i$.*

**Proof.** Cyclic primary groups are indecomposable. We assume that $G$ is finite so that it has both chain conditions on its subgroups. •

When $kG$ is semisimple, Theorem C-2.33 says that there are only finitely many simple modules, which implies that there are only finitely many indecomposables. This is not true in the modular case, however. For example, if $k$ is an algebraically closed field of characteristic 2, $k\mathbf{V}$ and $kA_4$ have infinitely many nonisomorphic indecomposable modules.

A finite-dimensional $k$-algebra $R$ over a field $k$ is said to have ***finite representation type*** if there are only finitely many nonisomorphic finite-dimensional indecomposable $R$-modules. D. G. Higman proved, for a finite group $G$, that $kG$ has finite representation type for every field $k$ if and only if all its Sylow subgroups $G$ are cyclic (Curtis–Reiner [**48**], p. 431). In the 1950s, the following two problems, known as the ***Brauer–Thrall conjectures***, were posed. Let $R$ be a ring not of finite representation type.

(I) Are the dimensions of the indecomposable $R$-modules unbounded?

(II) Is there a strictly increasing sequence $n_1 < n_2 < \cdots$ with infinitely many nonisomorphic indecomposable $R$-modules of dimension $n_i$ for every $i$?

[8] The Krull–Schmidt Theorem also holds for direct products of nonabelian groups (Rotman [**188**], p. 149). The proof generalizes to ***groups with operators***, a special case of which is modules.

The positive solution of the first conjecture, by Roiter in 1968, had a great impact. Shortly thereafter, Gabriel [**78**] introduced graph-theoretic methods, associating finite-dimensional algebras to certain oriented graphs, called *quivers*. He proved that a connected quiver has a finite number of nonisomorphic finite-dimensional representations if and only if the quiver is a Dynkin diagram of type $A_n$, $D_n$, $E_6$, $E_7$, or $E_8$. ***Dynkin diagrams*** are multigraphs that classify simple complex Lie algebras; in addition to those just mentioned, there are diagrams $B_n$, $C_n$, $F_4$, and $G_2$ (see page 167). Gabriel's result can be rephrased in terms of *left hereditary* $k$-algebras (all left ideals are projective modules). Dlab and Ringel generalized this classification to include other left hereditary algebras, thereby extending Gabriel's result to all Dynkin diagrams.

A positive solution of the Brauer–Thrall conjecture (II) above for all finite-dimensional algebras over an algebraically closed field follows from results of Bautista, Gabriel, Roiter, and Salmerón. M. Auslander and Reiten created a theory involving *almost split sequences* and *Auslander–Reiten quivers*. As of this writing, Auslander–Reiten theory is the most powerful tool in the study of representations of finite-dimensional algebras. We refer the reader to Artin–Nesbitt–Thrall [**9**], Dlab–Ringel [**54**], Drozd–Kirichenko [**57**], Jacobson [**114**], and Rowen [**194**] for a discussion of these ideas.

## Exercises

**C-2.18.** If $k$ is a division ring whose center is a field of characteristic $p > 0$, prove that every finite subgroup $G$ of $k^\times$ is cyclic.

**Hint.** Consider $\mathbb{F}_p G$, and use Theorem C-2.31.

**C-2.19.** Find $\mathbb{C}G$ if $G = D_8$, the dihedral group of order 8.

**C-2.20.** Find $\mathbb{C}G$ if $G = A_4$.

**Hint.** $A_4$ has four conjugacy classes.

**C-2.21.** (i) Let $k$ be a field, and view $\operatorname{sgn}\colon S_n \to \{\pm 1\} \subseteq k$. Define $\operatorname{Sig}(k)$ to be $k$ made into a $kS_n$-module (as in Proposition C-2.13): if $\gamma \in S_n$ and $a \in k$, then $\gamma a = \operatorname{sgn}(\gamma)a$. Prove that if $\operatorname{Sig}(k)$ is an irreducible $kS_n$-module and $k$ does not have characteristic 2, then $\operatorname{Sig}(k) \ncong V_0(k)$.

(ii) Find $\mathbb{C}S_5$.
**Hint.** There are five conjugacy classes in $S_5$.

**C-2.22.** Let $G$ be a finite group, and let $k$ and $K$ be algebraically closed fields whose characteristics $p$ and $q$, respectively, do not divide $|G|$.

(i) Prove that $kG$ and $KG$ have the same number of simple components.

(ii) Prove that the degrees of the irreducible representations of $G$ over $k$ are the same as the degrees of the irreducible representations of $G$ over $K$.

## C-2.6. Introduction to Lie Algebras

We have just discussed semisimple rings in general; the next step in studying group representations is to discuss group algebras $kG$ in particular. Since we have mentioned Dynkin diagrams, however, we say a bit about them here. Readers who prefer more groups now, however, may skip this section and proceed to studying characters.

There are interesting examples of nonassociative algebras, the most important of which are Lie algebras. In the late nineteenth century, Lie (pronounced LEE) studied the solution space $S$ of a system of partial differential equations using a group $G$ of transformations of $S$. The underlying set of $G$ is a differentiable manifold and its group operation is a $C^\infty$-function; such groups are called ***Lie groups***. The solution space is intimately related to its Lie group $G$; in turn, $G$ is studied using its *associated Lie algebra*, a considerably simpler object, which arises as the tangent space at the identity element of $G$. Moreover, the classification of the simple finite-dimensional complex Lie algebras, due to Killing and Cartan at the turn of the twentieth century, served as a model for the recent classification of all finite simple groups. Chevalley recognized that one could construct analogous families of finite simple groups by imitating the construction of simple Lie algebras, and Chevalley's groups are called ***groups of Lie type***. Aside from this fundamental reason for their study, Lie algebras turn out to be the appropriate way to deal with families of linear transformations on a vector space (in contrast to the study of canonical forms of a single linear transformation given in the first sections of Chapter C-1).

Before defining Lie algebras, we first generalize the notion of derivation.

**Definition.** A ***not-necessarily-associative $k$-algebra*** $A$ over a commutative ring $k$ is a $k$-module equipped with a binary operation $A \times A \to A$, denoted by $(a, b) \mapsto ab$, such that

(i) $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ for all $a, b, c \in A$;

(ii) $ua = au$ for all $u \in k$ and $a \in A$;

(iii) $a(ub) = (au)b = u(ab)$ for all $u \in k$ and $a, b \in A$.

A ***derivation*** of $A$ is a $k$-map $d\colon A \to A$ such that, for all $a, b \in A$,

$$d(ab) = (da)b + a(db).$$

Aside from ordinary differentiation in calculus, which is a derivation because the product rule $(fg)' = f'g + fg'$ holds, another example is provided by the $\mathbb{R}$-algebra $A$ of all real-valued functions $f(x_1, \ldots, x_n)$ of several variables: the partial derivatives $\partial/\partial x_i$ are derivations, for $i = 1, \ldots, n$.

The composite of two derivations need not be a derivation. For example, if $d\colon A \to A$ is a derivation, then $d^2 = d \circ d\colon A \to A$ satisfies the equation

$$d^2(fg) = d^2(f)g + 2d(f)d(g) + fd^2(g);$$

thus, the mixed term $2d(f)d(g)$ is the obstruction to $d^2$ being a derivation. The Leibniz formula for ordinary differentiation on the ring of all $C^\infty$-functions generalizes to derivations on any not-necessarily-associative algebra $A$. If $f, g \in A$,

then

$$d^n(fg) = \sum_{i=0}^{n} \binom{n}{i} d^i f \cdot d^{n-i} g.$$

It is worthwhile to compute the composite of two derivations $d_1$ and $d_2$. If $A$ is a not-necessarily-associative algebra and $f, g \in A$, then

$$\begin{aligned} d_1 d_2(fg) &= d_1\left[(d_2 f)g + f(d_2 g)\right] \\ &= (d_1 d_2 f)g + (d_2 f)(d_1 g) + (d_1 f)(d_2 g) + f(d_1 d_2 g). \end{aligned}$$

Of course,

$$d_2 d_1(fg) = (d_2 d_1 f)g + (d_1 f)(d_2 g) + (d_2 f)(d_1 g) + f(d_2 d_1 g).$$

If we denote $d_1 d_2 - d_2 d_1$ by $[d_1, d_2]$, then

$$[d_1, d_2](fg) = ([d_1, d_2]f)g + f([d_1, d_2]g);$$

that is, $[d_1, d_2] = d_1 d_2 - d_2 d_1$ is a derivation.

**Example C-2.54.** If $k$ is a commutative ring, equip $\mathrm{Mat}_n(k)$ with the ***bracket operation***:

$$[A, B] = AB - BA.$$

Of course, $A$ and $B$ commute if and only if $[A, B] = 0$. It is easy to find examples showing that the bracket operation is not associative. However, for any fixed $n \times n$ matrix $M$, the function

$$\mathrm{ad}_M \colon \mathrm{Mat}_n(k) \to \mathrm{Mat}_n(k),$$

defined by

$$\mathrm{ad}_M \colon A \mapsto [M, A],$$

is a derivation:

$$[M, [A, B]] = [[M, A], B] + [A, [M, B]].$$

The verification of this identity should be done once in one's life. ◀

The definition of Lie algebra involves a vector space with a binary operation generalizing the bracket operation.

**Definition.** A ***Lie algebra*** is a vector space $L$ over a field $k$ equipped with a bilinear operation $L \times L \to L$, denoted by $(a, b) \mapsto [a, b]$ (and called ***bracket***), such that

(i) $[a, a] = 0$ for all $a \in L$;

(ii) for each $a \in L$, the function $\mathrm{ad}_a \colon b \mapsto [a, b]$ is a derivation.

For all $u, v \in L$, bilinearity gives

$$[u + v, u + v] = [u, u] + [u, v] + [v, u] + [v, v],$$

which, when coupled with the first axiom $[a, a] = 0$, gives

$$[u, v] = -[v, u];$$

that is, bracket is ***anticommutative***. The second axiom is often written out in more detail. If $b, c \in L$, then their product in $L$ is $[b, c]$; that $\mathrm{ad}_a$ is a derivation is to say that

$$[a, [b, c]] = [[a, b], c] + [b, [a, c]];$$

rewriting,

$$[a, [b, c]] - [b, [a, c]] - [[a, b], c] = 0.$$

The anticommutativity from the first axiom now gives the ***Jacobi identity***:

$$[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0 \quad \text{for all } a, b, c \in L$$

(just cyclically permute $a, b, c$). Thus, a vector space $L$ is a Lie algebra if and only if $[a, a] = 0$ for all $a \in L$ and the Jacobi identity holds.

Here are some examples of Lie algebras.

**Example C-2.55.**

(i) If $V$ is a vector space over a field $k$, define $[a, b] = 0$ for all $a, b \in V$. It is obvious that $V$ so equipped is a Lie algebra; it is called an ***abelian*** Lie algebra.

(ii) In $\mathbb{R}^3$, define $[u, v] = u \times v$, the ***cross product*** (or *vector product*) defined in calculus. It is routine to check that $v \times v = 0$ and that the Jacobi identity holds, so that $\mathbb{R}^3$ is a Lie algebra. This example may be generalized: for every field $k$, cross product can be defined on the vector space $k^3$ making it a Lie algebra.

(iii) A ***subalgebra*** $S$ of a Lie algebra $L$ is a subspace that is closed under bracket: if $a, b \in S$, then $[a, b] \in S$. It is easy to see that every subalgebra of a Lie algebra is itself a Lie algebra.

(iv) If $k$ is a field, then $\mathrm{Mat}_n(k)$ is a Lie algebra with bracket defined by

$$[A, B] = AB - BA.$$

We usually denote this Lie algebra by $\mathfrak{gl}(n, k)$. This example is quite general, for it is a theorem of Ado that every finite-dimensional Lie algebra over a field $k$ of characteristic 0 is isomorphic to a subalgebra of $\mathfrak{gl}(n, k)$ for some $n$ (Jacobson [**113**], p. 202).

(v) An interesting subalgebra of $\mathfrak{gl}(n, k)$ is $\mathfrak{sl}(n, k)$, which consists of all $n \times n$ matrices of trace 0. In fact, if $G$ is a Lie group whose associated Lie algebra is $\mathfrak{g}$, then there is an analog of exponentiation $\mathfrak{g} \to G$. In particular, if $\mathfrak{g} = \mathfrak{gl}(n, \mathbb{C})$, then this map is exponentiation $A \mapsto e^A$. Thus, Proposition B-3.71(viii) in Part 1 shows that exponentiation sends $\mathfrak{sl}(n, \mathbb{C})$ into $\mathrm{SL}(n, \mathbb{C})$.

(vi) If $A$ is any algebra over a field $k$, then

$$\mathfrak{Der}(A/k) = \{\text{all derivations } d\colon A \to A\},$$

with bracket $[d_1, d_2] = d_1 d_2 - d_2 d_1$, is a Lie algebra.

It follows from the Leibniz rule that if $k$ has characteristic $p > 0$, then $d^p$ is a derivation for every $d \in \mathfrak{Der}(A/k)$, since $\binom{p}{i} \equiv 0 \bmod p$ whenever

$0 < i < p$. (This is an example of what is called a ***restricted Lie algebra*** of characteristic $p$.)

There is a Galois theory for certain purely inseparable extensions (Jacobson [**111**], pp. 533–536). If $k$ is a field of characteristic $p > 0$ and $E/k$ is a finite purely inseparable extension of *height* 1, that is, $\alpha^p \in k$ for all $\alpha \in E$, then there is a bijection between the family of all intermediate fields and the restricted Lie subalgebras of $\mathfrak{Der}(E/k)$, given by

$$B \mapsto \mathfrak{Der}(E/B);$$

the inverse of this function is given by

$$\mathfrak{L} \mapsto \big\{e \in E : D(e) = 0 \text{ for all } D \in \mathfrak{L}\big\}. \quad \blacktriangleleft$$

Not surprisingly, all Lie algebras over a field $k$ form a category.

**Definition.** Let $L$ and $L'$ be Lie algebras over a field $k$. Then a function $f\colon L \to L'$ is a ***Lie homomorphism*** if $f$ is a $k$-linear transformation that preserves bracket: for all $a, b \in L$,

$$f([a, b]) = [fa, fb].$$

**Definition.** An ***ideal*** of a Lie algebra $L$ is a subspace $I$ such that $[x, a] \in I$ for every $x \in L$ and $a \in I$.

Even though a Lie algebra need not be commutative, its anticommutativity shows that every left ideal (as just defined) is necessarily a right ideal; that is, every ideal is two-sided.

A Lie algebra $L$ is called ***simple*** if $L \neq \{0\}$ and $L$ has no nonzero proper ideals.

**Definition.** If $I$ is an ideal in $L$, then the ***quotient*** $L/I$ is the quotient space (considering $L$ as a vector space and $I$ as a subspace) with bracket defined by

$$[a + I, b + I] = [a, b] + I.$$

It is easy to check that this bracket on $L/I$ is well-defined. If $a' + I = a + I$ and $b' + I = b + I$, then $a - a' \in I$ and $b - b' \in I$, and so

$$\begin{aligned}[a', b'] - [a, b] &= [a', b'] - [a', b] + [a', b] - [a, b] \\ &= [a', b' - b] + [a' - a, b'] \in I.\end{aligned}$$

**Example C-2.56.**

(i) If $f\colon L \to L'$ is a Lie homomorphism, then its ***kernel*** is defined as usual:

$$\ker f = \{a \in L : f(a) = 0\}.$$

It is easy to see that $\ker f$ is an ideal in $L$.

Conversely, the ***natural map*** $\nu\colon L \to L/I$, defined by $a \mapsto a + I$, is a Lie homomorphism whose kernel is $I$. Thus, a subspace of $L$ is an ideal if and only if it is the kernel of some Lie homomorphism.

(ii) If $I$ and $J$ are ideals in a Lie algebra $L$, then

$$IJ = \Big\{ \sum_r [i_r, j_r] : i_r \in I \text{ and } j_r \in J \Big\}.$$

In particular, $L^2 = LL$ is the analog for Lie algebras of the commutator subgroup of a group: $L^2 = \{0\}$ if and only if $L$ is abelian.

(iii) There is an analog for Lie algebras of the derived series of a group. The ***derived series*** of a Lie algebra $L$ is defined inductively:

$$L^{(0)} = L, \quad L^{(n+1)} = (L^{(n)})^2.$$

A Lie algebra $L$ is called ***solvable*** if there is some $n \geq 0$ with $L^{(n)} = \{0\}$.

(iv) There is an analog for Lie algebras of the descending central series of a group. The ***descending central series*** is defined inductively:

$$L_1 = L, \quad L_{n+1} = LL_n.$$

A Lie algebra $L$ is called ***nilpotent*** if there is some $n \geq 0$ with $L_n = \{0\}$. ◀

We merely mention the first two theorems in the subject. If $L$ is a Lie algebra and $a \in L$, then $\mathrm{ad}_a \colon L \to L$, given by $\mathrm{ad}_a \colon x \mapsto [a, x]$, is a linear transformation on $L$ (viewed merely as a vector space). We say that $a$ is ***ad-nilpotent*** if $\mathrm{ad}_a$ is a nilpotent operator; that is, $(\mathrm{ad}_a)^m = 0$ for some $m \geq 1$.

**Theorem C-2.57 (Engel's Theorem).**

(i) *Let $L$ be a finite-dimensional Lie algebra over any field $k$. Then $L$ is nilpotent if and only if every $a \in L$ is ad-nilpotent.*

(ii) *Let $L$ be a Lie subalgebra of $\mathfrak{gl}(n, k)$ all of whose elements $A$ are nilpotent matrices. Then $L$ can be put into strict upper triangular form (all diagonal entries are 0); that is, there is a nonsingular matrix $P$ so that $PAP^{-1}$ is strictly upper triangular for every $A \in L$.*

**Proof.** Humphreys [**101**], p. 12. •

Compare Engel's Theorem with Exercise B-3.40 in Part 1, which is the much simpler version for a single nilpotent matrix. Nilpotent Lie algebras are so called because of Engel's Theorem; it is likely that nilpotent *groups* are so called by analogy with Engel's Theorem. Corollary C-1.47, which states that every finite $p$-group can be imbedded as a subgroup of unitriangular matrices over $\mathbb{F}_p$, may be viewed as a group-theoretic analog of Engel's Theorem.

**Theorem C-2.58 (Lie's Theorem).** *Every solvable subalgebra $L$ of $\mathfrak{gl}(n, k)$, where $k$ is an algebraically closed field, can be put into (not necessarily strict) upper triangular form; that is, there is a nonsingular matrix $P$ so that $PAP^{-1}$ is upper triangular for every $A \in L$.*

**Proof.** Humphreys [**101**], p. 16. •

Further study of Lie algebras leads to the classification of all finite-dimensional simple Lie algebras over an algebraically closed field of characteristic 0, due to Cartan and Killing (see Humphreys [**101**], Chapter IV, and Jacobson [**113**], Chapter IV). To each such algebra, they associated a certain geometric configuration called a *root system*, which is characterized by a *Cartan matrix*. Cartan matrices are, in turn, characterized by the ***Dynkin diagrams***. Thus, Dynkin diagrams arise from simple Lie algebras over $\mathbb{C}$, and two such algebras are isomorphic if and only if they have the same Dynkin diagram. This study uses representations analogous to representations of groups. A ***representation*** of a Lie algebra $L$ is a Lie homomorphism $\sigma\colon L \to \mathfrak{gl}(n, \mathbb{C})$. The ***radical*** $\mathrm{Rad}(L)$ of a Lie algebra $L$ is the largest solvable ideal ($\mathrm{Rad}(L)$ exists and is unique). Amongst other results, there is an analog of Wedderburn's Theorem: ***Weyl's Theorem*** says that every Lie algebra $L$ with $\mathrm{Rad}(L) = \{0\}$ is a direct sum of simple Lie algebras ([**101**], p. 28). Recently, Block, Premet, Strade, and Wilson classified all finite-dimensional simple Lie algebras over algebraically closed fields of characteristic $p \geq 5$: there are two other types aside from analogs given by Dynkin diagrams (see Strade [**215**]).

Dynkin diagrams have also appeared in work of Gabriel [**78**] in classifying finite-dimensional algebras over fields.

There are other not-necessarily-associative algebras of interest. ***Jordan algebras*** are algebras $A$ which are commutative (instead of anticommutative) and in which the Jacobi identity is replaced by

$$(x^2y)x = x^2(yx)$$

for all $x, y \in A$. These algebras were introduced by P. Jordan to provide an algebraic setting for doing quantum mechanics. An example of a Jordan algebra is a subspace of all $n \times n$ matrices, over a field of characteristic not 2, equipped with the binary operation $A * B$, where

$$A * B = \tfrac{1}{2}(AB + BA).$$

Another source of not-necessarily-associative algebras comes from combinatorics. The usual construction of a projective plane $P(k)$ over a field $k$, as the family of all lines in $k^3$ passing through the origin, leads to descriptions of its points by "homogeneous coordinates" $[x, y, z]$, where $x, y, z \in k$. Define an abstract ***projective plane*** to be an ordered pair $(X, \mathcal{L})$, where $X$ is a finite set and $\mathcal{L}$ is a family of subsets of $X$, called *lines*, subject to the following axioms:

(i) All lines have the same number of points.

(ii) Given any two points in $X$, there is a unique line containing them.

We want to introduce homogeneous coordinates to describe the points of such a projective plane, but there is no field $k$ given at the outset. Instead, we look at a collection $\mathcal{K}$ of functions on $X$, called *collineations*, and we equip $\mathcal{K}$ with two binary operations (called addition and multiplication). In general, $\mathcal{K}$ is a not-necessarily-associative algebra, but certain algebraic properties follow from the geometry of the plane. A theorem of Desargues holds if and only if multiplication is associative and $\mathcal{K}$ is a division ring; a theorem of Pappus also holds if and only if multiplication is commutative and $\mathcal{K}$ is a field (Reid–Szendrői [**179**], p. 88). Thus, Wedderburn's

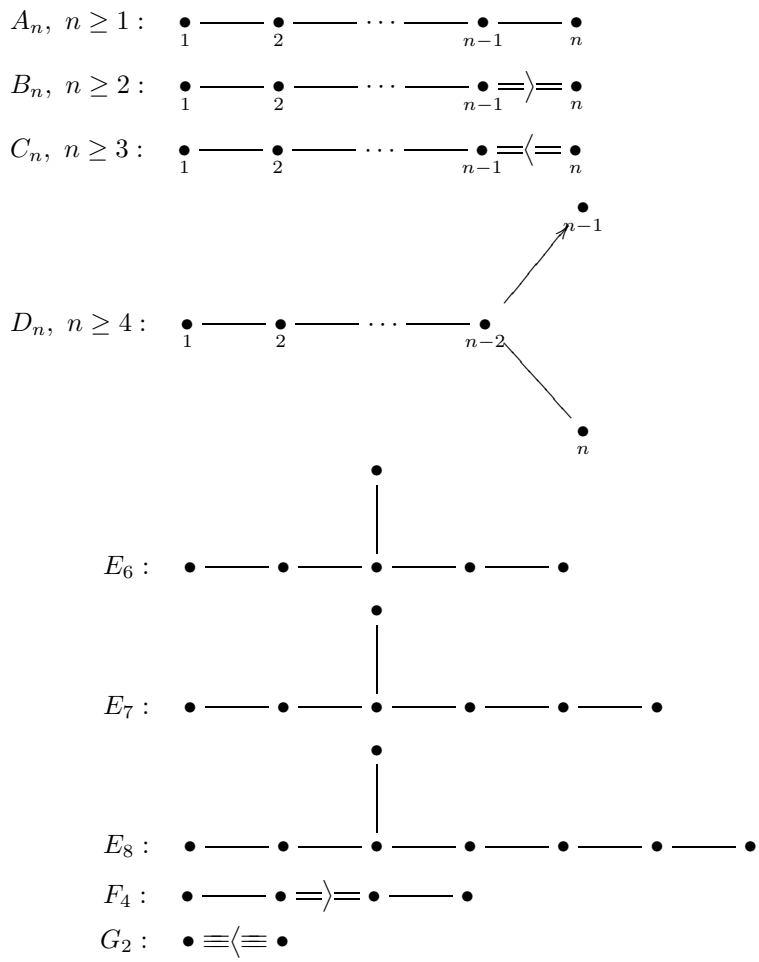



**Figure C-2.1.** Dynkin diagrams.

Theorem C-2.31 implies that if Desargues's Theorem holds in a finite projective plane, then so does the Pappus Theorem.

An interesting nonassociative algebra is the algebra of ***Cayley numbers*** (sometimes called *octonions*), which is an eight-dimensional real vector space containing the quaternions as a subalgebra (see the article by Curtis in Albert [**3**]). Indeed, the Cayley numbers almost form a real division algebra ("almost", for even though every nonzero element has a multiplicative inverse, multiplication in the Cayley numbers is not associative). The algebra of Cayley numbers acquires added interest (as do other not-necessarily-associative algebras) because its automorphism group has interesting properties. For example, the exceptional simple Lie algebra

$E_8$ is isomorphic to the Lie algebra of all the derivations of the Cayley numbers, while the Monster, the largest sporadic finite simple group, is the automorphism group of a certain nonassociative algebra constructed by Griess.

## Exercises

∗ **C-2.23.** Consider the ***de Rham complex*** when $n = 2$:

$$0 \to \Delta^0(X) \xrightarrow{d^0} \Delta^1(X) \xrightarrow{d^1} \Delta^2(X) \to 0.$$

Prove that if $f(x, y) \in A(X) = \Delta^0(X)$, then

$$d^0 f = \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy,$$

and that if $P dx + Q dy$ is a 1-form, then

$$d^1(P dx + Q dy) = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy.$$

**C-2.24.** Prove that if $L$ and $L'$ are nonabelian two-dimensional Lie algebras, then $L \cong L'$.

**C-2.25.** (i) Prove that the ***center*** of a Lie algebra $L$, defined by

$$Z(L) = \{a \in L : [a, x] = 0 \text{ for all } x \in L\},$$

is an abelian ideal in $L$.

(ii) Give an example of a Lie algebra $L$ for which $Z(L) = \{0\}$.

(iii) If $L$ is nilpotent and $L \neq \{0\}$, prove that $Z(L) \neq \{0\}$.

**C-2.26.** Prove that if $L$ is an $n$-dimensional Lie algebra, then $Z(L)$ cannot have dimension $n - 1$. (Compare Exercise A-4.79 on page 172 in Part 1.)

**C-2.27.** Equip $\mathbb{C}^3$ with a cross product (using the same formula as the cross product on $\mathbb{R}^3$). Prove that

$$\mathbb{C}^3 \cong \mathfrak{sl}(2, \mathbb{C}).$$

## C-2.7. Characters

*Characters* will enable us to use the preceding results about group algebras to produce numerical invariants whose arithmetic properties help to prove theorems about finite groups. The first important instance of this technique is the following theorem.

**Theorem C-2.59 (Burnside).** *Every group of order $p^m q^n$, where $p$ and $q$ are primes, is a solvable group.*

Notice that Burnside's Theorem cannot be improved to groups having orders with only three distinct prime factors, for $A_5$ is a simple group of order $60 = 2^2 \cdot 3 \cdot 5$.

**Proposition C-2.60.** *If Burnside's Theorem is false, then there exists a simple group $G$ of order $p^m q^n$ for primes $p$ and $q$ which has a conjugacy class of $G$ whose size is a power of $p > 1$.*

**Proof.** Assume that Burnside's Theorem is false, and let $G$ be a "least criminal"; that is, $G$ is a nonsolvable group of smallest order $p^m q^n$. If $G$ has a proper normal subgroup $H$ with $H \neq \{1\}$, then both $H$ and $G/H$ are solvable, for their orders are smaller than $|G|$ and are of the form $p^i q^j$. By Proposition A-5.25 in Part 1, $G$ is solvable, and this is a contradiction. We may assume, therefore, that $G$ is a nonabelian simple group.

Let $Q$ be a Sylow $q$-subgroup of $G$. If $Q = \{1\}$, then $G$ is a $p$-group, contradicting $G$ being a nonabelian simple group; hence, $Q \neq \{1\}$. Since the center of $Q$ is nontrivial, by Theorem C-1.22, there exists an element $x \neq 1$ with $x \in Z(Q)$. Now $Q \subseteq C_G(x)$, for every element in $Q$ commutes with $x$, and so

$$[G : Q] = [G : C_G(x)][C_G(x) : Q];$$

that is, $[G : C_G(x)]$ is a divisor of $[G : Q] = p^m$. Of course, $[G : C_G(x)]$ is the number of elements in the conjugacy class $x^G$ of $x$ (Corollary C-1.18). •

We will use characters to prove Theorem C-2.106, which says no such simple group exists.

We now specialize the definition of $k$-representation from arbitrary fields $k$ to the complex numbers $\mathbb{C}$; we abbreviate "$\mathbb{C}$-representation" as "representation".

**Definition.** A ***representation*** of a group $G$ is a homomorphism

$$\sigma\colon G \to \mathrm{GL}(V),$$

where $V$ is a vector space over $\mathbb{C}$. The ***degree*** of $\sigma$ is $\dim_{\mathbb{C}}(V)$.

***For the rest of this section, groups $G$ are finite and representations $\sigma\colon G \to \mathrm{GL}(V)$ have $V$ a finite-dimensional vector space over $\mathbb{C}$.***

We may view a representation $\sigma\colon G \to \mathrm{GL}(V)$ either as a left $\mathbb{C}G$-module $V^\sigma$ (its ***corresponding module***) or as a matrix representation. The scalar multiplication on $V^\sigma$ is given, for each $g \in G$ and $v \in V$, by

$$gv = \sigma(g)(v).$$

One of the most important representations is the *regular representation*.

**Definition.** If $G$ is a group, then the representation $\rho\colon G \to \mathrm{GL}(\mathbb{C}G)$ defined, for all $g, h \in G$, by

$$\rho(g)\colon h \mapsto gh$$

is called the ***regular representation***.

The module corresponding to the regular representation is $\mathbb{C}G$ considered as a left module over itself, for the original scalar multiplication on $\mathbb{C}G$ coincides with the scalar multiplication given by $\rho$; that is, $\rho(g)(v)$ is just the product $gv$ in $\mathbb{C}G$.

**Example C-2.61.** A choice of basis of $V$ allows each $\sigma(g)$ to be regarded as an $n \times n$ nonsingular complex matrix $A(g) = [a_{ij}(g)]$. We remind the reader of a remark made on page 136: $G$-sets give representations. If $X = v_1, \ldots, v_n$ is a basis of a complex vector space $V$ and $\tau \in S_X$, then there is a unique nonsingular linear transformation $T\colon V \to V$ with $T(v_j) = v_{\tau(j)}$ for all $j$, and $\tau \mapsto T$ is an injective

homomorphism $S_X \to \mathrm{GL}(V)$. In more detail, the matrix of $T$ is a ***permutation matrix***: it arises by permuting the columns of the identity matrix $I$ by $\tau$; thus, it has exactly one entry equal to 1 in each row and column while all its other entries are 0. ◀

Two representations $\sigma\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$ can be added.

**Definition.** If $\sigma\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$ are representations, then their ***sum*** $\sigma + \tau\colon G \to \mathrm{GL}(V \oplus W)$ is defined by

$$(\sigma + \tau)(g)\colon (v, w) \mapsto (\sigma(g)v, \tau(g)w)$$

for all $g \in G$, $v \in V$, and $w \in W$.

In matrix terms, if $\sigma\colon G \to \mathrm{GL}(n, \mathbb{C})$ and $\tau\colon G \to \mathrm{GL}(m, \mathbb{C})$, then

$$\sigma + \tau\colon G \to \mathrm{GL}(n + m, \mathbb{C}),$$

and if $g \in G$, then $(\sigma + \tau)(g)$ is the direct sum of blocks $\sigma(g) \oplus \tau(g)$; that is,

$$(\sigma + \tau)(g) = \begin{bmatrix} \sigma(g) & 0 \\ 0 & \tau(g) \end{bmatrix}.$$

The following terminology is the common one used in group representations.

**Definition.** A representation $\sigma$ of a group $G$ is ***irreducible*** if the corresponding $\mathbb{C}G$-module is simple. A representation $\sigma$ is ***completely reducible*** if $\sigma = \tau_1 + \cdots + \tau_m$, where each $\tau_i$ is irreducible.[9] A representation $\sigma$ is ***linear*** if $\mathrm{degree}(\sigma) = 1$.

**Example C-2.62.**

(i) The trivial representation of any group $G$ is linear, for the principal module $V_0(\mathbb{C})$ is one-dimensional. If $G = S_n$, then $\mathrm{sgn}\colon G \to \{\pm 1\}$ is also a linear representation.

(ii) Every linear representation is irreducible, for the corresponding $\mathbb{C}G$-module must be simple; after all, every submodule is a subspace, and $\{0\}$ and $V$ are the only subspaces of a one-dimensional vector space $V$. It follows that the trivial representation of any group $G$ is irreducible, as is the representation sgn of $S_n$. ◀

Recall the proof of the Wedderburn–Artin Theorem: there are pairwise non-isomorphic minimal left ideals $L_1, \ldots, L_r$ in $\mathbb{C}G$ and $\mathbb{C}G = B_1 \oplus \cdots \oplus B_r$, where $B_i$ is generated by all minimal left ideals isomorphic to $L_i$. By Corollary C-2.45, $B_i \cong \mathrm{Mat}_{n_i}(\mathbb{C})$, where $n_i = \dim_{\mathbb{C}}(L_i)$. But all minimal left ideals in $\mathrm{Mat}_{n_i}(\mathbb{C})$ are isomorphic, by Corollary C-2.41, so that $L_i \cong \mathrm{COL}(1) \cong \mathbb{C}^{n_i}$ (see Example C-2.6(iii)). Therefore,

$$B_i \cong \mathrm{End}_{\mathbb{C}}(L_i).$$

[9] Since *representation* here means $\mathbb{C}$-representation, Maschke's Theorem (with Molien's assistance) says that every representation is completely reducible.

**Proposition C-2.63.**

(i) *For each minimal left ideal $L_i$ in $\mathbb{C}G$, there is an irreducible representation $\lambda_i\colon G \to \mathrm{GL}(L_i)$, given by left multiplication*:

$$\lambda_i(g)\colon u_i \mapsto gu_i,$$

*where $g \in G$ and $u_i \in L_i$; moreover,* $\mathrm{degree}(\lambda_i) = n_i = \dim_{\mathbb{C}}(L_i)$.

(ii) *The representation $\lambda_i$ extends to a $\mathbb{C}$-algebra map $\widetilde{\lambda}_i\colon \mathbb{C}G \to \mathrm{End}_{\mathbb{C}}(L_i)$ if we define*

$$\widetilde{\lambda}_i(g)u_j = \begin{cases} gu_i & \text{if } j = i, \\ 0 & \text{if } j \neq i \end{cases}$$

*for $g \in G$ and $u_j \in B_j$.*

**Proof.**

(i) Since $L_i$ is a left ideal in $\mathbb{C}G$, each $g \in G$ acts on $L_i$ by left multiplication, and so the representation $\lambda_i$ of $G$ is as stated. Thus, $\lambda_i$ is the restriction of the regular representation, so that the corresponding module $L_i^{\sigma}$ is $L_i$, and $\lambda_i$ is an irreducible representation (because $L_i$, being a minimal left ideal, is a simple module).

(ii) If we regard $\mathbb{C}G$ and $\mathrm{End}_{\mathbb{C}}(L_i)$ as vector spaces over $\mathbb{C}$, then $\lambda_i$ extends to a linear transformation $\widetilde{\lambda}_i\colon \mathbb{C}G \to \mathrm{End}_{\mathbb{C}}(L_i)$ (because the elements of $G$ are a basis of $\mathbb{C}G$):

$$\widetilde{\lambda}_i\colon \sum_{g\in G} c_g g \mapsto \sum_{g\in G} c_g \lambda_i(g);$$

remember that $\lambda_i(g) \in \mathrm{GL}(L_i) \subseteq \mathrm{End}_{\mathbb{C}}(L_i)$. Let us show that $\widetilde{\lambda}_i\colon \mathbb{C}G \to \mathrm{End}_{\mathbb{C}}(L_i)$ is an algebra map. Now $\mathbb{C}G = B_1 \oplus \cdots \oplus B_r$, where the $B_j$ are two-sided ideals. To prove that $\widetilde{\lambda}_i$ is multiplicative, it suffices to check its values on products of basis elements. If $u_j \in B_j$ and $g, h \in G$, then

$$\widetilde{\lambda}_i(gh)\colon u_j \mapsto (gh)u_j,$$

while

$$\widetilde{\lambda}_i(g)\widetilde{\lambda}_i(h)\colon u_j \mapsto hu_j \mapsto g(hu_j);$$

these are the same, by associativity. Thus,

$$\widetilde{\lambda}_i(gh) = \widetilde{\lambda}_i(g)\widetilde{\lambda}_i.$$

Finally, $\widetilde{\lambda}_i(1) = \lambda_i(1) = 1_{L_i}$, and so $\widetilde{\lambda}_i$ is an algebra map. •

It is natural to call two representations *equivalent* if their corresponding modules are isomorphic. The following definition arises from Corollary C-2.16, which gives a criterion that $\mathbb{C}G$-modules $(\mathbb{C}^n)^{\sigma}$ and $(\mathbb{C}^n)^{\tau}$ are isomorphic as $\mathbb{C}G$-modules.

**Definition.** Let $\sigma, \tau\colon G \to \mathrm{GL}(n, \mathbb{C})$ be representations of a group $G$. Then $\sigma$ and $\tau$ are ***equivalent***, denoted by $\sigma \sim \tau$, if there is a nonsingular $n \times n$ matrix $P$ that intertwines them; that is, for every $g \in G$,

$$P\sigma(g)P^{-1} = \tau(g).$$

**Corollary C-2.64.**

(i) *Every irreducible representation of a finite group $G$ is equivalent to one of the representations $\lambda_i$ given in* Proposition C-2.63(i).

(ii) *If $\sigma\colon G \to \mathrm{GL}(V)$ is a matrix representation of a finite group $G$, then $\sigma(g)$ is similar to a diagonal matrix for each $g \in G$.*

**Proof.**

(i) If $\sigma\colon G \to \mathrm{GL}(V)$ is an irreducible representation $\sigma$, then the corresponding $\mathbb{C}G$-module $V^\sigma$ is a simple module. Therefore, $V^\sigma \cong L_i$, for some $i$, by Theorem C-2.33. But $L_i \cong V^{\lambda_i}$, so that $V^\sigma \cong V^{\lambda_i}$ and $\sigma \sim \lambda_i$.

(ii) If $\tau = \sigma|\langle g\rangle$, then $\tau(g) = \sigma(g)$. Now $\tau$ is a representation of the abelian group $\langle g\rangle$, and so part (i) implies that the module $V^\tau$ is a direct sum of one-dimensional submodules. If $V^\tau = \langle v_1\rangle \oplus \cdots \oplus \langle v_m\rangle$, then the matrix of $\sigma(g)$ with respect to the basis $v_1, \ldots, v_m$ is diagonal. •

**Example C-2.65.**

(i) By Theorem C-2.50, every irreducible representation of a finite abelian group is linear.

(ii) The Wedderburn–Artin Theorems can be restated to say that every representation $\tau\colon G \to \mathrm{GL}(V)$ is completely reducible: $\tau = \sigma_1 + \cdots + \sigma_k$, where each $\sigma_j$ is irreducible; moreover, the multiplicity of each $\sigma_j$ is uniquely determined by $\tau$. Since each $\sigma_j$ is equivalent to the irreducible representation $\lambda_i$ arising from a minimal left ideal $L_i$, we usually collect terms and write $\tau \sim \sum_i m_i\lambda_i$, where the multiplicities $m_i$ are nonnegative integers.

(iii) The regular representation $\rho\colon G \to \mathrm{GL}(\mathbb{C}G)$ is important because every irreducible representation is a summand of it. Now $\rho$ is equivalent to the sum

$$\rho \sim n_1\lambda_1 + \cdots + n_r\lambda_r,$$

where $n_i$ is the degree of $\lambda_i$ (recall that $\mathbb{C}G = \bigoplus_i B_i$, where $B_i \cong \mathrm{End}_{\mathbb{C}}(L_i) \cong \mathrm{Mat}_{n_i}(\mathbb{C})$; as a $\mathbb{C}G$-module, the simple module $L_i$ can be viewed as the first columns of $n_i \times n_i$ matrices, and so $B_i$ is a direct sum of $n_i$ copies of $L_i$). ◀

Recall that the *trace* of an $n \times n$ matrix $A = [a_{ij}]$ with entries in a commutative ring $k$ is the sum of the diagonal entries: $\mathrm{tr}(A) = \sum_{i=1}^n a_{ii}$. We remind the reader of some elementary facts about the trace.

**Proposition C-2.66.**

(i) *If $I$ is the $n \times n$ identity matrix and $k$ is a field of characteristic* 0, *then* $\mathrm{tr}(I) = n$.

(ii) $\mathrm{tr}(A) = -\sum_{i=1}^n \alpha_i$, *where $\alpha_1, \ldots, \alpha_n$ are the eigenvalues of $A$ (with multiplicities).*

(iii) *If* $A = [a_{ij}]$ *and* $B = [b_{ij}]$ *are* $n \times n$ *matrices with entries in a commutative ring* $k$*, then*

$$\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B) \quad \textit{and} \quad \operatorname{tr}(AB) = \operatorname{tr}(BA).$$

(iv) *If* $B = PAP^{-1}$*, then* $\operatorname{tr}(B) = \operatorname{tr}(A)$.

**Proof.**

(i) The sum of the diagonal entries is $n$, which is not 0 because $k$ has characteristic 0.

(ii) Proposition B-3.54 in Part 1.

(iii) The additivity of trace follows from the diagonal entries of $A + B$ being $a_{ii} + b_{ii}$. If $(AB)_{ii}$ denotes the $ii$ entry of $AB$, then

$$(AB)_{ii} = \sum_j a_{ij} b_{ji},$$

and so

$$\operatorname{tr}(AB) = \sum_i (AB)_{ii} = \sum_{i,j} a_{ij} b_{ji}.$$

Similarly,

$$\operatorname{tr}(BA) = \sum_{j,i} b_{ji} a_{ij}.$$

The entries commute because they lie in the commutative ring $k$, and so $a_{ij} b_{ji} = b_{ji} a_{ij}$ for all $i, j$. It follows that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, as desired.

(iv) Using (ii), we have

$$\operatorname{tr}(B) = \operatorname{tr}\big((PA)P^{-1}\big) = \operatorname{tr}\big(P^{-1}(PA)\big) = \operatorname{tr}(A). \quad \bullet$$

It follows from Proposition C-2.66(iii) that we can define the trace of a linear transformation $T\colon V \to V$, where $V$ is a vector space over a field $k$, as the trace of any matrix arising from it: if $A$ and $B$ are matrices of $T$, determined by two choices of bases of $V$, then $B = PAP^{-1}$ for some nonsingular matrix $P$, and so $\operatorname{tr}(B) = \operatorname{tr}(A)$.

**Definition.** If $\sigma\colon G \to \mathrm{GL}(V)$ is a representation, then its ***character*** is the function $\chi_\sigma\colon G \to \mathbb{C}$ defined by

$$\chi_\sigma(g) = \operatorname{tr}(\sigma(g)).$$

We call $\chi_\sigma$ the character ***afforded*** by $\sigma$. An ***irreducible character*** is a character afforded by an irreducible representation. The ***degree*** of $\chi_\sigma$ is defined to be the degree of $\sigma$; that is,

$$\operatorname{degree}(\chi_\sigma) = \operatorname{degree}(\sigma) = \dim(V).$$

A character of degree 1 is called ***linear***.

If $\sigma\colon G \to \mathrm{GL}(1,k) = k^\times$ is a linear representation, then the character afforded by $\sigma$ is linear.

**Remark.** Where does the definition of character come from? Linear characters $\varphi\colon G \to \mathrm{GL}_1(k)$ are really homomorphisms $\varphi\colon G \to k^\times$, the multiplicative group of nonzero elements of $k$. In particular, when $k = \mathbb{C}$, values are just roots of unity, for these are the only elements of finite order in $\mathbb{C}^\times$. Such functions arose in number theory, in work of Gauss (and earlier), for example. In the 1880s, in his study of discriminants of algebraic number fields, Dedekind saw an analogy with groups. If $G$ is a finite group of order $n$, he considered the $|G| \times |G|$ matrix $A$ over $\mathbb{C}[X]$, where $X = \{x_g : g \in G\}$ is a set of $n$ commuting indeterminates, and $A = [a_{gh}] = [x_{gh^{-1}}]$. For $G$ a cyclic group, this matrix is reminiscent of ***circulants*** (matrices familiar to nineteenth-century mathematicians and whose eigenvalues were explicitly known). Denote $\det(A)$ by $\Theta(G)$; it is a polynomial of total degree $n$, and it was called the ***group determinant***. Dedekind was able to factor $\Theta(G)$ for *abelian* $G$ as a product of linear forms whose coefficients are characters; in fact, if $\widehat{G}$ is the ***character group*** $\mathrm{Hom}(G, \mathbb{C}^\times)$, then

$$\Theta(G) = \prod_{\chi \in \widehat{G}} \Big( \sum_{g \in G} \chi(g) x_g \Big).$$

Dedekind also computed $\Theta(G)$ for certain nonabelian groups $G$ of small order (factors are not necessarily linear). In 1896, not making progress understanding such factorizations, Dedekind wrote two letters to Frobenius. As Lam writes in [**134**], these letters "became the catalyst for the creation of the character theory for abstract nonabeliean groups." ◀

**Example C-2.67.**

(i) Every linear character is irreducible, for every linear representation is simple.

(ii) The representation $\lambda_i\colon G \to \mathrm{GL}(L_i)$, given by $\lambda_i\colon u_i \mapsto g u_i$ if $u_i \in L_i$, is irreducible (see Proposition C-2.63(i)). Thus, the character $\chi_i$ afforded by $\lambda_i$, defined by

$$\chi_i = \chi_{\lambda_i},$$

is irreducible.

(iii) In light of Proposition C-2.63(ii), it makes sense to speak of $\chi_i(u)$ for every $u \in \mathbb{C}G$. If we write $u = u_1 + \cdots + u_r \in B_1 \oplus \cdots \oplus B_r$, where $u_j \in B_j$, define $\chi_i(u) = \widetilde{\lambda}_i(u_i)$. In particular, $\chi_i(u_i) = \mathrm{tr}(\widetilde{\lambda}_i(u_i))$ and $\chi_i(u_j) = 0$ if $j \neq i$.

(iv) If $\sigma\colon G \to \mathrm{GL}(V)$ is a representation, then $\sigma(1)$ is the identity matrix. Hence, Proposition C-2.66(i) gives $\chi_\sigma(1) = n$, where $n$ is the degree of $\sigma$.

(v) Let $\sigma\colon G \to S_X$ be a homomorphism; as in Example C-2.61, we may regard $\sigma$ as a representation on $V$, where $V$ is the vector space over $\mathbb{C}$ with basis $X$. For every $g \in G$, the matrix $\sigma(g)$ is a permutation matrix, and its $x$th diagonal entry is 1 if $\sigma(g)x = x$; otherwise, it is 0. Thus,

$$\chi_\sigma(g) = \mathrm{tr}(\sigma(g)) = \mathrm{Fix}(\sigma(g)),$$

the number of $x \in X$ fixed by $\sigma(g)$. In other words, if $X$ is a $G$-set, then each $g \in G$ acts on $X$, and the number of ***fixed points*** of the action of $g$ is a character value (see Example C-2.87 for a related discussion). ◀

**Proposition C-2.68.** *If $\chi_\sigma$ is the character afforded by a representation $\sigma\colon G \to \mathrm{GL}(V)$, then for each $g \in G$, $\chi_\sigma(g)$ is a sum of roots of unity.*

**Proof.** Since $G$ is finite (our standing assumption), each $g \in G$ has finite order. If $g^n = 1$, then $\sigma(g)^n = I$, so that every eigenvalue of $\sigma(g)$ is an $n$th root of unity. The result now follows from Proposition B-3.54 in Part 1: $\mathrm{tr}(\sigma(g))$ is the sum of the eigenvalues of $\sigma(g)$. •

Characters are compatible with addition of representations. If $\sigma\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$, then $\sigma + \tau\colon G \to \mathrm{GL}(V \oplus W)$, and

$$\mathrm{tr}((\sigma+\tau)(g)) = \mathrm{tr}\left(\begin{bmatrix}\sigma(g) & 0\\ 0 & \tau(g)\end{bmatrix}\right) = \mathrm{tr}(\sigma(g)) + \mathrm{tr}(\tau(g)).$$

Therefore,

$$\chi_{\sigma+\tau} = \chi_\sigma + \chi_\tau.$$

If $\sigma$ and $\tau$ are equivalent representations, then

$$\mathrm{tr}(\sigma(g)) = \mathrm{tr}(P\sigma(g)P^{-1}) = \mathrm{tr}(\tau(g))$$

for all $g \in G$; that is, they have the same characters: $\chi_\sigma = \chi_\tau$. It follows that if $\sigma\colon G \to \mathrm{GL}(V)$ is a representation, then its character $\chi_\sigma$ can be computed relative to any convenient basis of $V$.

**Theorem C-2.69.**

(i) *The only irreducible characters of $G$ are $\chi_1, \ldots, \chi_r$, the characters afforded by the irreducible representations $\lambda_i$.*

(ii) *Every character $\chi_\sigma$ is a linear combination $\chi_\sigma = \sum_i m_i\chi_i$, where $m_i \geq 0$ are nonnegative integers and*

$$\chi_i = \chi_{\lambda_i}$$

*is the irreducible character afforded by the irreducible representation $\lambda_i$ arising from the minimal left ideal $L_i$.*

(iii) *Equivalent representations have the same character.*

**Proof.**

(i) This follows from Corollary C-2.64(i).

(ii) The character $\chi_\sigma$ arises from a representation $\sigma$ of $G$, which, in turn, arises from a $\mathbb{C}G$-module $V$. But $V$ is a semisimple module (because $\mathbb{C}G$ is a semisimple ring), and so $V$ is a direct sum of simple modules: $V = \bigoplus_j S_j$. By (i), each $S_j \cong L_i$ for some $L_i$. If, for each $i$, we let $m_i \geq 0$ be the number of $S_j$ isomorphic to $L_i$, then $\chi_\sigma = \sum_i m_i\lambda_i$.

(iii) This follows from of Proposition C-2.66(ii) and Corollary C-2.64(i). •

As a consequence of Theorem C-2.69, we call $\chi_1, \ldots, \chi_r$ ***the irreducible characters*** of $G$.

**Example C-2.70.**

(i) The (linear) character $\chi_1$ afforded by the trivial representation $\sigma\colon G \to \mathbb{C}$ with $\sigma(g) = 1$ for all $g \in G$ is called the ***trivial character***. Thus, $\chi_1(g) = 1$ for all $g \in G$.

(ii) Let us compute the ***regular character*** $\psi = \chi_\rho$ afforded by the regular representation $\rho\colon G \to \mathrm{GL}(\mathbb{C}G)$, where $\rho(g)\colon u \mapsto gu$ for all $g \in G$ and $u \in \mathbb{C}G$. Any basis of $\mathbb{C}G$ can be used for this computation; we choose the usual basis comprised of the elements of $G$. If $g = 1$, then Example C-2.67(iv) shows that $\psi(1) = \dim(\mathbb{C}G) = |G|$. On the other hand, if $g \neq 1$, then for all $h \in G$, we have $gh$ a basis element distinct from $h$. Therefore, the matrix of $\rho(g)$ has 0's on the diagonal, and so its trace is 0. Thus,

$$\psi(g) = \begin{cases} 0 & \text{if } g \neq 1, \\ |G| & \text{if } g = 1. \end{cases} \quad \blacktriangleleft$$

## Exercises

**C-2.28.** Let $\varphi\colon G \to \mathrm{GL}(V)$ be a representation of a finite group $G$. If $V = W \oplus W'$ and $\operatorname{im}\varphi \subseteq W$, prove that a matrix afforded by $\varphi$ is given by

$$x \mapsto \begin{bmatrix} A(x) & C(x) \\ 0 & B(x) \end{bmatrix}.$$

## C-2.8. Class Functions

We proved, in Theorem C-2.69, that equivalent representations have the same character. The coming discussion will give the converse (Theorem C-2.72): two representations have the same character if and only if they are equivalent.

We paraphrase a remark of Isaacs [**105**], p. 14. Representations contain too much information. If $\sigma, \tau\colon G \to \mathrm{GL}(n, \mathbb{C})$ are matrix representations, then for every $g \in G$, the matrices $\sigma(g)$ and $\tau(g)$ each contain $n^2$ entries, many of which determine whether $\sigma$ and $\tau$ are similar. The trace eliminates much of this unnecessary data; indeed, Theorem C-2.72 shows that it eliminates just the right amount.

**Definition.** A function $\varphi\colon G \to \mathbb{C}$ is a ***class function*** if it is constant on conjugacy classes; that is, if $h = xgx^{-1}$, then $\varphi(h) = \varphi(g)$.

Every character $\chi_\sigma$ afforded by a representation $\sigma$ is a class function: if $h = xgx^{-1}$, then

$$\sigma(h) = \sigma(xgx^{-1}) = \sigma(x)\sigma(g)\sigma(x)^{-1},$$

and so $\mathrm{tr}(\sigma(h)) = \mathrm{tr}(\sigma(g))$; that is,

$$\chi_\sigma(h) = \chi_\sigma(g).$$

The converse is not true; not every class function is a character. For example, if $\chi$ is a character, then $-\chi$ is a class function; it is not a character because $-\chi(1)$ is negative, and so it cannot be a degree of a representation.

**Definition.** We denote the set of all class functions $G \to \mathbb{C}$ by $\mathrm{CF}(G)$:

$$\mathrm{CF}(G) = \{\varphi\colon G \to \mathbb{C} : \varphi(g) = \varphi(xgx^{-1}) \text{ for all } x, g \in G\}.$$

It is easy to see that $\mathrm{CF}(G)$ is a vector space over $\mathbb{C}$.

An element $u = \sum_{g\in G} c_g g \in \mathbb{C}G$ is an $n$-tuple $(c_g)$ of complex numbers; that is, $u$ is a function $u\colon G \to \mathbb{C}$ with $u(g) = c_g$ for all $g \in G$. From this viewpoint, we see that $\mathrm{CF}(G)$ is a subring of $\mathbb{C}G$. Note that a class function is a scalar multiple of a class sum; therefore, Lemma C-2.48 says that $\mathrm{CF}(G)$ is the center $Z(\mathbb{C}G)$, and so

$$\dim(\mathrm{CF}(G)) = r,$$

where $r$ is the number of conjugacy classes in $G$ (Theorem C-2.49).

**Definition.** Write $\mathbb{C}G = B_1 \oplus \cdots \oplus B_r$, where $B_i \cong \mathrm{End}_{\mathbb{C}}(L_i)$, and let $e_i$ denote the identity element of $B_i$; hence,

$$1 = e_1 + \cdots + e_r,$$

where 1 is the identity element of $\mathbb{C}G$. The elements $e_i$ are called the ***idempotents*** in $\mathbb{C}G$.

Not only is each $e_i$ an idempotent, that is, $e_i \neq 0$ and $e_i^2 = e_i$, but it is easy to see that

$$e_i e_j = \delta_{ij} e_i,$$

where $\delta_{ij}$ is the Kronecker delta.

**Lemma C-2.71.** *The irreducible characters $\chi_1, \ldots, \chi_r$ form a basis of* $\mathrm{CF}(G)$.

**Proof.** We have just seen that $\dim(\mathrm{CF}(G)) = r$, and so it suffices to prove that $\chi_1, \ldots, \chi_r$ is a linearly independent list, by Corollary A-7.20 in Part 1. We have already noted that $\chi_i(u_j) = 0$ for all $j \neq i$; in particular, $\chi_i(e_j) = 0$. On the other hand, $\chi_i(e_i) = n_i$, where $n_i$ is the degree of $\chi_i$, for it is the trace of the $n_i \times n_i$ identity matrix.

Suppose now that $\sum_i c_i\chi_i = 0$. It follows, for all $j$, that

$$0 = \Big(\sum_i c_i\chi_i\Big)(e_j) = c_j\chi_j(e_j) = c_j n_j.$$

Therefore, all $c_j = 0$, as desired. •

Since $\chi_i(1)$ is the trace of the $n_i \times n_i$ identity matrix, we have

$$n_i = \chi_i(1) = \sum_j \chi_i(e_j) = \chi_i(e_i), \tag{1}$$

where $e_i$ is the identity element of $B_i$.

**Theorem C-2.72.** *Two representations $\sigma, \tau$ of a finite group $G$ are equivalent if and only if they afford the same character*: $\chi_\sigma = \chi_\tau$.

**Proof.** We have already proved necessity, in Theorem C-2.69(iii). For sufficiency, Theorem C-2.69(ii) says that every representation is completely reducible: there are nonnegative integers $m_i$ and $\ell_i$ with $\sigma \sim \sum_i m_i\lambda_i$ and $\tau \sim \sum_i \ell_i\lambda_i$. By hypothesis, the corresponding characters coincide:

$$\sum_i m_i\chi_i = \chi_\sigma = \chi_\tau = \sum_i \ell_i\chi_i.$$

As the irreducible characters $\chi_1, \ldots, \chi_r$ are a basis of $\mathrm{CF}(G)$, $m_i = \ell_i$ for all $i$, and so $\sigma \sim \tau$. •

There are relations between the irreducible characters that facilitate their calculation. We begin by finding the expression of the idempotents $e_i$ in terms of the basis $G$ of $\mathbb{C}G$. Observe, for all $y \in G$, that

$$\chi_i(e_i y) = \chi_i(y), \tag{2}$$

for $y = \sum_j e_j y$, and so $\chi_i(y) = \sum_j \chi_i(e_j y) = \chi_i(e_i y)$, because $e_j y \in B_j$.

**Proposition C-2.73.** *If $e_i = \sum_{g\in G} a_{ig} g$, where $a_{ig} \in \mathbb{C}$, then*

$$a_{ig} = \frac{n_i\chi_i(g^{-1})}{|G|}.$$

**Proof.** Let $\psi$ be the regular character; that is, $\psi$ is the character afforded by the regular representation. Now $e_i g^{-1} = \sum_h a_{ih} h g^{-1}$, so that

$$\psi(e_i g^{-1}) = \sum_{h\in G} a_{ih}\psi(hg^{-1}).$$

By Example C-2.70(ii), $\psi(1) = |G|$ when $h = g$ and $\psi(hg^{-1}) = 0$ when $h \neq g$. Therefore,

$$a_{ig} = \frac{\psi(e_i g^{-1})}{|G|}.$$

On the other hand, since $\psi = \sum_j n_j\chi_j$, we have

$$\psi(e_i g^{-1}) = \sum_j n_j\chi_j(e_i g^{-1}) = n_i\chi_i(e_i g^{-1}),$$

by Proposition C-2.63(ii). But $\chi_i(e_i g^{-1}) = \chi_i(g^{-1})$, by Eq. (1). Therefore, $a_{ig} = n_i\chi_i(g^{-1})/|G|$. •

It is now convenient to equip $\mathrm{CF}(G)$ with an inner product.

**Definition.** If $\alpha, \beta \in \mathrm{CF}(G)$, define

$$(\alpha, \beta) = \frac{1}{|G|}\sum_{g\in G} \alpha(g)\overline{\beta(g)},$$

where $\overline{c}$ denotes the complex conjugate of a complex number $c$.

It is easy to see that we have defined an inner product[10]; that is, for all $c_1, c_2 \in \mathbb{C}$,

(i) $(c_1\alpha_1 + c_2\alpha_2, \beta) = c_1(\alpha_1, \beta) + c_2(\alpha_2, \beta)$;

(ii) $(\beta, \alpha) = \overline{(\alpha, \beta)}$.

Note that $(\alpha, \alpha)$ is real, by (ii), and the inner product is ***definite***; that is, $(\alpha, \alpha) > 0$ if $\alpha \neq 0$.

**Theorem C-2.74.** *With respect to the inner product just defined, the irreducible characters $\chi_1, \ldots, \chi_r$ form an orthonormal basis; that is,*

$$(\chi_i, \chi_j) = \delta_{ij}.$$

**Proof.** By Proposition C-2.73, we have

$$e_j = \frac{1}{|G|} \sum_g n_j \chi_j(g^{-1}) g.$$

Hence,

$$\chi_i(e_j)/n_j = \frac{1}{|G|} \sum_g \chi_j(g^{-1})\chi_i(g) = \frac{1}{|G|} \sum_g \chi_i(g)\overline{\chi_j(g)} = (\chi_i, \chi_j);$$

the next to last equation follows from Exercise C-2.30 on page 198, for $\chi_j$ is a character (not merely a class function), and so $\chi_j(g^{-1}) = \overline{\chi_j(g)}$.[11] The result now follows, for $\chi_i(e_j)/n_j = \delta_{ij}$, by Eqs. (1) and (2). •

The inner product on $\mathrm{CF}(G)$ can be used to check irreducibility.

**Definition.** A ***generalized character*** $\varphi$ on a finite group $G$ is a $\mathbb{Z}$-linear combination

$$\varphi = \sum_i m_i \chi_i,$$

where $\chi_1, \ldots, \chi_r$ are the irreducible characters of $G$ and all $m_i \in \mathbb{Z}$.

If $\theta$ is a character, then $\theta = \sum_i m_i\chi_i$, where all the coefficients are *nonnegative* integers, by Theorem C-2.69.

**Corollary C-2.75.** *A generalized character $\theta$ of a group $G$ is an irreducible character if and only if $\theta(1) > 0$ and $(\theta, \theta) = 1$.*

**Proof.** If $\theta$ is an irreducible character, then $\theta = \chi_i$ for some $i$, and so $(\theta, \theta) = (\chi_i, \chi_i) = 1$. Moreover, $\theta(1) = \deg(\chi_i) > 0$.

Conversely, let $\theta = \sum_j m_j\chi_j$, where $m_j \in \mathbb{Z}$, and suppose that $(\theta, \theta) = 1$. Then $1 = \sum_j m_j^2$; hence, some $m_i^2 = 1$ and all other $m_j = 0$. Therefore, $\theta = \pm\chi_i$, and so $\theta(1) = \pm\chi_i(1)$. Since $\chi_i(1) = \deg(\chi_i) > 0$, the hypothesis $\theta(1) > 0$ gives $m_i = 1$. Therefore, $\theta = \chi_i$, and so $\theta$ is an irreducible character. •

[10]This is not really an inner product, for it is *not* symmetric: $(\beta, \alpha) = \overline{(\alpha, \beta)}$, not $(\alpha, \beta)$. It is, however, a *hermitian form*. In spite of this, we will continue to call it an inner product.

[11]Recall that if $z$ lies on the unit circle, in particular, if $z$ is a root of unity, then $z^{-1} = \overline{z}$.

## C-2.9. Character Tables and Orthogonality Relations

Let us assemble the notation we will use from now on.

**Notation.** If $G$ is a finite group, we denote its conjugacy classes by

$$C_1,\ \ldots,\ C_r,$$

a choice of elements, one from each conjugacy class, by

$$g_1 \in C_1,\ \ldots,\ g_r \in C_r,$$

its irreducible characters by

$$\chi_1,\ \cdots,\ \chi_r,$$

their degrees by

$$n_1 = \chi_1(1),\ \ldots,\ n_r = \chi_r(1),$$

and the sizes of the conjugacy classes by

$$h_1 = |C_1|,\ \ldots,\ h_r = |C_r|.$$

The matrix $[\chi_i(g_j)]$ is a useful way to display information.

**Definition.** The ***character table*** of $G$ is the $r \times r$ complex matrix whose $ij$ entry is $\chi_i(g_j)$.

We always assume that $C_1 = \{1\}$ and that $\chi_1$ is the trivial character. Thus, the first row consists of all 1's, while the first column consists of the degrees of the characters: $\chi_i(1) = n_i$ for all $i$, by Example C-2.67(iv). The $i$th row of the character table consists of the values

$$\chi_i(1), \chi_i(g_2), \ldots, \chi_i(g_r).$$

There is no obvious way of labeling the other conjugacy classes (or the other irreducible characters), so that a finite group $G$ has many character tables. Nevertheless, we usually speak of "the" character table of $G$.

Since the inner product on $\mathrm{CF}(G)$ is summed over all $g \in G$, not just the chosen $g_i$ (one from each conjugacy class), it can be rewritten as a "weighted" inner product:

$$(\chi_i, \chi_j) = \frac{1}{|G|} \sum_{k=1}^{r} h_k \chi_i(g_k) \overline{\chi_j(g_k)}.$$

Theorem C-2.74 says that the weighted inner product of distinct rows in the character table is 0, while the weighted inner product of any row with itself is 1.

**Example C-2.76.**

(i) A character table can have complex entries. For example, it is easy to see that the character table for a cyclic group $G = \langle x \rangle$ of order 3 is given in Table 1, where $\omega = e^{2\pi i/3}$ is a primitive cube root of unity.

(ii) Write the four-group in additive notation:

$$\mathbf{V} = \{0,\ a,\ b,\ a+b\}.$$

As a vector space over $\mathbb{F}_2$, $\mathbf{V}$ has basis $a, b$, and the "coordinate functions" on $\mathbf{V}$, which take values in $\{1, -1\} \subseteq \mathbb{C}$, are linear; hence, they are irreducible

| $g_i$ | 1 | $x$ | $x^2$ |
|---|---|---|---|
| $h_i$ | 1 | 1 | 1 |
| $\chi_1$ | 1 | 1 | 1 |
| $\chi_2$ | 1 | $\omega$ | $\omega^2$ |
| $\chi_3$ | 1 | $\omega^2$ | $\omega$ |

**Table 1.** Character table of $\mathbb{Z}_3$.

representations. For example, the character $\chi_2$ arising from the function that is nontrivial on $a$ and trivial on $b$ is

$$\chi_2(v) = \begin{cases} -1 & \text{if } v = a \quad \text{or} \quad v = a + b, \\ 1 & \text{if } v = 0 \quad \text{or} \quad v = b. \end{cases}$$

Table 2 is the character table for $\mathbf{V}$.

| $g_i$ | 0 | $a$ | $b$ | $a+b$ |
|---|---|---|---|---|
| $h_i$ | 1 | 1 | 1 | 1 |
| $\chi_1$ | 1 | 1 | 1 | 1 |
| $\chi_2$ | 1 | $-1$ | 1 | $-1$ |
| $\chi_3$ | 1 | 1 | $-1$ | $-1$ |
| $\chi_4$ | 1 | $-1$ | $-1$ | 1 |

**Table 2.** Character table of $\mathbf{V}$.

(iii) We now discuss Table 3, the character table for the symmetric group $G = S_3$. Since two permutations in $S_n$ are conjugate if and only if they have the same cycle structure, there are three conjugacy classes, and we choose elements 1, (1 2), and (1 2 3) from each. (In Example C-2.52(i), we saw that there are three irreducible representations: $\lambda_1$ = the trivial representation, $\lambda_2$ = sgn, and a third representation $\lambda_3$ of degree 2.) Since $\chi_2$ = sgn, the second row records the fact that (1) and (1 2 3) are even while (1 2) is odd. The third row has entries

$$2 \quad a \quad b,$$

| $g_i$ | 1 | (1 2) | (1 2 3) |
|---|---|---|---|
| $h_i$ | 1 | 3 | 2 |
| $\chi_1$ | 1 | 1 | 1 |
| $\chi_2$ | 1 | $-1$ | 1 |
| $\chi_3$ | 2 | 0 | $-1$ |

**Table 3.** Character table of $S_3$.

where $a$ and $b$ are to be found. The weighted inner products of row 3 with the other two rows give the equations

$$2 + 3a + 2b = 0,$$
$$2 - 3a + 2b = 0.$$

It follows easily that $a = 0$ and $b = -1$. ◀

The following lemma will be used to describe the inner products of the columns of the character table.

**Lemma C-2.77.** *If $A$ is the character table of a finite group $G$, then $A$ is nonsingular and its inverse $A^{-1}$ has $ij$ entry*

$$(A^{-1})_{ij} = \frac{h_i\overline{\chi_j(g_i)}}{|G|}.$$

**Proof.** If $B$ is the matrix whose $ij$ entry is displayed in the statement, then

$$(AB)_{ij} = \frac{1}{|G|}\sum_k \chi_i(g_k)h_k\overline{\chi_j(g_k)} = \frac{1}{|G|}\sum_g \chi_i(g)\overline{\chi_j(g)} = (\chi_i, \chi_j) = \delta_{ij},$$

because $h_k\overline{\chi_j(g_k)} = \sum_{y\in C_k}\overline{\chi_j(y)}$. Therefore, $AB = I$. •

The next result is fundamental.

**Theorem C-2.78 (Orthogonality Relations).** *Let $G$ be a finite group of order $n$ with conjugacy classes $C_1, \ldots, C_r$ of cardinalities $h_1, \ldots, h_r$, respectively, and choose elements $g_i \in C_i$. Let the irreducible characters of $G$ be $\chi_1, \ldots, \chi_r$, and let $\chi_i$ have degree $n_i$. Then the following relations hold*:

(i)
$$\sum_{k=1}^{r} h_k\chi_i(g_k)\overline{\chi_j(g_k)} = \begin{cases} 0 & \text{if } i \neq j, \\ |G| & \text{if } i = j. \end{cases}$$

(ii)
$$\sum_{i=1}^{r} \chi_i(g_k)\overline{\chi_i(g_\ell)} = \begin{cases} 0 & \text{if } k \neq \ell, \\ |G|/h_k & \text{if } k = \ell. \end{cases}$$

**Proof.**

(i) This is just a restatement of Theorem C-2.74.

(ii) If $A$ is the character table of $G$ and $B = [h_i\overline{\chi_j(g_i)}/|G|]$, we proved, in Lemma C-2.77, that $AB = I$. It follows that $BA = I$; that is, $(BA)_{k\ell} = \delta_{k\ell}$. Therefore,
$$\frac{1}{|G|}\sum_i h_k\overline{\chi_i(g_k)}\chi_i(g_\ell) = \delta_{k\ell}.$$ •

In terms of the character table, the second orthogonality relation says that the usual (unweighted, but with complex conjugation) inner product of distinct columns is 0 while, for every $k$, the usual inner product of column $k$ with itself is $|G|/h_k$.

The orthogonality relations yield the following special cases.

**Corollary C-2.79.**

(i) $|G| = \sum_{i=1}^{r} n_i^2$.
(ii) $\sum_{i=1}^{r} n_i \chi_i(g_k) = 0 \quad \textit{if } k > 1$.
(iii) $\sum_{k=1}^{r} h_k \chi_i(g_k) = 0 \quad \textit{if } i > 1$.
(iv) $\sum_{k=1}^{r} h_k |\chi_i(g_k)|^2 = |G|$.

**Proof.**

(i) This equation records the inner product of column 1 with itself: it is Theorem C-2.78(ii) when $k = \ell = 1$.
(ii) This is the special case of Theorem C-2.78(ii) with $\ell = 1$, for $\chi_i(1) = n_i$.
(iii) This is the special case of Theorem C-2.78(i) in which $j = 1$.
(iv) This is the special case of Theorem C-2.78(i) in which $j = i$. •

We can now give another proof of Burnside's Lemma, Theorem C-1.28, which counts the number of orbits of a $G$-set.

**Theorem C-2.80 (Burnside's Lemma Again).** *Let $G$ be a finite group and let $X$ be a finite $G$-set. If $N$ is the number of orbits of $X$, then*

$$N = \frac{1}{|G|} \sum_{g \in G} \mathrm{Fix}(g),$$

*where* $\mathrm{Fix}(g)$ *is the number of $x \in X$ with $gx = x$.*

**Proof.** Let $V$ be the complex vector space having $X$ as a basis. As in Example C-2.61, the $G$-set $X$ gives a representation $\sigma\colon G \to \mathrm{GL}(V)$ by $\sigma(g)(x) = gx$ for all $g \in G$ and $x \in X$; moreover, if $\chi_\sigma$ is the character afforded by $\sigma$, then Example C-2.67(v) shows that $\chi_\sigma(g) = \mathrm{Fix}(g)$.

Let $\mathcal{O}_1, \ldots, \mathcal{O}_N$ be the orbits of $X$. We begin by showing that $N = \dim(V^G)$, where $V^G$ is the space of *fixed points*:

$$V^G = \{v \in V : gv = v \text{ for all } g \in G\}.$$

For each $i$, define $s_i$ to be the sum of all the $x$ in $\mathcal{O}_i$; it suffices to prove that these elements form a basis of $V^G$. It is plain that $s_1, \ldots, s_N$ is a linearly independent list in $V^G$, and it remains to prove that they span $V^G$. If $u \in V^G$, then $u = \sum_{x \in X} c_x x$, so that $gu = \sum_{x \in X} c_x (gx)$. Since $gu = u$, however, $c_x = c_{gx}$. Thus, given $x \in X$ with $x \in \mathcal{O}_j$, each coefficient of $gx$, where $g \in G$, is equal to $c_x$; that is, all the $x$ lying in the orbit $\mathcal{O}_j$ have the same coefficient, say, $c_j$, and so $u = \sum_j c_j s_j$. Therefore,

$$N = \dim(V^G).$$

Now define a linear transformation $T\colon V \to V$ by

$$T = \frac{1}{|G|} \sum_{g \in G} \sigma(g).$$

It is routine to check that $T$ is a $\mathbb{C}G$-map, that $T|(V^G) =$ identity, and that $\operatorname{im} T = V^G$. Since $\mathbb{C}G$ is semisimple, $V = V^G \oplus W$ for some submodule $W$. We claim that $T|W = 0$. If $w \in W$, then $\sigma(g)(w) \in W$ for all $g \in G$, because $W$ is a submodule, and so $T(w) \in W$. On the other hand, $T(w) \in \operatorname{im} T = V^G$, and so $T(w) \in V^G \cap W = \{0\}$, as claimed.

If $w_1, \ldots, w_\ell$ is a basis of $W$, then $s_1, \ldots, s_N, w_1, \ldots, w_\ell$ is a basis of $V = V^G \oplus W$. Note that $T$ fixes each $s_i$ and annihilates each $w_j$. Since trace preserves sums,

$$\operatorname{tr}(T) = \frac{1}{|G|} \sum_{g \in G} \operatorname{tr}(\sigma(g)) = \frac{1}{|G|} \sum_{g \in G} \chi_\sigma(g) = \frac{1}{|G|} \sum_{g \in G} \operatorname{Fix}(g).$$

It follows that

$$\operatorname{tr}(T) = \dim(V^G),$$

for the matrix of $T$ with respect to the chosen basis is the direct sum of an identity block and a zero block, and so $\operatorname{tr}(T)$ is the size of the identity block, namely, $\dim(V^G) = N$. Therefore,

$$N = \frac{1}{|G|} \sum_{g \in G} \operatorname{Fix}(g). \quad \bullet$$

Character tables can be used to detect normal subgroups (after all, normal subgroups are unions of conjugacy classes).

**Definition.** If $\chi_\tau$ is the character afforded by a representation $\tau\colon G \to \mathrm{GL}(V)$, then

$$\ker \chi_\tau = \ker \tau.$$

**Proposition C-2.81.** *Let $\theta = \chi_\tau$ be the character of a finite group $G$ afforded by a representation $\tau\colon G \to \mathrm{GL}(V)$.*

(i) *For each $g \in G$, we have*

$$|\theta(g)| \le \theta(1).$$

(ii)

$$\ker \theta = \{g \in G : \theta(g) = \theta(1)\}.$$

(iii) *If $\theta = \sum_j m_j \chi_j$, where the $m_j$ are positive integers, then*

$$\ker \theta = \bigcap_j \ker \chi_j.$$

(iv) *If $N$ is a normal subgroup of $G$, there are irreducible characters $\chi_{i_1}, \ldots, \chi_{i_s}$ with $N = \bigcap_{j=1}^{s} \ker \chi_{i_j}$.*

**Proof.**

(i) By Lagrange's Theorem, $g^{|G|} = 1$ for every $g \in G$; it follows that the eigenvalues $\varepsilon_1, \ldots, \varepsilon_d$ of $\tau(g)$, where $d = \theta(1)$, are $|G|$th roots of unity, and so $|\varepsilon_j| = 1$ for all $j$. By the triangle inequality in $\mathbb{C}$,

$$\big|\theta(g)\big| = \Big|\sum_{j=1}^{d} \varepsilon_j\Big| \le d = \theta(1).$$

(ii) If $g \in \ker\theta = \ker\tau$, then $\tau(g) = I$, the identity matrix, and $\theta(g) = \mathrm{tr}(I) = \theta(1)$. Conversely, suppose that $\theta(g) = \theta(1) = d$; that is, $\sum_{j=1}^{d} \varepsilon_j = d$. By Proposition A-3.109 in Part 1, all the eigenvalues $\varepsilon_j$ are equal, say, $\varepsilon_j = \omega$ for all $j$. Therefore, $\tau(g) = \omega I$, by Corollary C-2.64(ii), and so

$$\theta(g) = \theta(1)\omega.$$

But $\theta(g) = \theta(1)$, by hypothesis, and so $\omega = 1$; that is, $\tau(g) = I$ and $g \in \ker\tau$.

(iii) For all $g \in G$, we have

$$\theta(g) = \sum_j m_j \chi_j(g);$$

in particular,

$$\theta(1) = \sum_j m_j \chi_j(1).$$

By (ii), if $g \in \ker\theta$, then $\theta(g) = \theta(1)$. Suppose that $\chi_{j'}(g) \neq \chi_{j'}(1)$ for some $j'$. Since $\chi_{j'}(g)$ is a sum of roots of unity (by Proposition C-2.68), Proposition A-3.109 in Part 1 applies to force $|\chi_{j'}(g)| < \chi_{j'}(1)$, and so $|\theta(g)| \leq \sum_j m_j|\chi_j(g)| < \sum_j \chi_j(1) = 1$, which implies that $\theta(g) \neq \theta(1)$, a contradiction. Therefore, $g \in \bigcap_j \ker\chi_j$. For the reverse inclusion, if $g \in \ker\chi_j$, then $\chi_j(g) = \chi_j(1)$ for some $j$, and so

$$\theta(g) = \sum_j m_j \chi_j(g) = \sum_j m_j \chi_j(1) = \theta(1);$$

hence, $g \in \ker\theta$, by (ii).

(iv) It suffices to find a representation of $G$ whose kernel is $N$. By (ii) and Example C-2.70(ii), the regular representation $\rho$ of $G/N$ is faithful (i.e., is an injection), and so its kernel is $\{1\}$. If $\pi\colon G \to G/N$ is the natural map, then $\rho\pi$ is a representation of $G$ having kernel $N$. If $\theta$ is the character afforded by $\rho\pi$, then $\theta = \sum_j m_j\chi_j$, where the $m_j$ are positive integers, by Lemma C-2.71, and so (iii) applies. •

**Example C-2.82.**

(i) Table 6 on page 193 is the character table of $S_4$. We can see there that $\ker\chi_2 = A_4$ and $\ker\chi_3 = \mathbf{V}$ are the only two normal subgroups of $S_4$ (other than $\{1\}$ and $S_4$).

(ii) In Example C-2.83, we can see that $\ker\chi_2 = \{1\} \cup z^G \cup y^G$ (where $z^G$ denotes the conjugacy class of $z$ in $G$) and $\ker\chi_3 = \{1\} \cup z^G \cup x^G$. Another normal subgroup occurs as $\ker\chi_2 \cap \ker\chi_3 = \{1\} \cup z^G$.

(iii) A normal subgroup described by characters is given as a union of conjugacy classes; this viewpoint can give another proof of the simplicity of $A_5$. In Exercise C-1.12 on page 15, we saw that $A_5$ has five conjugacy classes, of sizes 1, 12, 12, 15, and 20. Since every subgroup contains the identity element, the order of a normal subgroup of $A_5$ is the sum of some of these numbers, including 1. But it is easy to see that 1 and 60 are the only such sums that are divisors of 60, and so the only normal subgroups are $\{1\}$ and $A_5$ itself.

◀

## C-2.10. Induced Characters

The orthogonality relations help to complete a character table but, obviously, it would be useful to have a supply of characters. One important class of characters consists of those arising by *lifting* a representation of a quotient group. Another, *induced representations*, arises from representations of a subgroup $H$ of $G$.

**Definition.** Let $H \lhd G$ and let $\sigma\colon G/H \to \mathrm{GL}(V)$ be a representation. If $\pi\colon G \to G/H$ is the natural map, then the representation $\sigma\pi\colon G \to \mathrm{GL}(V)$ is called a ***lifting*** of $\sigma$.

Scalar multiplication of $G$ on a $\mathbb{C}(G/H)$-module $V$ is given, for $v \in V$, by

$$gv = (gH)v.$$

Thus, every $\mathbb{C}(G/H)$-submodule of $V$ is also a $\mathbb{C}G$-submodule; hence, if $V$ is a simple $\mathbb{C}(G/H)$-module, then it is also a simple $\mathbb{C}G$-module. It follows that if $\sigma\colon G/H \to \mathrm{GL}(V)$ is an irreducible representation of $G/H$, then its lifting $\sigma\pi$ is also an irreducible representation of $G$.

**Example C-2.83.** We know that $D_8$ and $\mathbf{Q}$ are nonisomorphic nonabelian groups of order 8; we now show that they have the same character tables.

If $G$ is a nonabelian group of order 8, then its center has order 2, say, $Z(G) = \langle z \rangle$. Now $G/Z(G)$ is not cyclic, by Exercise A-4.79 on page 172 in Part 1, and so $G/Z(G) \cong \mathbf{V}$. Therefore, if $\sigma\colon \mathbf{V} \to \mathbb{C}$ is an irreducible representation of $\mathbf{V}$, then its lifting $\sigma\pi$ is an irreducible representation of $G$. This gives four (necessarily irreducible) linear characters of $G$, each of which takes value 1 on $z$. As $G$ is not abelian, there must be an irreducible character $\chi_5$ of degree $n_5 > 1$ (if all $n_i = 1$, then $\mathbb{C}G$ is commutative and $G$ is abelian). Since $\sum_i n_i^2 = 8$, we see that $n_5 = 2$. Thus, there are five irreducible representations and, hence, five conjugacy classes; choose representatives $g_i$ to be $1, z, x, y, w$. Table 4 is the character table. The values for $\chi_5$ are computed from the orthogonality relations of the columns. For example, if the last row of the character table is

$$2 \quad a \quad b \quad c \quad d,$$

then the inner product of columns 1 and 2 gives the equation $4 + 2a = 0$, so that $a = -2$. The reader may verify that $0 = b = c = d$.

| $g_i$ | 1 | $z$ | $x$ | $y$ | $w$ |
|---|---|---|---|---|---|
| $h_i$ | 1 | 1 | 2 | 2 | 2 |
| $\chi_1$ | 1 | 1 | 1 | 1 | 1 |
| $\chi_2$ | 1 | 1 | $-1$ | 1 | $-1$ |
| $\chi_3$ | 1 | 1 | 1 | $-1$ | $-1$ |
| $\chi_4$ | 1 | 1 | $-1$ | $-1$ | 1 |
| $\chi_5$ | 2 | $-2$ | 0 | 0 | 0 |

**Table 4.** Character table of $D_8$ and of $\mathbf{Q}$.

◀

It is more difficult to find representations of a group arising from representations of its subgroups. The original construction of *induced representations*, due to Frobenius, is rather complicated. Tensor products make this construction more natural. The ring $\mathbb{C}G$ is a $(\mathbb{C}G, \mathbb{C}H)$-bimodule (for $\mathbb{C}H$ is a subring of $\mathbb{C}G$), so that if $V$ is a left $\mathbb{C}H$-module, then the tensor product $\mathbb{C}G \otimes_{\mathbb{C}H} V$ is defined; Proposition B-4.82 in Part 1 says that this tensor product is, in fact, a left $\mathbb{C}G$-module.

**Definition.** Let $H$ be a subgroup of a group $G$. If $\rho\colon H \to \mathrm{GL}(V)$ is a representation with left $\mathbb{C}H$-module $V$, then the ***induced module*** is the left $\mathbb{C}G$-module

$$V\uparrow^G = \mathbb{C}G \otimes_{\mathbb{C}H} V.$$

The corresponding representation $\rho\uparrow^G\colon G \to V^G$ is called the ***induced representation***. The character of $G$ afforded by $\rho\uparrow^G$ is called the ***induced character***, and it is denoted by $\chi_\rho\uparrow^G$.

Let us recognize at the outset that the character of an induced representation need not restrict to the original representation of the subgroup. For example, we have seen that there is an irreducible character $\chi$ of $A_3 \cong \mathbb{Z}_3$ having complex values, whereas every irreducible character of $S_3$ has (real) integer values. A related observation is that two elements may be conjugate in a group but not conjugate in a subgroup (for example, 3-cycles are conjugate in $S_3$, for they have the same cycle structure, but they are not conjugate in the abelian group $A_3$).

The next lemma will help us compute the character afforded by an induced representation.

**Lemma C-2.84.**

(i) *If $H \subseteq G$, then $\mathbb{C}G$ is a free right $\mathbb{C}H$-module on $[G : H]$ generators.*

(ii) *If a left $\mathbb{C}H$-module $V$ has a (vector space) basis $e_1, \ldots, e_m$, then a (vector space) basis of the induced module $V\uparrow^G = \mathbb{C}G \otimes_{\mathbb{C}H} V$ is the family of all $t_i \otimes e_j$, where $t_1, \ldots, t_n$ is a transversal of $H$ in $G$.*

**Proof.**

(i) Since $t_1, \ldots, t_n$ is a transversal of $H$ in $G$ (of course, $n = [G : H]$), we see that $G$ is the disjoint union

$$G = \bigcup_i t_i H;$$

thus, for every $g \in G$, there is a unique $i$ and a unique $h \in H$ with $g = t_i h$. We claim that $t_1, \ldots, t_n$ is a basis of $\mathbb{C}G$ viewed as a right $\mathbb{C}H$-module.

If $u \in \mathbb{C}G$, then $u = \sum_g a_g g$, where $a_g \in \mathbb{C}$. Rewrite each term

$$a_g g = a_g t_i h = t_i a_g h$$

(scalars in $\mathbb{C}$ commute with everything), collect terms involving the same $t_i$, and obtain $u = \sum_i t_i \eta_i$, where $\eta_i \in \mathbb{C}H$.

To prove uniqueness of this expression, suppose that $0 = \sum_i t_i \eta_i$, where $\eta_i \in \mathbb{C}H$. Now $\eta_i = \sum_{h\in H} a_{ih} h$, where $a_{ih} \in \mathbb{C}$. Substituting,

$$0 = \sum_{i,h} a_{ih} t_i h.$$

But $t_i h = t_j h'$ if and only if $i = j$ and $h = h'$, so that $0 = \sum_{i,h} a_{ih} t_i h = \sum_{g\in G} a_{ih} g$, where $g = t_i h$. Since the elements of $G$ form a basis of $\mathbb{C}G$ (viewed as a vector space over $\mathbb{C}$), we have $a_{ih} = 0$ for all $i, h$, and so $\eta_i = 0$ for all $i$.

(ii) By Theorem B-4.86 in Part 1,

$$\mathbb{C}G \otimes_{\mathbb{C}H} V \cong \bigoplus_i t_i \mathbb{C}H \otimes_{\mathbb{C}H} V.$$

It follows that every $u \in \mathbb{C}G \otimes_{\mathbb{C}H} V$ has a unique expression as a $\mathbb{C}$-linear combination of $t_i \otimes e_j$, and so these elements comprise a basis. •

**Notation.** If $H \subseteq G$ and $\chi\colon H \to \mathbb{C}$ is a function, then $\dot\chi\colon G \to \mathbb{C}$ is given by

$$\dot\chi(g) = \begin{cases} 0 & \text{if } g \notin H, \\ \chi(g) & \text{if } g \in H. \end{cases}$$

**Theorem C-2.85.** *If $\chi_\sigma$ is the character afforded by a representation $\sigma\colon H \to \mathrm{GL}(V)$ of a subgroup $H$ of a group $G$, then the induced character $\chi_\sigma\uparrow^G$ is given by*

$$\chi_\sigma\uparrow^G(g) = \frac{1}{|H|} \sum_{a\in G} \dot\chi_\sigma(a^{-1}ga).$$

**Proof.** Let $t_1, \ldots, t_n$ be a transversal of $H$ in $G$, so that $G$ is the disjoint union $G = \bigcup_i t_i H$, and let $e_1, \ldots, e_m$ be a (vector space) basis of $V$. By Lemma C-2.84, a basis for the vector space $V^G = \mathbb{C}G \otimes_{\mathbb{C}H} V$ consists of all $t_i \otimes e_j$. If $g \in G$, we compute the matrix of left multiplication by $g$ relative to this basis. Note that

$$g t_i = t_{k(i)} h_i,$$

where $h_i \in H$, and so

$$g(t_i \otimes e_j) = (g t_i) \otimes e_j = t_{k(i)} h_i \otimes e_j = t_{k(i)} \otimes \sigma(h_i) e_j$$

(the last equation holds because we can slide any element of $H$ across the tensor sign). Now $g(t_i \otimes e_j)$ is written as a $\mathbb{C}$-linear combination of *all* the basis elements of $V\uparrow^G$, for the coefficients $t_p \otimes e_j$ for $p \neq k(i)$ are all 0. Hence, $\sigma\uparrow^G(g)$ gives the $nm \times nm$ matrix whose $m$ columns labeled by $t_i \otimes e_j$, for fixed $i$, are all zero except for an $m \times m$ block equal to

$$[a_{pq}(h_i)] = [a_{pq}(t_{k(i)}^{-1} g t_i)].$$

Thus, the big matrix is partitioned into $m \times m$ blocks, most of which are 0, and a nonzero block is on the diagonal of the big matrix if and only if $k(i) = i$; that is,

$$t_{k(i)}^{-1} g t_i = t_i^{-1} g t_i = h_i \in H.$$

The induced character is the trace of the big matrix, which is the sum of the traces of these blocks on the diagonal. Therefore,

$$\chi_\sigma\uparrow^G(g) = \sum_{t_i^{-1}gt_i\in H} \operatorname{tr}([a_{pq}(t_i^{-1}gt_i)]) = \sum_i \dot\chi_\sigma(t_i^{-1}gt_i)$$

(remember that $\dot\chi_\sigma$ is 0 outside of $H$). We now rewrite the summands (to get a formula that does not depend on the choice of the transversal): if $t_i^{-1}gt_i \in H$, then $(t_ih)^{-1}g(t_ih) = h^{-1}(t_i^{-1}gt_i)h$ in $H$, so that, for fixed $i$,

$$\sum_{h\in H} \dot\chi_\sigma\big((t_ih)^{-1}g(t_ih)\big) = |H|\dot\chi_\sigma(t_i^{-1}gt_i),$$

because $\chi_\sigma$ is a class function on $H$. Therefore,

$$\chi_\sigma\uparrow^G(g) = \sum_i \dot\chi_\sigma(t_i^{-1}gt_i) = \frac{1}{|H|}\sum_{i,h} \dot\chi_\sigma\big((t_ih)^{-1}g(t_ih)\big) = \frac{1}{|H|}\sum_{a\in G} \dot\chi_\sigma(a^{-1}ga). \quad \bullet$$

**Remark.** We have been considering induced characters, but it is easy to generalize the discussion to ***induced class functions***. If $H \subseteq G$, then a class function $\theta$ on $H$ has a unique expression as a $\mathbb{C}$-linear combination of irreducible characters of $H$, say, $\theta = \sum c_i\chi_i$, and so we can define

$$\theta\uparrow^G = \sum c_i\chi_i\uparrow^G.$$

It is plain that $\theta\uparrow^G$ is a class function on $G$ and that the formula in Theorem C-2.85 extends to induced class functions. ◀

If, for $h \in H$, the matrix of $\sigma(h)$ (with respect to the basis $e_1, \dots, e_m$ of $V$) is $B(h)$, then define $m \times m$ matrices $\dot B(g)$, for all $g \in G$, by

$$\dot B(g) = \begin{cases} 0 & \text{if } g \notin H, \\ B(g) & \text{if } g \in H. \end{cases}$$

The proof of Theorem C-2.85 allows us to picture the matrix of the induced representation in block form

$$\sigma\uparrow^G(g) = \begin{bmatrix} \dot B(t_1^{-1}gt_1) & \dot B(t_1^{-1}gt_2) & \cdots & \dot B(t_1^{-1}gt_n) \\ \dot B(t_2^{-1}gt_1) & \dot B(t_2^{-1}gt_2) & \cdots & \dot B(t_2^{-1}gt_n) \\ \vdots & \vdots & \vdots & \vdots \\ \dot B(t_n^{-1}gt_1) & \dot B(t_n^{-1}gt_2) & \cdots & \dot B(t_n^{-1}gt_n) \end{bmatrix}.$$

**Corollary C-2.86.** *Let $H$ be a subgroup of a finite group $G$ and let $\chi$ be a character on $H$.*

(i) $\chi\uparrow^G(1) = [G : H]\chi(1)$.

(ii) *If $H \lhd G$, then $\chi\uparrow^G(g) = 0$ for all $g \notin H$.*

**Proof.**

(i) For all $a \in G$, we have $a^{-1}1a = 1$, so that there are $|G|$ terms in the sum $\chi\uparrow^G(1) = \frac{1}{|H|}\sum_{a\in G}\dot{\chi}(a^{-1}ga)$ that are equal to $\chi(1)$; hence,

$$\chi\uparrow^G(1) = \frac{|G|}{|H|}\chi(1) = [G:H]\chi(1).$$

(ii) If $H \lhd G$, then $g \notin H$ implies that $a^{-1}ga \notin H$ for all $a \in G$. Therefore, $\dot{\chi}(a^{-1}ga) = 0$ for all $a \in G$, and so $\chi\uparrow^G(g) = 0$. •

**Example C-2.87.** Let $H \subseteq G$ be a subgroup of index $n$, let $X = \{t_1H, \ldots, t_nH\}$ be the family of left cosets of $H$, and let $\varphi\colon G \to S_X$ be the (permutation) representation of $G$ on the cosets of $H$. As in Example C-2.67(v), we may regard $\varphi\colon G \to \mathrm{GL}(V)$, where $V$ is the vector space over $\mathbb{C}$ having basis $X$; that is, $\varphi$ is a representation in the sense of this section.

We claim that if $\chi_\varphi$ is the character afforded by $\varphi$, then $\chi_\varphi = \epsilon\uparrow^G$, where $\epsilon$ is the trivial character on $H$. On the one hand, Example C-2.67(v) shows that

$$\chi_\varphi(g) = \mathrm{Fix}(\varphi(g))$$

for every $g \in G$. On the other hand, suppose $\varphi(g)$ is the permutation (in two-rowed notation)

$$\varphi(g) = \begin{pmatrix} t_1H & \ldots & t_nH \\ gt_1H & \ldots & gt_nH \end{pmatrix}.$$

Now $gt_iH = t_iH$ if and only if $t_i^{-1}gt_i \in H$. Thus, $\dot{\epsilon}(t_i^{-1}gt_i) \neq 0$ if and only if $gt_iH = t_iH$, and so

$$\epsilon\uparrow^G(g) = \mathrm{Fix}(\varphi(g)). \quad \blacktriangleleft$$

Even though a character $\lambda$ of a subgroup $H$ is irreducible, its induced character need not be irreducible. For example, let $G = S_3$ and let $H$ be the cyclic subgroup generated by (1 2). The linear representation $\sigma = \mathrm{sgn}\colon H \to \mathbb{C}$ is irreducible, and it affords the character $\chi_\sigma$ with

$$\chi_\sigma(1) = 1 \quad \text{and} \quad \chi_\sigma((1\,2)) = -1.$$

Using the formula for the induced character, we find that

$$\chi_\sigma\uparrow^{S_3}(1) = 3, \quad \chi_\sigma\uparrow^{S_3}((1\,2)) = -1, \quad \text{and} \quad \chi_\sigma\uparrow^{S_3}((1\,2\,3)) = 0.$$

Corollary C-2.75 shows that $\chi_\sigma\uparrow^{S_3}$ is not irreducible, for $(\chi_\sigma\uparrow^{S_3}, \chi_\sigma\uparrow^{S_3}) = 2$. It is easy to see that $\chi_\sigma\uparrow^{S_3} = \chi_2 + \chi_3$, the latter being the nontrivial irreducible characters of $S_3$.

Here is an important result of Brauer. Call a subgroup $E$ of a finite group $G$ ***elementary*** if $E = Z \times P$, where $Z$ is cyclic and $P$ is a $p$-group for some prime $p$.

**Theorem C-2.88 (Brauer).** *Every complex character $\theta$ on a finite group $G$ has the form*

$$\theta = \sum_i m_i\mu_i\uparrow^G,$$

*where $m_i \in \mathbb{Z}$ and the $\mu_i$ are linear characters on elementary subgroups of $G$.*

**Proof.** See Curtis-Reiner [**48**], p. 283, or Serre [**201**], Chapter 10. •

**Definition.** If $H$ is a subgroup of a group $G$, then every representation $\sigma\colon G \to \mathrm{GL}(V)$ gives, by restriction, a representation $\sigma|H\colon H \to \mathrm{GL}(V)$. (In terms of modules, every left $\mathbb{C}G$-module $V$ can be viewed as a left $\mathbb{C}H$-module.) We call $\sigma|H$ the ***restriction*** of $\sigma$, and we denote it by $\sigma\downarrow_H$. The character of $H$ afforded by $\sigma\downarrow_H$ is denoted by $\chi_\sigma\downarrow_H$.

The next result displays the relation between characters on a group and characters on a subgroup.

**Theorem C-2.89 (Frobenius Reciprocity).** *Let $H$ be a subgroup of a group $G$, let $\chi$ be a class function on $G$, and let $\theta$ be a class function on $H$. Then*

$$(\theta\uparrow^G, \chi)_G = (\theta, \chi\downarrow_H)_H,$$

*where $(\square,\square)_G$ denotes the inner product on $\mathrm{CF}(G)$ and $(\square,\square)_H$ denotes the inner product on $\mathrm{CF}(H)$.*

**Proof.** We have

$$\begin{aligned}(\theta\uparrow^G, \chi)_G &= \frac{1}{|G|}\sum_{g\in G} \theta\uparrow^G(g)\overline{\chi(g)}\\ &= \frac{1}{|G|}\sum_{g\in G}\frac{1}{|H|}\sum_{a\in G}\dot\theta(a^{-1}ga)\overline{\chi(g)}\\ &= \frac{1}{|G|}\frac{1}{|H|}\sum_{a,g\in G}\dot\theta(a^{-1}ga)\overline{\chi(a^{-1}ga)},\end{aligned}$$

the last equation occurring because $\chi$ is a class function. For fixed $a \in G$, as $g$ ranges over $G$, then so does $a^{-1}ga$. Therefore, writing $x = a^{-1}ga$, the equations continue:

$$\begin{aligned}&= \frac{1}{|G|}\frac{1}{|H|}\sum_{a,x\in G}\dot\theta(x)\overline{\chi(x)}\\ &= \frac{1}{|G|}\frac{1}{|H|}\sum_{a\in G}\Big(\sum_{x\in G}\dot\theta(x)\overline{\chi(x)}\Big)\\ &= \frac{1}{|G|}\frac{1}{|H|}|G|\sum_{x\in G}\dot\theta(x)\overline{\chi(x)}\\ &= \frac{1}{|H|}\sum_{x\in G}\dot\theta(x)\overline{\chi(x)}\\ &= (\theta, \chi\downarrow_H)_H,\end{aligned}$$

the next to last equation holding because $\dot\theta(x)$ vanishes off the subgroup $H$. •

The following elementary remark facilitates the computation of induced class functions.

**Lemma C-2.90.** *Let $H$ be a subgroup of a finite group $G$, and let $\chi$ be a class function on $H$. Then*

$$\chi\downarrow^G(g) = \frac{1}{|H|}\sum_i |C_G(g_i)|\dot{\chi}(g_i^{-1}gg_i).$$

**Proof.** Let $|C_G(g_i)| = m_i$. If $a_0^{-1}g_ia_0 = g$, we claim that there are exactly $m_i$ elements $a \in G$ with $a^{-1}g_ia = g$. There are at least $m_i$ elements in $G$ conjugating $g_i$ to $g$, namely, all $aa_0$ for $a \in C_G(g_i)$. There are at most $m_i$ elements, for if $b^{-1}g_ib = g$, then $b^{-1}g_ib = a_0^{-1}g_ia_0$, and so $a_0b \in C_G(g_i)$. The result now follows by collecting terms involving $g_i$'s in the formula for $\chi\downarrow^G(g)$. •

| $g_i$ | (1) | (1 2 3) | (1 3 2) | (1 2)(3 4) |
|---|---|---|---|---|
| $h_i$ | 1 | 4 | 4 | 3 |
| $\chi_1$ | 1 | 1 | 1 | 1 |
| $\chi_2$ | 1 | $\omega$ | $\omega^2$ | 1 |
| $\chi_3$ | 1 | $\omega^2$ | $\omega$ | 1 |
| $\chi_4$ | 3 | 0 | 0 | $-1$ |

**Table 5.** Character table of $A_4$.

**Example C-2.91.** The group $A_4$ consists of the identity, eight 3-cycles, and three products of disjoint transpositions. In $S_4$, all the 3-cycles are conjugate; if $g = (1\ 2\ 3)$, then $[S_4 : C_{S_4}(g)] = 8$. It follows that $|C_{S_4}(g)| = 3$, and so $C_{S_4}(g) = \langle g\rangle$. Therefore, in $A_4$, the number of conjugates of $g$ is $[A_4 : C_{A_4}(g)] = 4$ (we know that $C_{A_4}(g) = A_4 \cap C_{S_4}(g) = \langle g\rangle$). The reader may show that $g$ and $g^{-1}$ are not conjugate, and so we have verified the first two rows of the character table. Table 5 is the character table of $A_4$, where $\omega = e^{2\pi i/3}$ is a primitive cube root of unity.

The rows for $\chi_2$ and $\chi_3$ are liftings of linear characters of $A_4/\mathbf{V} \cong \mathbb{Z}_3$. Note that if $h = (1\ 2)(3\ 4)$, then $\chi_2(h) = \chi_2(1) = 1$, because $\mathbf{V}$ is the kernel of the lifted representation; similarly, $\chi_3(h) = 1$. Now $\chi_4(1) = 3$, because $3 + (n_4)^2 = 12$. The bottom row arises from orthogonality of the columns. (We can check, using Corollary C-2.75, that the character of degree 3 is irreducible.) ◀

**Example C-2.92.** Table 6 is the character table of $S_4$. We know, for all $n$, that two permutations in $S_n$ are conjugate if and only if they have the same cycle structure; the sizes of the conjugacy classes in $S_4$ were computed in Table 1 on page 121 in Part 1.

The rows for $\chi_2$ and $\chi_3$ are liftings of irreducible characters of $S_4/\mathbf{V} \cong S_3$. The entries in the fourth column of these rows arise from $(1\ 2)\mathbf{V} = (1\ 2\ 3\ 4)\mathbf{V}$; the entries in the last column of these rows arise from $\mathbf{V}$ being the kernel (in either case), so that $\chi_j((1\ 2)(3\ 4)) = \chi_j(1)$ for $j = 2, 3$.

We complete the first column using $24 = 1+1+4+n_4^2+n_5^2$; thus, $n_4 = 3 = n_5$. Let us see whether $\chi_4$ is an induced character; if it is, then Corollary C-2.86(i)

| $g_i$ | (1) | (1 2) | (1 2 3) | (1 2 3 4) | (1 2)(3 4) |
|---|---|---|---|---|---|
| $h_i$ | 1 | 6 | 8 | 6 | 3 |
| $\chi_1$ | 1 | 1 | 1 | 1 | 1 |
| $\chi_2$ | 1 | −1 | 1 | −1 | 1 |
| $\chi_3$ | 2 | 0 | −1 | 0 | 2 |
| $\chi_4$ | 3 | 1 | 0 | −1 | −1 |
| $\chi_5$ | 3 | −1 | 0 | 1 | −1 |

**Table 6.** Character table of $S_4$.

shows that it arises from a linear character of a subgroup $H$ of index 3. Such a subgroup has order 8, and so it is a Sylow 2-subgroup; that is, $H \cong D_8$. Let us choose one such subgroup:

$$H = \langle \mathbf{V}, (1\ 3) \rangle = \mathbf{V} \cup \{(1\ 3), (2\ 4), (1\ 2\ 3\ 4), (1\ 4\ 3\ 2)\}.$$

The conjugacy classes are

$$\begin{aligned} C_1 &= \{1\}, \\ C_2 &= \{(1\ 3)(2\ 4)\}, \\ C_3 &= \{(1\ 2)(3\ 4), (1\ 4)(2\ 3)\}, \\ C_4 &= \{(1\ 3), (2\ 4)\}, \\ C_5 &= \{(1\ 2\ 3\ 4), (1\ 4\ 3\ 2)\}. \end{aligned}$$

Let $\theta$ be the character on $H$ defined by

| $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|
| 1 | 1 | −1 | 1 | −1. |

Define $\chi_4 = \theta|^{S_4}$. Using the formula for induced characters, assisted by Lemma C-2.90, we obtain the fourth row of the character table. However, before going on to row 5, we observe that Corollary C-2.75 shows that $\chi_4$ is irreducible, for $(\chi_4, \chi_4) = 1$. Finally, the orthogonality relations allow us to compute row 5. ◀

## C-2.11. Algebraic Integers Interlude

At this point in the story, we must introduce algebraic integers. Recall that a complex number $z$ is an ***algebraic number*** if it is a root of a nonzero $f(x) \in \mathbb{Q}[x]$. An ***algebraic integer*** is a complex number $z$ which is a root of a monic $g(x) \in \mathbb{Z}[x]$. Since $G$ is a finite group, Lagrange's Theorem gives $g^{|G|} = 1$ for all $g \in G$. It follows that if $\sigma\colon G \to \mathrm{GL}(V)$ is a representation, then $\sigma(g)^{|G|} = I$ for all $g$; hence, all the eigenvalues of $\sigma(g)$ are $|G|$th roots of unity, and so all the eigenvalues are algebraic integers. The trace of $\sigma(g)$, being the sum of the eigenvalues, is also an algebraic integer.

**Definition.** A complex number $\alpha$ is called an ***algebraic integer*** if $\alpha$ is a root of a monic $f(x) \in \mathbb{Z}[x]$.

We note that it is crucial, in the definition of algebraic integer, that $f(x) \in \mathbb{Z}[x]$ be monic. Every algebraic number $z$, that is, every complex number $z$ that is a root of some polynomial $g \in \mathbb{Q}[x]$, is necessarily a root of some polynomial $h \in \mathbb{Z}[x]$; just clear the denominators of the coefficients of $g$.

Of course, every ordinary integer is an algebraic integer. To contrast ordinary integers with more general algebraic integers, elements of $\mathbb{Z}$ may be called ***rational integers***. Theorem A-3.101 in Part 1 can be rephrased: an algebraic integer is either an integer or it is irrational.

The next proposition shows that the sum and product of algebraic integers are themselves algebraic integers. If $\alpha$ and $\beta$ are algebraic integers, it is not too difficult to show there are monic polynomials in $\mathbb{Q}[x]$ having $\alpha + \beta$ and $\alpha\beta$ as roots, but it is harder to show there are such polynomials in $\mathbb{Z}[x]$.

We now characterize algebraic integers.

**Proposition C-2.93.** *Let $\alpha \in \mathbb{C}$ and define $\mathbb{Z}[\alpha] = \big\{g(\alpha)\colon g(x) \in \mathbb{Z}[x]\big\}$.*

(i) *$\mathbb{Z}[\alpha]$ is a subring of $\mathbb{C}$.*

(ii) *$\alpha$ is an algebraic integer if and only if $\mathbb{Z}[\alpha]$ is a finitely generated additive abelian group.*

(iii) *The set $\mathbb{A}$ of all the algebraic integers is a subring of $\mathbb{C}$, and $\mathbb{A} \cap \mathbb{Q} = \mathbb{Z}$.*

**Proof.**

(i) If $g = 1$ is the constant polynomial, then $g \in \mathbb{Z}[x]$; hence, $1 = g(\alpha)$ and so $1 \in \mathbb{Z}[\alpha]$. Suppose that $f(\alpha), g(\alpha) \in \mathbb{Z}[\alpha]$, where $f(x), g(x) \in \mathbb{Z}[x]$. Now $f+g$ and $fg$ lie in $\mathbb{Z}[x]$, so that $f(\alpha) + g(\alpha), f(\alpha)g(\alpha) \in \mathbb{Z}[\alpha]$. Therefore, $\mathbb{Z}[\alpha]$ is a subring of $\mathbb{C}$.

(ii) If $\alpha$ is an algebraic integer, there is a monic polynomial $f(x) \in \mathbb{Z}[x]$ having $\alpha$ as a root. We claim that if $\deg(f) = n$, then $\mathbb{Z}[\alpha] = G$, where $G$ is the set of all linear combinations $m_0 + m_1\alpha + \cdots + m_{n-1}\alpha^{n-1}$ with $m_i \in \mathbb{Z}$. Clearly, $G \subseteq \mathbb{Z}[\alpha]$. For the reverse inclusion, each element $u \in \mathbb{Z}[\alpha]$ has the form $u = g(\alpha)$, where $g(x) \in \mathbb{Z}[x]$. Since $f$ is monic, the Division Algorithm gives $q(x), r(x) \in \mathbb{Z}[x]$ with $g = qf + r$, where either $r = 0$ or $\deg(r) < \deg(f) = n$ (see Corollary A-3.48 in Part 1). Therefore,

$$u = g(\alpha) = q(\alpha)f(\alpha) + r(\alpha) = r(\alpha) \in G.$$

Thus, the additive group of $\mathbb{Z}[\alpha]$ is finitely generated.

Conversely, if the additive group of the commutative ring $\mathbb{Z}[\alpha]$ is finitely generated, that is, $\mathbb{Z}[\alpha] = \langle g_1, \ldots, g_m \rangle$ as an abelian group, then each $g_j$ is a $\mathbb{Z}$-linear combination of powers of $\alpha$. Let $m$ be the largest power of $\alpha$ occurring in any of these $g$'s. Since $\mathbb{Z}[\alpha]$ is a commutative ring, $\alpha^{m+1} \in \mathbb{Z}[\alpha]$; hence, $\alpha^{m+1}$ can be expressed as a $\mathbb{Z}$-linear combination of smaller powers of $\alpha$; say, $\alpha^{m+1} = \sum_{i=0}^{m} b_i\alpha^i$, where $b_i \in \mathbb{Z}$. Therefore, $\alpha$ is a root of $f(x) = x^{m+1} - \sum_{i=0}^{m} b_i x^i$, which is a monic polynomial in $\mathbb{Z}[x]$, and so $\alpha$ is an algebraic integer.

(iii) Suppose $\alpha$ and $\beta$ are algebraic integers; let $\alpha$ be a root of a monic $f \in \mathbb{Z}[x]$ of degree $n$, and let $\beta$ be a root of a monic $g \in \mathbb{Z}[x]$ of degree $m$. Now $\mathbb{Z}[\alpha\beta]$

is an additive subgroup of $G = \langle \alpha^i \beta^j : 0 \le i < n, 0 \le j < m \rangle$. Since $G$ is finitely generated, so is its subgroup $\mathbb{Z}[\alpha\beta]$, by Theorem B-2.28 in Part 1, and so $\alpha\beta$ is an algebraic integer. Similarly, $\mathbb{Z}[\alpha + \beta]$ is an additive subgroup of $\langle \alpha^i \beta^j : i + j \le n + m - 2 \rangle$, and so $\alpha + \beta$ is also an algebraic integer.

The last statement is Theorem A-3.101 in Part 1: if an algebraic integer is not an integer, it is irrational. •

**Theorem C-2.94.** *If $G$ is a finite group, then $\chi(g)$ is an algebraic integer for every character $\chi$ and every $g \in G$.*

**Proof.** This follows from Proposition C-2.68 and Proposition C-2.93(i). •

**Corollary C-2.95.**

(i) *If $M$ is a finitely generated abelian group that is a faithful left $R$-module for some ring $R$, then the additive group of $R$ is finitely generated.*

(ii) *Let $\mathbb{Z}[\alpha]$ be the subring of $\mathbb{C}$ generated by a complex number $\alpha$. If there is a faithful $\mathbb{Z}[\alpha]$-module $M$ that is finitely generated as an abelian group, then $\alpha$ is an algebraic integer.*

**Proof.**

(i) The ring $R$ is isomorphic to a subring of $\mathrm{End}_{\mathbb{Z}}(M)$, by Proposition B-1.22 in Part 1, for $M$ is faithful. Since $M$ is finitely generated, Exercise B-4.16 on page 474 in Part 1 shows that $\mathrm{End}_{\mathbb{Z}}(M) = \mathrm{Hom}_{\mathbb{Z}}(M, M)$ is finitely generated. Therefore, the additive group of $R$ is finitely generated, by Proposition C-2.93.

(ii) It suffices to prove that the ring $\mathbb{Z}[\alpha]$ is finitely generated as an abelian group, by Proposition C-2.93, and this follows from part (i). •

This last corollary gives a technique for proving that an integer $a$ is a divisor of an integer $b$. If we can prove that $b/a$ is an algebraic integer, then it must be an integer, for it is obviously rational.

Since $\mathbb{Z}[x] \subseteq \mathbb{Q}[x]$, every algebraic integer $\alpha$ has a unique minimal polynomial $m(x) = \mathrm{irr}(\alpha, \mathbb{Q}) \in \mathbb{Q}[x]$, and $m$ is irreducible in $\mathbb{Q}[x]$.

**Corollary C-2.96.** *If $\alpha$ is an algebraic integer, then* $\mathrm{irr}(\alpha, \mathbb{Q})$ *lies in* $\mathbb{Z}[x]$.

**Proof.** Let $p(x) \in \mathbb{Z}[x]$ be the monic polynomial of least degree having $\alpha$ as a root. If $p(x) = G(x)H(x)$ in $\mathbb{Q}[x]$, where $\deg(G) < \deg(p)$ and $\deg(H) < \deg(p)$, then $\alpha$ is a root of either $G$ or $H$. By Gauss's Lemma A-3.65 in Part 1, there is a factorization $p = gh$ in $\mathbb{Z}[x]$ with $\deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$; in fact, there are rationals $c$ and $d$ with $g = cG$ and $h = dH$. If $a$ is the leading coefficient of $g$ and $b$ is the leading coefficient of $h$, then $ab = 1$, for $p$ is monic. Therefore, we may assume that $a = 1 = b$, for $a, b \in \mathbb{Z}$ (the only other option is $a = -1 = b$); that is, we may assume that both $g$ and $h$ are monic. Since $\alpha$ is a root of $g$ or $h$, we have contradicted $p$ being a monic polynomial in $\mathbb{Z}[x]$ of least degree having $\alpha$ as a root. It follows that $p(x) = \mathrm{irr}(\alpha, \mathbb{Q})$, for the latter is the unique monic irreducible polynomial in $\mathbb{Q}[x]$ having $\alpha$ as a root. •

**Remark.** We define the (algebraic) ***conjugates*** of $\alpha$ to be the roots of $\mathrm{irr}(\alpha, \mathbb{Q})$, and we define the ***norm*** of $\alpha$ to be the absolute value of the product of the conjugates of $\alpha$. Of course, the norm of $\alpha$ is just the absolute value of the constant term of $\mathrm{irr}(\alpha, \mathbb{Q})$, and so it is an (ordinary) integer. ◀

We can now prove the following interesting result.

**Theorem C-2.97.** *The degrees $n_i$ of the irreducible characters of a finite group $G$ are divisors of $|G|$.*

**Proof.** By Theorem A-3.101 in Part 1, the rational number $\alpha = |G|/n_i$ is an integer if it is also an algebraic integer. Now Corollary C-2.95 says that $\alpha$ is an algebraic integer if there is a faithful $\mathbb{Z}[\alpha]$-module $M$ that is a finitely generated abelian group, where $\mathbb{Z}[\alpha]$ is the smallest subring of $\mathbb{C}$ containing $\alpha$.

By Proposition C-2.73, we have

$$e_i = \sum_{g \in G} \frac{n_i}{|G|} \chi_i(g^{-1}) g = \sum_{g \in G} \frac{1}{\alpha} \chi_i(g^{-1}) g.$$

Hence, $\alpha e_i = \sum_{g \in G} \chi_i(g^{-1}) g$. But $e_i$ is an idempotent: $e_i^2 = e_i$, and so

$$\alpha e_i = \sum_{g \in G} \chi_i(g^{-1}) g e_i.$$

Define $M$ to be the abelian subgroup of $\mathbb{C}G$ generated by all elements of the form $\zeta g e_i$, where $\zeta$ is a $|G|$th root of unity and $g \in G$; of course, $M$ is a finitely generated abelian group.

To see that $M$ is a $\mathbb{Z}[\alpha]$-module, it suffices to show that $\alpha M \subseteq M$. But

$$\alpha \zeta g e_i = \zeta g \alpha e_i = \zeta g \sum_{h \in G} \chi_i(h^{-1}) h e_i = \sum_{h \in G} \chi_i(h^{-1}) \zeta g h e_i.$$

This last element lies in $M$, however, because $\chi_i(h^{-1})$ is a sum of $|G|$th roots of unity.

Finally, if $\beta \in \mathbb{C}$ and $u \in \mathbb{C}G$, then $\beta u = 0$ if and only if $\beta = 0$ or $u = 0$. Since $\mathbb{Z}[\alpha] \subseteq \mathbb{C}$ and $M \subseteq \mathbb{C}G$, however, it follows that $M$ is a faithful $\mathbb{Z}[\alpha]$-module, as desired. •

In Tables 3 and 6, we saw that all character values of the symmetric groups $S_3$ and $S_4$ are integers, while Table 4 shows that this is also true of $D_8$ and $\mathbf{Q}$. We now show that this is true for all the symmetric groups.

**Lemma C-2.98.** *If $\chi$ is a character of a finite group $G$ and $g \in G$, then $\chi(g)$ is an integer if and only if $\chi(g)$ is rational.*

**Proof.** Since $\chi(g)$ is an algebraic integer, by Theorem C-2.94, the result follows from Proposition C-2.93(iii). •

Here is the key group-theoretic property.

**Definition.** A finite group $G$ is ***generator-conjugate*** if, for all pairs $g, g' \in G$ with $\langle g \rangle = \langle g' \rangle$, the elements $g$ and $g'$ are conjugate.

**Lemma C-2.99.** *If $G$ is a finite generator-conjugate group, then $\chi(g) \in \mathbb{Q}$ for every $g \in G$ and character $\chi$.*

**Remark.** The converse is also true. ◀

**Proof.** Let $\tau\colon G \to \mathrm{GL}(V)$ be a representation affording the character $\chi_\tau$. If $g$ has order $m$ and $\zeta$ is a primitive $m$th root of unity, then all the eigenvalues of $\tau(g)$ are powers of $\zeta$:

$$\chi_\tau(g) = \mathrm{tr}(\tau(g)) = \epsilon_1 + \cdots + \epsilon_t,$$

where each $\epsilon_i = \zeta^{j_i}$. If $g'$ is another generator of $\langle g\rangle$, then $g' = g^k$ for some $k$ with $\gcd(k, m) = 1$. Since $g'$ has order $m$, all the eigenvalues of $\tau(g')$ are also $m$th roots of unity; these eigenvalues are $\epsilon_i^k$: if $\tau(g)(v) = \epsilon_i v$, then $\tau(g^k)(v) = \tau(g)^k(v) = \epsilon_i^k v$).

$$\chi_\tau(g') = \chi_\tau(g^k) = \mathrm{tr}(\tau(g^k)) = \epsilon_1^k + \cdots + \epsilon_t^k.$$

The extension field $\mathbb{Q}(\zeta)/\mathbb{Q}$ is a Galois extension; indeed, it is the splitting field of $x^m - 1$ over $\mathbb{Q}$. Recall Proposition A-5.12 in Part 1: $\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$ is a cyclic group with generator $\sigma_k\colon \zeta \mapsto \zeta^k$, where $k$ satisfies $\gcd(k, m) = 1$. Now

$$\begin{aligned}\sigma_k(\chi_\tau(g)) &= \sigma_k(\epsilon_1 + \cdots + \epsilon_t)\\ &= \sigma_k(\epsilon_1) + \cdots + \sigma_k(\epsilon_t)\\ &= \epsilon_1^k + \cdots + \epsilon_t^k\\ &= \chi_\tau(g').\end{aligned}$$

By hypothesis, $g$ and $g'$ are conjugate, so that $\chi_\tau(g') = \chi_\tau(g') = \chi_\tau(g)$, because $\chi_\tau$ is a class function, and so

$$\sigma_k(\chi_\tau(g)) = \chi_\tau(g).$$

Therefore, $\chi_\tau(g) \in \mathbb{Q}$, the fixed field, for $\sigma_k$ generates $\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$. •

**Theorem C-2.100.** *If $G$ is a finite generator-conjugate group, then all the entries in its character table are integers.*

**Remark.** The converse is also true. ◀

**Proof.** Lemma C-2.99 says that the statement is true if *integers* is replaced by *rationals*, while Lemma C-2.98 allows this replacement. •

**Corollary C-2.101.** *All the character values of $S_n$, $D_8$, and $\mathbf{Q}$ lie in $\mathbb{Z}$.*

**Proof.** See Exercise C-2.42 below. •

We will present two important applications of character theory in the next section; for other applications, as well as a more serious study of representations, the interested reader should look at the books by Curtis–Reiner [**48**], Feit [**65**], Huppert [**102**], Isaacs [**105**], and Serre [**201**].

Representation theory is used throughout the proof of the Classification Theorem of Finite Simple Groups. An account of this theorem, describing the infinite families of such groups as well as the 26 sporadic simple groups, can be found in the ATLAS [**44**]. This book contains the character tables of every simple group of order under $10^{25}$ as well as the character tables of all the sporadic groups.

## Exercises

**C-2.29.** Prove that if $\theta$ is a generalized character of a finite group $G$, then there are characters $\chi$ and $\psi$ with $\theta = \chi - \psi$.

∗ **C-2.30.** Prove that if $G$ is a finite group and $\sigma\colon G \to \mathrm{GL}(V)$ is a representation, then

$$\chi_\sigma(g^{-1}) = \overline{\chi_\sigma(g)}$$

for all $g \in G$.

**Hint.** Use the fact that every eigenvalue of $\sigma(g)$ is a root of unity, as well as the fact that if $A$ is a nonsingular matrix over a field $k$ and if $u_1, \ldots, u_n$ are the eigenvalues of $A$ (with multiplicities), then the eigenvalues of $A^{-1}$ are $u_1^{-1}, \ldots, u_n^{-1}$; that is, $\overline{u_1}, \ldots, \overline{u_n}$.

**C-2.31.** If $\sigma\colon G \to \mathrm{GL}(n, \mathbb{C})$ is a representation, its ***contragredient*** $\sigma^*\colon G \to \mathrm{GL}(n, \mathbb{C})$ is the function given by

$$\sigma^*(g) = \sigma(g^{-1})^\top,$$

where $\square^\top$ denotes transpose.

(i) Prove that the contragredient of a representation $\sigma$ is a representation that is irreducible when $\sigma$ is irreducible.

(ii) Prove that the character $\chi_{\sigma^*}$ afforded by the contragredient $\sigma^*$ is

$$\chi_{\sigma^*}(g) = \overline{\chi_\sigma(g)},$$

where $\overline{\chi_\sigma(g)}$ is the complex conjugate. Conclude that if $\chi$ is a character of $G$, then $\overline{\chi}$ is also a character.

∗ **C-2.32.** Construct an irreducible representation of $S_3$ of degree 2.

**C-2.33.** (i) Let $g \in G$, where $G$ is a finite group. Prove that $g$ is conjugate to $g^{-1}$ if and only if $\chi(g)$ is real for every character $\chi$ of $G$.

(ii) Prove that every character of $S_n$ is real-valued. (It is a theorem of Frobenius that every character of $S_n$ is integer-valued.)

**C-2.34.** (i) Recall that the ***character group*** $G^*$ of a finite abelian group $G$ is

$$G^* = \mathrm{Hom}(G, \mathbb{C}^\times),$$

where $\mathbb{C}^\times$ is the multiplicative group of nonzero complex numbers. Prove that $G^* \cong G$.

**Hint.** Use the Fundamental Theorem of Finite Abelian Groups.

(ii) Prove that $\mathrm{Hom}(G, \mathbb{C}^\times) \cong \mathrm{Hom}(G, \mathbb{Q}/\mathbb{Z})$ when $G$ is a finite abelian group.

**Hint.** $\mathbb{C}^\times \cong \mathbb{R}/\mathbb{Z}$, by Corollary B-4.75 in Part 1.

∗ **C-2.35.** Prove that the only linear character of a simple group is the trivial character. Conclude that if $\chi_i$ is not the trivial character, then $n_i = \chi_i(1) > 1$.

∗ **C-2.36.** Let $\theta = \chi_\sigma$ be the character afforded by a representation $\sigma$ of a finite group $G$.

(i) If $g \in G$, prove that $|\theta(g)| = \theta(1)$ if and only if $\sigma(g)$ is a scalar matrix.

**Hint.** Use Proposition A-3.109 in Part 1.

(ii) If $\theta$ is an irreducible character, prove that

$$Z(G/\ker\theta) = \{g \in G : |\theta(g)| = \theta(1)\}$$

($Z(H)$ denotes the center of a group $H$).

∗ **C-2.37.** If $G$ is a finite group, prove that the number of its (necessarily irreducible) linear representations is $[G : G']$.

**C-2.38.** Let $G$ be a finite group.

(i) If $g \in G$, show that its centralizer $|C_G(g)| = \sum_{i=1}^{r} |\chi_i(g)|^2$. Conclude that the character table of $G$ gives $|C_G(g)|$.

(ii) Show how to use the character table of $G$ to see whether $G$ is abelian.

(iii) Show how to use the character table of $G$ to find the lattice of normal subgroups of $G$ and their orders.

(iv) If $G$ is a finite group, show how to use its character table to find the commutator subgroup $G'$.

**Hint.** If $K \lhd G$, then the character table of $G/K$ is a submatrix of the character table of $G$, and so we can find the abelian quotient of $G$ having largest order.

(v) Show how to use the character table of a finite group $G$ to determine whether $G$ is solvable.

**C-2.39.** (i) Show how to use the character table of $G$ to find $|Z(G)|$.

(ii) Show how to use the character table of a finite group $G$ to determine whether $G$ is nilpotent.

**C-2.40.** Recall that the group $\mathbf{Q}$ of quaternions has the presentation

$$\mathbf{Q} = (a, b \mid a^4 = 1, a^2 = b^2, bab^{-1} = a^{-1}).$$

(i) Show that there is a representation $\sigma\colon \mathbf{Q} \to \mathrm{GL}(2, \mathbb{C})$ with

$$a \mapsto \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \text{ and } b \mapsto \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

(ii) Prove that $\sigma$ is an irreducible representation.

**C-2.41.** (i) If $\sigma\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$ are representations, prove that

$$\sigma \otimes \tau\colon G \to \mathrm{GL}(V \otimes W)$$

defined by

$$(\sigma \otimes \tau)(g) = \sigma(g) \otimes \tau(g)$$

is a representation.

(ii) Prove that the character afforded by $\sigma \otimes \tau$ is the pointwise product:

$$\chi_\sigma \chi_\tau\colon g \mapsto \mathrm{tr}(\sigma(g))\,\mathrm{tr}(\tau(g)).$$

(iii) Prove that $\mathrm{CF}(G)$ is a commutative ring (usually called the ***Burnside ring*** of $G$).

∗ **C-2.42.** (i) Prove that the groups $S_n$, for $n \geq 2$, $D_8$, and $\mathbf{Q}$ are generator-conjugate, and conclude that all the entries in their character tables are integers.

(ii) Show directly, without using Table 5, that $A_4$ is not generator-conjugate.

(iii) Find a generator-conjugate group $G$ not listed in part (i).

**C-2.43.** Prove the converse of Theorem C-2.100.

**Hint.** Use Lemma C-2.71.

## C-2.12. Theorems of Burnside and of Frobenius

Character theory will be used in this section to prove two important results in group theory: Burnside's $p^m q^n$ Theorem and a theorem of Frobenius. We begin with the following variation of Schur's Lemma.

**Proposition C-2.102 (Schur's Lemma II).** *If $\sigma\colon G \to \mathrm{GL}(V)$ is an irreducible representation and if a linear transformation $\varphi\colon V \to V$ satisfies*

$$\varphi\sigma(g) = \sigma(g)\varphi$$

*for all $g \in G$, then $\varphi$ is a scalar transformation: there exists $\omega \in \mathbb{C}$ with $\varphi = \omega 1_V$.*

**Proof.** The vector space $V$ is a $\mathbb{C}G$-module with scalar multiplication $gv = \sigma(g)(v)$ for all $v \in V$, and any linear transformation $\theta$ satisfying the equation $\theta\sigma(g) = \sigma(g)\theta$ for all $g \in G$ is a $\mathbb{C}G$-map $V^\sigma \to V^\sigma$ (for $\theta(gv) = \theta\sigma(g)(v) = \sigma(g)\theta(v) = g\theta(v)$). Since $\sigma$ is irreducible, the $\mathbb{C}G$-module $V^\sigma$ is simple.

Schur's Lemma (Theorem C-2.30) says that $\mathrm{End}(V^\sigma)$ is a division ring, and so every nonzero element in it is nonsingular. Now $\varphi - \omega 1_V \in \mathrm{End}(V^\sigma)$ for every $\omega \in \mathbb{C}$; in particular, this is so when $\omega$ is an eigenvalue of $\varphi$ (which lies in $\mathbb{C}$ because $\mathbb{C}$ is algebraically closed). The definition of eigenvalue says that $\varphi - \omega 1_V$ is singular, and so $\varphi - \omega 1_V$ is zero; that is, $\varphi = \omega 1_V$, as desired. •

Recall that if $L_i$ is a minimal left ideal in $\mathbb{C}G$ and $\lambda_i\colon G \to \mathrm{End}_{\mathbb{C}}(L_i)$ is the corresponding irreducible representation, then we extended $\lambda_i$ to a linear transformation $\widetilde{\lambda}_i\colon \mathbb{C}G \to \mathrm{End}_{\mathbb{C}}(L_i)$:

$$\widetilde{\lambda}_i(g)u_j = \begin{cases} gu_i & \text{if } j = i, \\ 0 & \text{if } j \neq i. \end{cases}$$

Thus, $\widetilde{\lambda}_i(g) = \lambda_i(g)$ for all $g \in G$. In Proposition C-2.63(ii), we proved that $\widetilde{\lambda}_i$ is a $\mathbb{C}$-algebra map.

**Corollary C-2.103.** *Let $L_i$ be a minimal left ideal in $\mathbb{C}G$, let $\lambda_i\colon G \to \mathrm{End}_{\mathbb{C}}(L_i)$ be the corresponding irreducible representation, and let $\widetilde{\lambda}_i\colon \mathbb{C}G \to \mathrm{End}_{\mathbb{C}}(L_i)$ be the algebra map of* Proposition C-2.63(ii).

(i) *If $z \in Z(\mathbb{C}G)$, then there is $\omega_i(z) \in \mathbb{C}$ with*

$$\widetilde{\lambda}_i(z) = \omega_i(z)I.$$

(ii) *The function $\omega_i\colon Z(\mathbb{C}G) \to \mathbb{C}$, given by $z \mapsto \omega_i(z)$, is an algebra map.*

**Proof.**

(i) Let $z \in Z(\mathbb{C}G)$. We verify the hypothesis of Schur's Lemma II in the special case $V = L_i$, $\sigma = \lambda_i$, and $\varphi = \widetilde{\lambda}_i(z)$. For all $g \in G$, we have $\widetilde{\lambda}_i(z)\lambda_i(g) = \lambda_i(zg)$ (for $\widetilde{\lambda}_i$ is a multiplicative map extending $\lambda_i$), while $\lambda_i(g)\widetilde{\lambda}_i(z) = \lambda_i(gz)$. These are equal, for $zg = gz$ since $z \in Z(\mathbb{C}G)$. Proposition C-2.102 now says that $\widetilde{\lambda}_i(z) = \omega_i(z)I$ for some $\omega_i(z) \in \mathbb{C}$.

(ii) This follows from the equation $\widetilde{\lambda}_i(z) = \omega_i(z)I$ and $\widetilde{\lambda}_i$ being an algebra map. •

Recall the notation on page 180. Lemma C-2.48 says that a basis for $Z(\mathbb{C}G)$ consists of the *class sums*

$$z_i = \sum_{g \in C_i} g,$$

where the conjugacy classes of $G$ are $C_1, \ldots, C_r$.

**Proposition C-2.104.** *Let $z_1, \ldots, z_r$ be the class sums of a finite group $G$.*

(i) *For each $i, j$, we have*

$$\omega_i(z_j) = \frac{h_j \chi_i(g_j)}{n_i},$$

*where $g_j \in C_j$.*

(ii) *There are nonnegative integers $a_{ij\nu}$ with*

$$z_i z_j = \sum_{\nu} a_{ij\nu} z_\nu.$$

(iii) *The complex numbers $\omega_i(z_j)$ are algebraic integers.*

**Proof.**

(i) Computing the trace of $\widetilde{\lambda}_i(z_j) = \omega_i(z_j)I$ gives

$$n_i\omega_i(z_j) = \chi_i(z_j) = \sum_{g \in C_j} \chi_i(g) = h_j\chi_i(g_j),$$

for $\chi_i$ is constant on the conjugacy class $C_j$. Therefore,

$$\omega_i(z_j) = h_j\chi_i(g_j)/n_i.$$

(ii) Choose $g_\nu \in C_\nu$. The definition of multiplication in the group algebra shows that the coefficient of $g_\nu$ in $z_i z_j$ is

$$|\{(g_i, g_j) \in C_i \times C_j : g_i g_j = g_\nu\}|,$$

the cardinality of a finite set, and hence it is a nonnegative integer. As all the coefficients of $z_\nu$ are equal (for we are in $Z(\mathbb{C}G)$), it follows that this number is $a_{ij\nu}$.

(iii) Let $M$ be the (additive) subgroup of $\mathbb{C}$ generated by all $\omega_i(z_j)$, for $j = 1, \ldots, r$. Since $\omega_i$ is an algebra map,

$$\omega_i(z_j)\omega_i(z_\ell) = \sum_{\nu} a_{j\ell\nu}\omega_i(z_\nu),$$

so that $M$ is a ring that is finitely generated as an abelian group (because $a_{ij\nu} \in \mathbb{Z}$). Hence, for each $j$, $M$ is a $\mathbb{Z}[\omega_i(z_j)]$-module that is a finitely generated abelian group. If $M$ is faithful, then Corollary C-2.95 will give $\omega_i(z_j)$ an algebraic integer. But $M \subseteq \mathbb{C}$, so that the product of nonzero elements is nonzero, and this implies that $M$ is a faithful $\mathbb{Z}[\omega_i(z_j)]$-module, as desired. •

We are almost ready to complete the proof of Burnside's Theorem.

**Proposition C-2.105.** *If $(n_i, h_j) = 1$ for some $i, j$, then either $|\chi_i(g_j)| = n_i$ or $\chi_i(g_j) = 0$.*

**Proof.** By hypothesis, there are integers $s$ and $t$ in $\mathbb{Z}$ with $sn_i + th_j = 1$, so that, for $g_j \in C_j$, we have

$$\frac{\chi_i(g_j)}{n_i} = s\chi_i(g_j) + \frac{th_j\chi_i(g_j)}{n_i}.$$

Hence, Proposition C-2.104 gives $\chi_i(g_j)/n_i$ an algebraic integer, and so $|\chi_i(g_j)| \leq n_i$, by Proposition C-2.81(i). Thus, it suffices to show that if $|\chi_i(g_j)/n_i| < 1$, then $\chi_i(g_j) = 0$.

Let $m(x) \in \mathbb{Z}[x]$ be the minimum polynomial of $\alpha = \chi_i(g_j)/n_i$; that is, $m(x)$ is the monic polynomial in $\mathbb{Z}[x]$ of least degree having $\alpha$ as a root. We proved, in Corollary C-2.96, that $m(x)$ is irreducible in $\mathbb{Q}[x]$. If $\alpha'$ is a root of $m(x)$, then Proposition A-5.14 in Part 1 shows that $\alpha' = \sigma(\alpha)$ for some $\sigma \in \mathrm{Gal}(E/\mathbb{Q})$, where $E/\mathbb{Q}$ is the splitting field of $m(x)(x^{|G|} - 1)$. But

$$\alpha = \frac{1}{n_i}\left(\varepsilon_1 + \cdots + \varepsilon_{n_i}\right),$$

where the $\varepsilon$'s are $|G|$th roots of unity, and so $\alpha' = \sigma(\alpha)$ is also such a sum. It follows that $|\alpha'| \leq 1$ (as in the proof of Proposition C-2.81(i)). Therefore, if $N(\alpha)$ is the norm of $\alpha$ (which is, by definition, the absolute value of the product of all the roots of $m(x)$; see the remark on page 196), then $N(\alpha) < 1$, for we are assuming that $|\alpha| < 1$. But $N(\alpha)$ is the absolute value of the constant term of $m(x)$, which is an integer. Therefore, $N(\alpha) = 0$, hence $\alpha = 0$, and so $\chi_i(g_j) = 0$, as claimed. •

At last, we can use Proposition C-2.60 to complete the proof of Burnside's Theorem.

**Theorem C-2.106.**

(i) *If $G$ is a nonabelian finite simple group, then $\{1\}$ is the only conjugacy class whose size is a prime power.*

(ii) *Burnside's Theorem is true: every group of order $p^m q^n$, where $p$ and $q$ are primes, is solvable.*

**Proof.**

(i) Assume, on the contrary, that $h_j = p^e > 1$ for some $j$. By Exercise C-2.36 on page 199, for all $i$, we have

$$Z(G/\ker\chi_i) = \{g \in G : |\chi_i(g)| = n_i\}.$$

Since $G$ is simple, $\ker \chi_i = \{1\}$ for all $i$, and so $Z(G/\ker \chi_i) = Z(G) = \{1\}$. By Proposition C-2.105, if $(n_i, h_j) = 1$, then either $|\chi_i(g_j)| = n_i$ or $\chi_i(g_j) = 0$. Of course, $\chi_1(g_j) = 1$ for all $j$, where $\chi_1$ is the trivial character. If $\chi_i$ is not the trivial character, then we have just seen that the first possibility cannot occur, and so $\chi_i(g_j) = 0$. On the other hand, if $(n_i, h_j) \neq 1$, then $p \mid n_i$ (for $h_j = p^e$). Thus, for every $i \neq 1$, either $\chi_i(g_j) = 0$ or $p \mid n_i$.

Consider the orthogonality relation Corollary C-2.79(ii):

$$\sum_{i=1}^{r} n_i \chi_i(g_j) = 0.$$

Now $n_1 = 1 = \chi_1(g_j)$, while each of the other terms is either 0 or of the form $p\alpha_i$, where $\alpha_i$ is an algebraic integer. It follows that

$$0 = 1 + p\beta,$$

where $\beta$ is an algebraic integer. This implies that the rational number $-1/p$ is an algebraic integer, hence lies in $\mathbb{Z}$, and we have the contradiction that $-1/p$ is an integer.

(ii) Proposition C-2.60. •

Another early application of characters is a theorem of Frobenius. We begin by recalling *doubly transitive* permutation groups. Let $G$ be a finite group and $X$ a finite $G$-set. Recall that if $x \in X$, then its *orbit* is $\mathcal{O}(x) = \{gx : g \in G\}$ and its *stabilizer* is $G_x = \{g \in G : gx = x\}$. Theorem C-1.16 shows that $|\mathcal{O}(x)||G_x| = |G|$. A $G$-set $X$ is *transitive* if it has only one orbit: if $x, y \in X$, then there exists $g \in G$ with $y = gx$; in this case, $\mathcal{O}(x) = X$. We will also mention the stabilizer of two points: $G_{x,y} = \{g \in G : gx = x \text{ and } gy = y\}$.

**Definition.** A transitive $G$-set $X$ is called ***regular*** if only the identity element of $G$ fixes any element of $X$; that is, $G_x = \{1\}$ for all $x \in X$.

For example, Cayley's Theorem shows that every group $G$ is isomorphic to a regular subgroup of $S_G$ (indeed, this is why the *regular representation* is so called).

**Notation.** If $G$ is a group, then $G^{\#} = \{g \in G : g \neq 1\}$.

We now consider transitive groups $G$ such that each $g \in G^{\#}$ has at most one fixed point.[12] In case every $g \in G^{\#}$ has no fixed points, we say that the action of $G$ is ***fixed-point-free***. Thompson proved that if a finite group $H$ has a fixed-point-free automorphism $\alpha$ of prime order (that is, the action of the group $G = \langle \alpha \rangle$ on $H^{\#}$ is fixed-point-free), then $H$ is nilpotent (Robinson [**181**], pp. 306–307). Thus, let us consider such actions in which there is some $g \in G^{\#}$ that has a fixed point; that is, the action of $G$ is not regular.

[12] Proving general theorems about groups $G$ by first normalizing is a common feature of group theory (and of mathematics!). We first prove special cases by assuming extra properties of $G$ until only the difficult case remains. This strategy often leaves groups satisfying what appears to be unnatural hypotheses; and if these hypotheses occur in other situations, we create a definition that also seems unnatural. Such is the case with Frobenius groups, for example.

**Definition.** A finite group $G$ is a ***Frobenius group*** if there exists a transitive $G$-set $X$ such that

(i) every $g \in G^{\#}$ has at most one fixed point;

(ii) there is some $g \in G^{\#}$ that does have a fixed point.

If $x \in X$, we call $G_x$ a ***Frobenius complement*** of $G$.

Note that condition (i) implies that the $G$-set $X$ in the definition is necessarily faithful. Let us rephrase the two conditions: (i) if every $g \in G^{\#}$ has at most one fixed point, then the stabilizer of two points $G_{x,y} = \{1\}$; (ii) if there is some $g \in G^{\#}$ that does have a fixed point, then the stabilizer of one point $G_x \neq \{1\}$.

**Example C-2.107.**

(i) The symmetric group $S_3$ is a Frobenius group: $X = \{1, 2, 3\}$ is a faithful transitive $S_3$-set; no $\alpha \in (S_3)^{\#}$ fixes two elements; each transposition $(i\ j)$ fixes one element. The cyclic subgroups $\langle (i\ j) \rangle$ are Frobenius complements (so Frobenius complements need not be unique). A permutation $\beta \in S_3$ has no fixed points if and only if $\beta$ is a 3-cycle. We are going to prove, in every Frobenius group, that 1 together with all those elements having no fixed points comprise a normal subgroup.

(ii) The example of $S_3$ in (i) can be generalized. Let $X$ be a $G$-set, with at least three elements, which is a *sharply doubly transitive* $G$-set; that is, $X$ is a doubly transitive $G$-set and $G_{x,y} = \{1\}$ for every pair of $x \neq y$. Thus, $X$ is transitive and $G_x \neq \{1\}$ (for if $x, y, z \in X$ are distinct, there exists $g \in G$ with $x = gx$ and $z = gy$). Therefore, every sharply doubly transitive group $G$ is a Frobenius group. ◀

The significance of the next proposition is that it translates the definition of Frobenius group from the language of $G$-sets into the language of abstract groups.

**Proposition C-2.108.** *A finite group $G$ is a Frobenius group if and only if it contains a proper nontrivial subgroup $H$ such that $H \cap gHg^{-1} = \{1\}$ for all $g \notin H$.*[13]

**Proof.** Let $X$ be a $G$-set as in the definition of Frobenius group. Choose $x \in X$, and define $H = G_x$. Now $H$ is a proper subgroup of $G$, for transitivity does not permit $gx = x$ for all $g \in G$. To see that $H$ is nontrivial, choose $g \in G^{\#}$ having a fixed point; say, $gy = y$. If $y = x$, then $g \in G_x = H$. If $y \neq x$, then transitivity provides $h \in G$ with $hy = x$, and Exercise C-1.24 on page 16 gives $H = G_x = hG_yh^{-1} \neq \{1\}$. If $g \notin H$, then $gx \neq x$. Now $g(G_x)g^{-1} = G_{gx}$. Hence, if $h \in H \cap gHg^{-1} = G_x \cap G_{gx}$, then $h$ fixes $x$ and $gx$; that is, $h \in G_{x,y} = \{1\}$.

For the converse, we take $X$ to be the $G$-set $G/H$ of all left cosets of $H$ in $G$, where $g\colon aH \mapsto gaH$ for all $g \in G$. Example C-1.7(i) says that $X$ is a transitive $G$-set and that the stabilizer of each $aH \in G/H$ is the subgroup $aHa^{-1}$ of $G$. Since

[13] A subset $H$ of a finite $G$-set $X$ is called a ***T.I. set*** (*trivial intersection set*) if, for all $g \in G$, either $H \cap gHg^{-1} = X$ or $H \cap gHg^{-1} \subseteq \{1\}$. Thus, if $G$ acts on itself by conjugation, then the subgroup $H$ is a T.I. set. See Exercise C-2.47.

$H \neq \{1\}$, we see that $G_{aH} \neq \{1\}$. Finally, if $aH \neq bH$, then

$$G_{aH,bH} = G_{aH} \cap G_{bH} = aHa^{-1} \cap bHb^{-1} = a\big(H \cap a^{-1}bHb^{-1}a\big)a^{-1} = \{1\},$$

because $a^{-1}b \notin H$. Therefore, $G$ is a Frobenius group. •

**Definition.** Let $X$ be a $G$-set. The ***Frobenius kernel*** of $G$ is the subset

$$N = \{1\} \cup \{g \in G : g \text{ has no fixed points}\}.$$

When $X$ is transitive, we can describe $N$ in terms of a stabilizer $G_x$. If $a \notin N^{\#}$, then there is some $y \in X$ with $ay = y$. Since $G$ acts transitively, there is $g \in G$ with $gx = y$, and $a \in G_y = gG_xg^{-1}$. Hence, $a \in \bigcup_{g\in G} gG_xg^{-1}$. For the reverse inclusion, if $a \in \bigcup_{g\in G} gG_xg^{-1}$, then $a \in gG_xg^{-1} = G_{gx}$ for some $g \in G$, and so $a$ has a fixed point; that is, $a \notin N$. We have proved that

$$N = \{1\} \cup \Big(G - \Big(\bigcup_{g\in G} gG_xg^{-1}\Big)\Big).$$

Exercise C-1.11 on page 14 shows that if $G_x$ is a proper subgroup of $G$, then $G \neq \bigcup_{g\in G} gG_xg^{-1}$, and so $N \neq \{1\}$ in this case.

**Proposition C-2.109.** *If $G$ is a Frobenius group with Frobenius complement $H$ and Frobenius kernel $N$, then $|N| = [G : H]$.*

**Proof.** By Proposition C-2.108, there is a disjoint union

$$G = \{1\} \cup \Big(\bigcup_{g\in G} gH^{\#}g^{-1}\Big) \cup N^{\#}.$$

Note that $N_G(H) = H$: if $g \notin H$, then $H \cap gHg^{-1} = \{1\}$, and so $gHg^{-1} \neq H$. Hence, the number of conjugates of $H$ is $[G : N_G(H)] = [G : H]$ (Corollary C-1.19). Therefore, $|\bigcup_{g\in G} gH^{\#}g^{-1}| = [G : H](|H| - 1)$, and so

$$|N| = |N^{\#}| + 1 = |G| - ([G : H](|H| - 1)) = [G : H]. \quad \bullet$$

The Frobenius kernel may not be a subgroup of $G$. It is very easy to check that if $g \in N$, then $g^{-1} \in N$ and $aga^{-1} \in N$ for every $a \in G$; the difficulty is in proving that $N$ is closed under multiplication. For example, if $V = k^n$ is the vector space of all $n \times 1$ column vectors over a field $k$, then $V^{\#}$, the set of nonzero vectors in $V$, is a faithful transitive $\mathrm{GL}(V)$-set. Now $A \in \mathrm{GL}(V)$ has a fixed point if and only if there is some $v \in V^{\#}$ with $Av = v$; that is, $A$ has a fixed point if and only if 1 is an eigenvalue of $A$. Thus, the Frobenius kernel now consists of the identity matrix together with all linear transformations which do not have 1 as an eigenvalue. Let $|k| \geq 4$, and let $\alpha$ be a nonzero element of $k$ with $\alpha^2 \neq 1$. Then $A = \left[\begin{smallmatrix} \alpha & 0 \\ 0 & \alpha \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & \alpha \end{smallmatrix}\right]$ lie in $N$, but their product $AB = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^2 \end{smallmatrix}\right]$ does not lie in $N$. However, if $G$ is a Frobenius group, then $N$ is a subgroup; the only known proof of this fact uses characters.

We have already remarked that if $\psi$ is a character on a subgroup $H$ of a group $G$, then the restriction $(\psi 1^G)_H$ of the induced character $\psi 1^G$ need not equal $\psi$. The next proof shows that irreducible characters of a Frobenius complement do extend to irreducible characters of $G$.

**Lemma C-2.110.** *Let $G$ be a Frobenius group with Frobenius complement $H$ and Frobenius kernel $N$. For every irreducible character $\psi$ on $H$ other than the trivial character $\psi_1$, define the generalized character*

$$\varphi = \psi - d\psi_1,$$

*where $d = \psi(1)$. Then $\psi^* = \varphi\uparrow^G + d\chi_1$ is an irreducible character on $G$, and $\psi^*_H = \psi$; that is, $\psi^*(h) = \psi(h)$ for all $h \in H$.*

**Proof.** Note first that $\varphi(1) = 0$. We claim that the induced generalized character $\varphi\uparrow^G$ satisfies the equation

$$(\varphi\uparrow^G)_H = \varphi.$$

If $t_1 = 1, \ldots, t_n$ is a transversal of $H$ in $G$, then for $g \in G$, the matrix of $\varphi\uparrow^G(g)$ on page 189 has the blocks $\dot{B}(t_i^{-1}gt_i)$ on its diagonal, where $\dot{B}(t_i^{-1}gt_i) = 0$ if $t_i^{-1}gt_i \notin H$ (this is just the matrix version of Theorem C-2.85). If $h \in H$, then $t_i^{-1}ht_i \notin H$ for all $i \neq 1$, and so $\dot{B}(t_i^{-1}ht_i) = 0$. Therefore, there is only one nonzero diagonal block, and

$$\operatorname{tr}(\varphi\uparrow^G(h)) = \operatorname{tr}(B(h));$$

that is,

$$\varphi\uparrow^G(h) = \varphi(h).$$

We have just seen that $\varphi\uparrow^G$ is a generalized character on $G$ such that $(\varphi\uparrow^G)_H = \varphi$. By Frobenius Reciprocity,

$$(\varphi\uparrow^G, \varphi\uparrow^G)_G = (\varphi, (\varphi\uparrow^G)_H)_H = (\varphi, \varphi)_H.$$

But $\varphi = \psi - d\psi_1$, so that orthogonality of $\psi$ and $\psi_1$ gives

$$(\varphi, \varphi)_H = 1 + d^2.$$

Similarly,

$$(\varphi\uparrow^G, \chi_1)_G = (\varphi, \psi_1)_H = -d,$$

where $\chi_1$ is the trivial character on $G$. Define

$$\psi^* = \varphi\uparrow^G + d\chi_1.$$

Now $\psi^*$ is a generalized character on $G$, and

$$(\psi^*, \psi^*)_G = (\varphi\uparrow^G, \varphi\uparrow^G)_G + 2d(\varphi\uparrow^G, \chi_1)_G + d^2 = 1 + d^2 - 2d^2 + d^2 = 1.$$

We have

$$(\psi^*)_H = (\varphi\uparrow^G)_H + d(\chi_1)_H = \varphi + d\psi_1 = (\psi - d\psi_1) + d\psi_1 = \psi.$$

Since $\psi^*(1) = \psi(1) > 0$, Corollary C-2.75 says that $\psi^*$ is an irreducible character on $G$. •

**Theorem C-2.111 (Frobenius).** *Let $G$ be a Frobenius group with Frobenius complement $H$ and Frobenius kernel $N$. Then $N$ is a normal subgroup of $G$, $N \cap H = \{1\}$, and $NH = G$.*

**Remark.** A group $G$ having a subgroup $Q$ and a normal subgroup $K$ such that $K \cap Q = \{1\}$ and $KQ = G$ is called a *semidirect product*. We will discuss such groups in the chapter on homology. ◄

**Proof.** For every irreducible character $\psi$ on $H$ other than the trivial character $\psi_1$, define the generalized character $\varphi = \psi - d\psi_1$, where $d = \psi(1)$. By Lemma C-2.110, $\psi^* = \varphi\uparrow^G + d\chi_1$ is an irreducible character on $G$. Define

$$N^* = \bigcap_{\psi \neq \psi_1} \ker \psi^*.$$

Of course, $N^*$ is a normal subgroup of $G$.

By Lemma C-2.110, $\psi^*(h) = \psi(h)$ for all $h \in H$; in particular, if $h = 1$, we have

$$\psi^*(1) = \psi(1) = d. \tag{1}$$

If $g \in N^{\#}$, then for all $a \in G$, we have $g \notin aHa^{-1}$ (for $g$ has no fixed points), and so $\dot{\varphi}(aga^{-1}) = 0$. The induced character formula, Theorem C-2.85, now gives $\varphi\uparrow^G(g) = 0$. Hence, if $g \in N^{\#}$, then Eq. (1) gives

$$\psi^*(g) = \varphi\uparrow^G(g) + d\chi_1(g) = d.$$

We conclude that if $g \in N$, then

$$\psi^*(g) = d = \psi^*(1);$$

that is, $g \in \ker \psi^*$. Therefore,

$$N \subseteq N^*.$$

The reverse inclusion will arise from a counting argument.

Let $h \in H \cap N^*$. Since $h \in H$, Lemma C-2.110 gives $\psi^*(h) = \psi(h)$. On the other hand, since $h \in N^*$, we have $\psi^*(h) = \psi^*(1) = d$. Therefore, $\psi(h) = \psi^*(h) = d = \psi(1)$, so that $h \in \ker \psi$ for every irreducible character $\psi$ on $H$. Consider the regular character, afforded by the regular representation $\rho$ on $H$: $\chi_\rho = \sum_i n_i\psi_i$. Now $\chi_\rho(h) = \sum_i n_i\psi_i(h) \neq 0$, so that Example C-2.70(ii) gives $h = 1$. Thus,

$$H \cap N^* = \{1\}.$$

Next, $|G| = |H|[G : H] = |H||N|$, by Proposition C-2.109. Note that $HN^*$ is a subgroup of $G$, because $N^* \lhd G$. Now $|HN^*||H \cap N^*| = |H||N^*|$, by the Second Isomorphism Theorem; since $H \cap N^* = \{1\}$, we have $|H||N| = |G| \geq |HN^*| = |H||N^*|$. Hence, $|N| \geq |N^*|$. But $|N| \leq |N^*|$, because $N \subseteq N^*$, and so $N = N^*$. Therefore, $N \lhd G$, $H \cap N = \{1\}$, and $HN = G$. •

Much more can be said about the structure of Frobenius groups. Every Sylow subgroup of a Frobenius complement is either cyclic or generalized quaternion (Huppert [**103**], p. 502), and it is a consequence of Thompson's Theorem on fixed-point-free automorphisms that every Frobenius kernel is nilpotent (Robinson [**181**], p. 306); that is, $N$ is the direct product of its Sylow subgroups. The reader is referred to Curtis–Reiner [**48**], pp. 242–246, or Feit [**65**], pp. 133–139.

## Exercises

**C-2.44.** Prove that the affine group $\mathrm{Aff}(1, \mathbb{F}_q)$ is sharply doubly transitive.

**C-2.45.** Assume that the family of left cosets $G/H$ of a subgroup $H \subseteq G$ is a $G$-set via the representation on cosets. Prove that $G/H$ is a faithful $G$-set if and only if $\bigcap_{a \in G} aHa^{-1} = \{1\}$. Give an example in which $G/H$ is not a faithful $G$-set.

**C-2.46.** Prove that every Sylow subgroup of $\mathrm{SL}(2, \mathbb{F}_5)$ is either cyclic or quaternion.

**C-2.47.** A subset $A$ of a group $G$ is a **T.I. set** (***trivial intersection set***) if $A \subseteq N_G(A)$ and $A \cap gAg^{-1} \subseteq \{1\}$ for all $g \notin N_G(A)$.

(i) Prove that a Frobenius complement $H$ in a Frobenius group $G$ is a T.I. set.

(ii) Let $A$ be a T.I. set in a finite group $G$, and let $N = N_G(A)$. If $\alpha$ is a class function vanishing on $N - A$ and $\beta$ is a class function on $N$ vanishing on $\big(\bigcup_{g \in G}(A^g \cap N)\big) - A$, prove, for all $g \in N^{\#}$, that $\alpha{\uparrow}^G(g) = \alpha(g)$ and $\beta{\uparrow}^G(g) = \beta(g)$.
**Hint.** See the proofs of Theorem C-2.111 and Lemma C-2.110.

(iii) If $\alpha(1) = 0$, prove that $(\alpha, \beta)_N = (\alpha{\uparrow}^G, \beta{\uparrow}^G)_G$.

(iv) Let $H$ be a *self-normalizing* subgroup of a finite group $G$; that is, $H = N_G(H)$. If $H$ is a T.I. set, prove that there is a normal subgroup $K$ of $G$ with $K \cap H = \{1\}$ and $KH = G$.
**Hint.** See Feit [**65**], p. 124.

∗ **C-2.48.** Prove that there are no nonabelian simple groups of order $n$, where $60 < n \leq 100$.

**Hint.** By Burnside's Theorem, the only candidates for $n$ in the given range are 66, 70, 78, 84, and 90; note that 90 was eliminated in Exercise C-1.46 on page 33.

**C-2.49.** Prove that there are no nonabelian simple groups of order $n$, where $101 \leq n < 168$. We remark that $\mathrm{PSL}(2, \mathbb{F}_7)$ is a simple group of order 168, and it is the unique such group up to isomorphism. In view of Proposition C-1.40, Corollary C-1.81, and Exercise C-2.48, we see that $A_5$ is the only nonabelian simple group of order strictly less than 168.

**Hint.** By Burnside's Theorem, the only candidates for $n$ in the given range are 102, 105, 110, 120, 126, 130, 132, 138, 140, 150, 154, 156, and 165. Use Exercise C-1.23 on page 16 and Exercise C-1.47 on page 33.

## C-2.13. Division Algebras

When applying the Wedderburn–Artin Theorems to group algebras $kG$, where $k$ is algebraically closed, we used Molien's Theorem (Corollary C-2.44) to assume that the matrix rings have entries in $k$. If $k$ is not algebraically closed, then (noncommutative) division rings can occur. At the moment, the only example we know of such a ring is the quaternions $\mathbb{H}$ (or certain subalgebras of $\mathbb{H}$; see Example B-1.1(x) in Part 1), and it is not at all obvious how to construct other examples.

Linear representations of a finite group over a field $k$ are the simplest ones, for every finite subgroup of the multiplicative group $k^{\times}$ of nonzero elements is cyclic (Theorem A-3.59 in Part 1). Herstein proved that every finite subgroup of $D^{\times}$ is

cyclic if $D$ is a division ring whose center is a field of characteristic $p > 0$, but it is false when $Z(D)$ has characteristic 0 (obviously, the group of quaternions is a subgroup of $\mathbb{H}^\times$). All finite subgroups of multiplicative groups of division rings were found by Amitsur [**4**].

That the tensor product of algebras is, again, an algebra, is used in the study of division rings.

**Definition.** A ***division algebra over a field*** $k$ is a division ring regarded as an algebra over its center $k$.

Let us begin by considering the wider class of simple algebras.

**Definition.** A finite-dimensional[14] $k$-algebra $A$ over a field $k$ is ***central simple*** if it is simple (no two-sided ideals other than $A$ and $\{0\}$) and its center $Z(A) = k$.

**Notation.** If $A$ is an algebra over a field $k$, then we write

$$[A : k] = \dim_k(A).$$

**Example C-2.112.**

(i) Every division algebra $k$ that is finite-dimensional over its center $k$ is a central simple $k$-algebra. The ring $\mathbb{H}$ of quaternions is a central simple $\mathbb{R}$-algebra, and every field is a central simple algebra over itself.

(ii) If $k$ is a field, then $\mathrm{Mat}_n(k)$ is a central simple $k$-algebra (it is simple, by Proposition C-2.29, and its center consists of all scalar matrices $\{aI : a \in k\}$, by Exercise B-1.8 on page 281 in Part 1).

(iii) If $A$ is a central simple $k$-algebra, then its opposite algebra $A^{\mathrm{op}}$ is also a central simple $k$-algebra. ◀

**Theorem C-2.113.** *Let $A$ be a central simple $k$-algebra. If $B$ is a simple $k$-algebra, then $A \otimes_k B$ is a central simple $Z(B)$-algebra. In particular, if $B$ is a central simple $k$-algebra, then $A \otimes_k B$ is a central simple $k$-algebra.*

**Proof.** Each $x \in A \otimes_k B$ has an expression of the form

$$x = a_1 \otimes b_1 + \cdots + a_n \otimes b_n, \tag{1}$$

where $a_i \in A$ and $b_i \in B$. For nonzero $x$, define the *length* of $x$ to be $n$ if there is no such expression having fewer than $n$ terms. We claim that if $x$ has length $n$, that is, if Eq. (1) is a shortest such expression, then $b_1, \ldots, b_n$ is a linearly independent list in $B$ (viewed as a vector space over $k$). Otherwise, there is some $j$ and $u_i \in k$, not all zero, with $b_j = \sum_i u_i b_i$. Substituting and collecting terms gives

$$x = \sum_{i \neq j} (a_i + u_i a_j) \otimes b_i,$$

which is a shorter expression for $x$.

Let $I \neq (0)$ be a two-sided ideal in $A \otimes_k B$. Choose $x$ to be a (nonzero) element in $I$ of smallest length, and assume that Eq. (1) is a shortest expression for $x$. Now

[14] We assume that central simple algebras are finite-dimensional, but some authors do not. Hilbert gave an example of an infinite-dimensional division algebra (Drozd–Kirichenko [**57**], p. 81).

$a_1 \neq 0$. Since $Aa_1A$ is a two-sided ideal in $A$, simplicity gives $A = Aa_1A$. Hence, there are elements $a'_p$ and $a''_p$ in $A$ with $1 = \sum_p a'_p a_1 a''_p$. Since $I$ is a two-sided ideal,

$$x' = \sum_p a'_p x a''_p = 1 \otimes b_1 + c_2 \otimes b_2 + \cdots + c_n \otimes b_n \tag{2}$$

lies in $I$, where, for $i \geq 2$, we have $c_i = \sum_p a'_p a_i a''_p$. At this stage, we do not know whether $x' \neq 0$, but we do know, for every $a \in A$, that $(a \otimes 1)x' - x'(a \otimes 1) \in I$. Now

$$(a \otimes 1)x' - x'(a \otimes 1) = \sum_{i \geq 2} (ac_i - c_i a) \otimes b_i. \tag{3}$$

First, this element is 0, lest it be an element in $I$ of length smaller than the length of $x$. Since $b_1, \ldots, b_n$ is a linearly independent list, the $k$-subspace it generates is $\langle b_1, \ldots, b_n \rangle = \langle b_1 \rangle \oplus \cdots \oplus \langle b_n \rangle$, and so

$$A \otimes_k \langle b_1, \ldots, b_n \rangle = A \otimes_k \langle b_1 \rangle \oplus \cdots \oplus A \otimes_k \langle b_n \rangle.$$

It follows from Eq. (3) that each term $(ac_i - c_i a) \otimes b_i$ must be 0. Hence, $ac_i = c_i a$ for all $a \in A$; that is, each $c_i \in Z(A) = k$. Eq. (2) becomes

$$\begin{aligned} x' &= 1 \otimes b_1 + c_2 \otimes b_2 + \cdots + c_n \otimes b_n \\ &= 1 \otimes b_1 + 1 \otimes c_2 b_2 + \cdots + 1 \otimes c_n b_n \\ &= 1 \otimes (b_1 + c_2 b_2 + \cdots + c_n b_n). \end{aligned}$$

Now $b_1 + c_2 b_2 + \cdots + c_n b_n \neq 0$, because $b_1, \ldots, b_n$ is a linearly independent list, and so $x' \neq 0$. Therefore, $I$ contains a nonzero element of the form $1 \otimes b$. But simplicity of $B$ gives $BbB = B$, and so there are $b'_q, b''_q \in B$ with $\sum_q b'_q b b''_q = 1$. Hence, $I$ contains $\sum_q (1 \otimes b'_q)(1 \otimes b)(1 \otimes b''_q) = 1 \otimes 1$, which is the unit in $A \otimes_k B$. Therefore, $I = A \otimes_k B$ and $A \otimes_k B$ is simple.

We now seek the center of $A \otimes_k B$. Clearly, $k \otimes_k Z(B) \subseteq Z(A \otimes_k B)$. For the reverse inequality, let $z \in Z(A \otimes_k B)$ be nonzero, and let

$$z = a_1 \otimes b_1 + \cdots + a_n \otimes b_n$$

be a shortest such expression for $z$. As in the preceding argument, $b_1, \ldots, b_n$ is a linearly independent list over $k$. For each $a \in A$, we have

$$0 = (a \otimes 1)z - z(a \otimes 1) = \sum_i (aa_i - a_i a) \otimes b_i.$$

It follows, as above, that $(aa_i - a_i a) \otimes b_i = 0$ for each $i$. Hence, $aa_i - a_i a = 0$, so that $aa_i = a_i a$ for all $a \in A$ and each $a_i \in Z(A) = k$. Thus, $z = 1 \otimes x$, where $x = a_1 b_1 + \cdots + a_n b_n \in B$. But if $b \in B$, then

$$0 = z(1 \otimes b) - (1 \otimes b)z = (1 \otimes x)(1 \otimes b) - (1 \otimes b)(1 \otimes x) = 1 \otimes (xb - bx).$$

Therefore, $xb - bx = 0$ and $x \in Z(B)$. We conclude that $z \in k \otimes_k Z(B)$. •

It is not generally true that the tensor product of simple $k$-algebras is again simple; we must pay attention to the centers. Exercise C-2.54 on page 221 shows that if $E/k$ is a field extension, then $E \otimes_k E$ need not be a field. The tensor product of division algebras need not be a division algebra, as we see in the next example.

**Example C-2.114.** The algebra $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H}$ is an eight-dimensional $\mathbb{R}$-algebra, but it is also a four-dimensional $\mathbb{C}$-algebra with basis

$$1 = 1 \otimes 1,\ 1 \otimes i,\ 1 \otimes j,\ 1 \otimes k.$$

We let the reader prove that the vector space isomorphism $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} \to \mathrm{Mat}_2(\mathbb{C})$ with

$$1 \otimes 1 \mapsto \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad 1 \otimes i \mapsto \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad 1 \otimes j \mapsto \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad 1 \otimes k \mapsto \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

is an isomorphism of $\mathbb{C}$-algebras. It follows that $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H}$, though a simple algebra, is not a division ring. ◀

Another way to see that $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathrm{Mat}_2(\mathbb{C})$ arises from Example C-2.52(ii). We remarked then that

$$\mathbb{R}\mathbf{Q} \cong \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{H};$$

tensoring by $\mathbb{C}$ gives

$$\mathbb{C}\mathbf{Q} \cong \mathbb{C} \otimes_{\mathbb{R}} \mathbb{R}\mathbf{Q} \cong \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times (\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H}).$$

It follows from the uniqueness in Wedderburn–Artin Theorem II that $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathrm{Mat}_2(\mathbb{C})$ (we give yet another proof of this in the next theorem).

The next theorem puts the existence of the isomorphism in Example C-2.114 into the context of central simple algebras.

**Theorem C-2.115.** *Let $k$ be a field and let $A$ be a central simple $k$-algebra.*

(i) *If $\overline{k}$ is the algebraic closure of $k$, then there is an integer $n$ with*

$$\overline{k} \otimes_k A \cong \mathrm{Mat}_n(\overline{k}).$$

*In particular, $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathrm{Mat}_2(\mathbb{C})$.*

(ii) *If $A$ is a central simple $k$-algebra, then there is an integer $n$ with*

$$[A : k] = n^2.$$

**Proof.**

(i) By Theorem C-2.113, $\overline{k} \otimes_k A$ is a simple $\overline{k}$-algebra. Hence, Wedderburn's Theorem (actually, Corollary C-2.42) gives $\overline{k} \otimes_k A \cong \mathrm{Mat}_n(D)$ for some $n \geq 1$ and some division ring $D$. Since $D$ is a finite-dimensional division algebra over $\overline{k}$, the argument in Molien's Theorem (Corollary C-2.44) shows that $D = \overline{k}$. In particular, since $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$, we have $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathrm{Mat}_n(\mathbb{C})$ for some $n$; as $\dim_{\mathbb{R}}(\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H}) = 8$, we have $n = 2$.

(ii) We claim that $[A : k] = [\overline{k} \otimes_k A : \overline{k}]$, for if $a_1, \ldots, a_m$ is a basis of $A$ over $k$, then $1 \otimes a_1, \ldots, 1 \otimes a_m$ is a basis of $\overline{k} \otimes_k A$ over $\overline{k}$ (essentially because tensor product commutes with direct sum). Therefore,

$$[A : k] = [\overline{k} \otimes_k A : \overline{k}] = [\mathrm{Mat}_n(\overline{k}) : \overline{k}] = n^2. \quad \bullet$$

**Definition.** A ***splitting field*** for a central simple $k$-algebra $A$ is a field extension $E/k$ for which there exists an integer $n$ such that $E \otimes_k A \cong \mathrm{Mat}_n(E)$.

Theorem C-2.115 says that the algebraic closure $\overline{k}$ of a field $k$ is a splitting field for every central simple $k$-algebra $A$. We are going to see that there always exists a splitting field that is a finite extension of $k$, but we first develop some tools in order to prove it.

**Definition.** If $A$ is a $k$-algebra and $X \subseteq A$ is a subset, then its ***centralizer***, $C_A(X)$, is defined by

$$C_A(X) = \{a \in A : ax = xa \text{ for every } x \in X\}.$$

It is easy to check that centralizers are always subalgebras.

The key idea in the next proof is that a subalgebra $B$ of $A$ makes $A$ into a $(B, A)$-bimodule and that the centralizer of $B$ can be described in terms of an endomorphism ring.

**Theorem C-2.116 (Double Centralizer).** *Let $A$ be a central simple algebra over a field $k$ and let $B$ be a simple subalgebra of $A$.*

(i) *$C_A(B)$ is a simple $k$-algebra.*

(ii) *$B \otimes_k A^{\mathrm{op}} \cong \mathrm{Mat}_s(D)$ and $C_A(B) \cong \mathrm{Mat}_r(D)$ for some division algebra $D$, where $r \mid s$.*

(iii) *$[B : k][C_A(B) : k] = [A : k]$.*

(iv) *$C_A(C_A(B)) = B$.*

**Proof.** Associativity of the multiplication in $A$ shows that $A$ can be viewed as a $(B, A)$-bimodule. As such, it is a left $(B \otimes_k A^{\mathrm{op}})$-module (Proposition B-5.9 in Part 1), where $(b \otimes a)x = bxa$ for all $x \in A$; we denote this module by $A^*$. But $B \otimes_k A^{\mathrm{op}}$ is a simple $k$-algebra, by Theorem C-2.113, so that Corollary C-2.42 gives $B \otimes_k A^{\mathrm{op}} \cong \mathrm{Mat}_s(D)$ for some integer $s$ and some division algebra $D$ over $k$; in fact, $B \otimes_k A^{\mathrm{op}}$ has a unique (up to isomorphism) minimal left ideal $L$, and $D^{\mathrm{op}} \cong \mathrm{End}_{B \otimes_k A^{\mathrm{op}}}(L)$. Therefore, as $(B \otimes_k A^{\mathrm{op}})$-modules, Corollary C-2.18 gives $A^* \cong L^r = L \oplus \cdots \oplus L$, the direct sum of $r$ copies of $L$, and so $\mathrm{End}_{B \otimes_k A^{\mathrm{op}}}(A^*) \cong \mathrm{Mat}_r(D)$.

We claim that

$$C_A(B) \cong \mathrm{End}_{B \otimes_k A^{\mathrm{op}}}(A^*) \cong \mathrm{Mat}_r(D);$$

this will prove (i) and most of (ii). If $\varphi \in \mathrm{End}_{B \otimes_k A^{\mathrm{op}}}(A^*)$, then it is, in particular, an endomorphism of $A$ as a right $A$-module. Hence, for all $a \in A$, we have

$$\varphi(a) = \varphi(1a) = \varphi(1)a = ua,$$

where $u = \varphi(1)$. In particular, if $b \in B$, then $\varphi(b) = ub$. On the other hand, taking the left action of $B$ into account, we have $\varphi(b) = \varphi(b1) = b\varphi(1) = bu$. Therefore, $ub = bu$ for all $b \in B$, and so $u \in C_A(B)$. Thus, $\varphi \mapsto \varphi(1)$ is a function $\mathrm{End}_{B \otimes_k A^{\mathrm{op}}}(A^*) \to C_A(B)$. It is routine to check that this function is an injective $k$-algebra map; it is also surjective, for if $u \in C_A(B)$, then the map $A \to A$, defined by $a \mapsto ua$, is a $(B \otimes_k A^{\mathrm{op}})$-map.

We now compute dimensions. Define $d = [D : k]$. Since $L$ is a minimal left ideal in $\mathrm{Mat}_s(D)$, we have $\mathrm{Mat}_s(D) \cong L^s$ (concretely, $L = \mathrm{COL}(1)$, all the first

columns of $s \times s$ matrices over $D$). Therefore, $[\mathrm{Mat}_s(D) : k] = s^2[D : k]$ and $[L^s : k] = s[L : k]$, so that

$$[L : k] = sd.$$

Also,

$$[A : k] = [A^* : k] = [L^r : k] = rsd.$$

It follows that

$$[A : k][B : k] = [B \otimes_k A^{\mathrm{op}} : k] = [\mathrm{Mat}_s(D) : k] = s^2 d.$$

Therefore, $[B : k] = s^2 d/rsd = s/r$, and so $r \mid s$. Hence,

$$[B : k][C_A(B) : k] = [B : k][\mathrm{Mat}_r(D) : k] = \frac{s}{r} \cdot r^2 d = rsd = [A : k],$$

because we have already proved that $C_A(B) \cong \mathrm{Mat}_r(D)$.

Finally, we prove (iv). It is easy to see that $B \subseteq C_A(C_A(B))$: after all, if $b \in B$ and $u \in C_A(B)$, then $bu = ub$, and so $b$ commutes with every such $u$. But $C_A(B)$ is a simple subalgebra, by (i), and so the equation in (iii) holds if we replace $B$ by $C_A(B)$:

$$[C_A(B) : k][C_A(C_A(B)) : k] = [A : k].$$

We conclude that $[B : k] = [C_A(C_A(B)) : k]$; together with $B \subseteq C_A(C_A(B))$, this equality gives $B = C_A(C_A(B))$. •

Here is a minor variant of the theorem.

**Corollary C-2.117.** *If $B$ is a simple subalgebra of a central simple $k$-algebra $A$, where $k$ is a field, then there is a division algebra $D_1$ with $B^{\mathrm{op}} \otimes_k A \cong \mathrm{Mat}_s(D_1)$.*

**Proof.** By Theorem C-2.116(ii), we have $B \otimes_k A^{\mathrm{op}} \cong \mathrm{Mat}_s(D)$ for some division algebra $D$. Hence, $(B \otimes_k A^{\mathrm{op}})^{\mathrm{op}} \cong (\mathrm{Mat}_s(D))^{\mathrm{op}}$. But $(\mathrm{Mat}_s(D))^{\mathrm{op}} \cong \mathrm{Mat}_s(D^{\mathrm{op}})$, by Proposition B-1.25 in Part 1, while $(B \otimes_k A^{\mathrm{op}})^{\mathrm{op}} \cong B^{\mathrm{op}} \otimes_k A$, by Exercise C-2.52 on page 220. Setting $D_1 = D^{\mathrm{op}}$ completes the proof. •

If $D$ is a division algebra over a field $k$ and $\delta \in D$, then the subdivision algebra generated by $k$ and $\delta$ is a field, because elements in the center $k$ commute with $\delta$. We are interested in maximal subfields of $D$.

**Lemma C-2.118.** *If $D$ is a division algebra over a field $k$, then a subfield $E$ of $D$ is a maximal subfield if and only if $C_D(E) = E$.*

**Proof.** If $E$ is a maximal subfield of $D$, then $E \subseteq C_k(E)$ because $E$ is commutative. For the reverse inclusion, it is easy to see that if $\delta \in C_k(E)$, then the division algebra $E'$ generated by $E$ and $\delta$ is a field. Hence, if $\delta \notin E$, then $E \subsetneq E'$, and the maximality of $E$ is contradicted.

Conversely, suppose that $E$ is a subfield with $C_k(E) = E$. If $E$ is not a maximal subfield of $D$, then there exists a subfield $E'$ with $E \subsetneq E'$. Now $E' \subseteq C_k(E)$, so that if there is some $a' \in E'$ with $a' \notin E$, then $E \neq C_k(E)$. Therefore, $E$ is a maximal subfield. •

After proving an elementary lemma about tensor products, we will extend the next result from division algebras to central simple algebras (Theorem C-2.128).

**Theorem C-2.119.** *If $D$ is a division algebra over a field $k$ and $E$ is a maximal subfield of $D$, then $E$ is a splitting field for $D$; that is, $E \otimes_k D \cong \mathrm{Mat}_s(E)$, where $s = [D : E] = [E : k]$.*

**Proof.** Let us specialize the algebras in Theorem C-2.116. Here, $A = D$, $B = E$, and $C_A(E) = E$, by Lemma C-2.118. Now the condition $C_A(B) \cong \mathrm{Mat}_r(k)$ becomes $E \cong \mathrm{Mat}_r(k)$; since $E$ is commutative, $r = 1$ and $k = E$. Thus, Corollary C-2.117 says that $E \otimes_k D = E^{\mathrm{op}} \otimes_k D \cong \mathrm{Mat}_s(E)$.

The equality in Theorem C-2.116(iii) is now $[D : k] = [E : k][E : k] = [E : k]^2$. But $[E \otimes_k D : k] = [\mathrm{Mat}_s(E) : k] = s^2[E : k]$, so that $s^2 = [D : k] = [E : k]^2$ and $s = [E : k]$. •

**Corollary C-2.120.** *If $D$ is a division algebra over a field $k$, then all maximal subfields have the same degree over $k$.*

**Remark.** It is not true that maximal subfields in arbitrary division algebras are isomorphic; see Exercise C-2.64 on page 222. ◀

**Proof.** For every maximal subfield $E$, we have $[E : k] = [D : E] = \sqrt{[D : k]}$. •

This corollary can be illustrated by Example C-2.114. The quaternions $\mathbb{H}$ is a four-dimensional $\mathbb{R}$-algebra, so that a maximal subfield must have degree 2 over $\mathbb{R}$; this is so, for $\mathbb{C}$ is a maximal subfield.

We now prove a technical theorem that will yield wonderful results. Recall that a *unit* in a noncommutative ring $A$ is an element having a two-sided inverse in $A$.

**Theorem C-2.121.** *Let $k$ be a field, let $B$ be a simple $k$-algebra, and let $A$ be a central simple $k$-algebra. If there are algebra maps $f, g\colon B \to A$, then there exists a unit $u \in A$ with*

$$g(b) = uf(b)u^{-1}$$

*for all $b \in B$.*

**Proof.** The map $f$ makes $A$ into a left $B$-module if we define the action of $b \in B$ on an element $a \in A$ as $f(b)a$. This action makes $A$ into a $(B, A)$-bimodule, for the associative law in $A$ gives $\big(f(b)x\big)a = f(b)(xa)$ for all $x \in A$. As usual, this $(B, A)$-bimodule is a left $(B \otimes_k A^{\mathrm{op}})$-module, where $(b \otimes a')a = baa'$ for all $a \in A$; denote it by ${}_fA$. Similarly, $g$ can be used to make $A$ into a left $(B \otimes_k A^{\mathrm{op}})$-module we denote by ${}_gA$. By Theorem C-2.113, $B \otimes_k A^{\mathrm{op}}$ is a simple $k$-algebra. Now

$$[{}_fA : k] = [A : k] = [{}_gA : k],$$

so that ${}_fA \cong {}_gA$ as $(B \otimes_k A^{\mathrm{op}})$-modules, by Corollary C-2.42. If $\varphi\colon {}_fA \to {}_gA$ is a $(B \otimes_k A^{\mathrm{op}})$-isomorphism, then

$$\varphi(f(b)aa') = g(b)\varphi(a)a' \tag{4}$$

for all $b \in B$ and $a, a' \in A$. Since $\varphi$ is an automorphism of $A$ as a right module over itself, $\varphi(a) = \varphi(1a) = ua$, where $u = \varphi(1) \in A$. To see that $u$ is a unit, note that $\varphi^{-1}(a) = u'a$ for all $a \in A$. Now $a = \varphi\varphi^{-1}(a) = \varphi(u'a) = uu'a$ for all $a \in A$; in

particular, when $a = 1$, we have $1 = uu'$. The equation $\varphi^{-1}\varphi = 1_A$ gives $1 = u'u$, as desired. Substituting into Eq. (4), we have

$$uf(b)a = \varphi(f(b)a) = g(b)\varphi(a) = g(b)ua$$

for all $a \in A$. In particular, if $a = 1$, then $uf(b) = g(b)u$ and $g(b) = uf(b)u^{-1}$. •

**Corollary C-2.122 (Skolem–Noether).** *Let $A$ be a central simple $k$-algebra over a field $k$, and let $B$ and $B'$ be isomorphic simple $k$-subalgebras of $A$. If $\psi\colon B \to B'$ is an isomorphism, then there exists a unit $u \in A$ with $\psi(b) = ubu^{-1}$ for all $b \in B$.*

**Proof.** In the theorem, take $f\colon B \to A$ to be the inclusion, define $B' = \operatorname{im}\psi$, and define $g = i\psi$, where $i\colon B' \to A$ is the inclusion. •

There is an analog of the Skolem–Noether Theorem in group theory. A theorem of Higman, Neumann, and Neumann says that if $B$ and $B'$ are isomorphic subgroups of a group $G$, say, $\varphi\colon B \to B'$ is an isomorphism, then there exists a group $G^*$ containing $G$ and an element $u \in G^*$ with $\varphi(b) = ubu^{-1}$ for every $b \in B$. There is a proof in Rotman [**188**], p. 404.

**Corollary C-2.123.** *Let $k$ be a field. If $\psi$ is an automorphism of $\mathrm{Mat}_n(k)$, then there exists a nonsingular matrix $P \in \mathrm{Mat}_n(k)$ with*

$$\psi(T) = PTP^{-1}$$

*for every matrix $T$ in $\mathrm{Mat}_n(k)$.*

**Proof.** The ring $A = \mathrm{Mat}_n(k)$ is a central simple $k$-algebra. Set $B = B' = A$ in the Skolem–Noether Theorem. •

Here is another proof of Wedderburn's Theorem C-2.31 in the present spirit.

**Theorem C-2.124 (Wedderburn).** *Every finite division ring $D$ is a field.*

**Proof** (**van der Waerden**). Let $Z = Z(D)$, and let $E$ be a maximal subfield of $D$. If $d \in D$, then $Z(d)$ is a subfield of $D$, and hence there is a maximal subfield $E_d$ containing $Z(d)$. By Corollary C-2.120, all maximal subfields have the same degree, hence have the same order. By Corollary A-3.100 in Part 1, all maximal subfields here are isomorphic (this is not generally true; see Exercise C-2.64 on page 222). For every $d \in D$, the Skolem–Noether Theorem says that there is $x_d \in D$ with $E_d = x_d E x_d^{-1}$. Therefore, $D = \bigcup_x xEx^{-1}$, and so

$$D^\times = \bigcup_x xE^\times x^{-1}.$$

If $E$ is a proper subfield of $D$, then $E^\times$ is a proper subgroup of $D^\times$, and this equation contradicts Exercise C-1.11 on page 14. Therefore, $D = E$ is commutative. •

**Theorem C-2.125 (Frobenius).** *If $D$ is a noncommutative finite-dimensional real division algebra, then $D \cong \mathbb{H}$.*

**Proof.** If $E$ is a maximal subfield of $D$, then $[D : E] = [E : \mathbb{R}] \leq 2$. If $[E : \mathbb{R}] = 1$, then $[D : \mathbb{R}] = 1^2 = 1$ and $D = \mathbb{R}$. Hence, $[E : \mathbb{R}] = 2$ and $[D : \mathbb{R}] = 4$. Let us identify $E$ with $\mathbb{C}$ (we know they are isomorphic). Now complex conjugation is an automorphism of $E$, so that the Skolem–Noether Theorem gives $x \in D$ with $\overline{z} = xzx^{-1}$ for all $z \in E$. In particular, $-i = xix^{-1}$. Hence,

$$x^2ix^{-2} = x(-i)x^{-1} = -xix^{-1} = i,$$

and so $x^2$ commutes with $i$. Therefore, $x^2 \in C_D(E) = E$, by Lemma C-2.118, and so $x^2 = a + bi$ for $a, b \in \mathbb{R}$. But

$$a + bi = x^2 = xx^2x^{-1} = x(a + bi)x^{-1} = a - bi,$$

so that $b = 0$ and $x^2 \in \mathbb{R}$. If $x^2 > 0$, then there is $t \in \mathbb{R}$ with $x^2 = t^2$. Now $(x + t)(x - t) = 0$ gives $x = \pm t \in \mathbb{R}$, contradicting $-i = xix^{-1}$. Therefore, $x^2 = -r^2$ for some real $r$. The element $j$, defined by $j = x/r$, satisfies $j^2 = -1$ and $ji = -ij$. The list $1, i, j, ij$ is linearly independent over $\mathbb{R}$: if $a+bi+cj+dij = 0$, then $(-di-c)j = a+ib \in \mathbb{C}$. Since $j \notin \mathbb{C}$ (lest $x \in \mathbb{C}$), we must have $-di-c = 0 = a+bi$. Hence, $a = b = 0 = c = d$. Since $[D : \mathbb{R}] = 4$, the list $1, i, j, ij$ is a basis of $D$. It is now routine to see that if we define $k = ij$, then $ki = j = -ik$, $jk = i = -kj$, and $k^2 = -1$, and so $D \cong \mathbb{H}$. •

In 1929, Brauer introduced the Brauer group in his study of division rings. Since construction of division rings was notoriously difficult, he considered the wider class of central simple algebras. Brauer introduced the following relation on central simple $k$-algebras.

**Definition.** Two central simple $k$-algebras $A$ and $B$ are ***similar***, denoted by

$$A \sim B,$$

if there are integers $n$ and $m$ with $A \otimes_k \mathrm{Mat}_n(k) \cong B \otimes_k \mathrm{Mat}_m(k)$.

If $A$ is a (finite-dimensional) central simple $k$-algebra, then Corollary C-2.39 and Wedderburn–Artin Theorem II show that $A \cong \mathrm{Mat}_n(D)$ for a unique $k$-division algebra $D$. We shall see that $A \sim B$ if and only if they determine the same division algebra.

**Lemma C-2.126.** *Let $A$ be a finite-dimensional algebra over a field $k$. If $S$ and $T$ are $k$-subalgebras of $A$ such that*

(i) *$st = ts$ for all $s \in S$ and $t \in T$,*

(ii) *$A = ST$,*

(iii) *$[A : k] = [S : k][T : k]$,*

*then $A \cong S \otimes_k T$.*

**Proof.** There is a $k$-linear transformation $f\colon S \otimes_k T \to A$ with $s \otimes t \mapsto st$, because $(s, t) \mapsto st$ is a $k$-bilinear function $S \times T \to A$. Condition (i) implies that $f$ is an algebra map, for

$$f\big((s \otimes t)(s' \otimes t')\big) = f(ss' \otimes tt') = ss'tt' = sts't' = f(s \otimes t)f(s' \otimes t').$$

Since $A = ST$, by condition (ii), the $k$-linear transformation $f$ is a surjection; since $\dim_k(S \otimes_k T) = \dim_k(A)$, by condition (iii), $f$ is a $k$-algebra isomorphism. •

**Lemma C-2.127.** *Let $k$ be a field.*

(i) *If $A$ is a $k$-algebra, then*

$$A \otimes_k \mathrm{Mat}_n(k) \cong \mathrm{Mat}_n(A).$$

(ii) $\mathrm{Mat}_n(k) \otimes_k \mathrm{Mat}_m(k) \cong \mathrm{Mat}_{nm}(k)$.

(iii) *$A \sim B$ is an equivalence relation.*

(iv) *If $A$ is a central simple algebra, then*

$$A \otimes_k A^{\mathrm{op}} \cong \mathrm{Mat}_n(k),$$

*where $n = [A : k]$.*

**Proof.**

(i) Define $k$-subalgebras of $\mathrm{Mat}_n(A)$ by

$$S = \mathrm{Mat}_n(k) \quad \text{and} \quad T = \{aI : a \in A\}.$$

If $s \in S$ and $t \in T$, then $st = ts$ (for the entries of matrices in $S$ commute with elements $a \in A$). Now $S$ contains every matrix unit $E_{ij}$ (whose $ij$ entry is 1 and whose other entries are 0), so that $ST$ contains all matrices of the form $a_{ij}E_{ij}$ for all $i, j$, where $a_{ij} \in A$; hence, $ST = \mathrm{Mat}_n(A)$. Finally, $[S : k][T : k] = n^2[A : k] = [\mathrm{Mat}_n(A) : k]$. Therefore, Lemma C-2.126 gives the desired isomorphism.

(ii) If $V$ and $W$ are vector spaces over $k$ of dimensions $n$ and $m$, respectively, it suffices to prove that $\mathrm{End}_k(V) \otimes_k \mathrm{End}_k(W) \cong \mathrm{End}_k(V \otimes_k W)$. Define $S$ to be all $f \otimes 1_W$, where $f \in \mathrm{End}_k(V)$, and define $T$ to be all $1_V \otimes g$, where $g \in \mathrm{End}_k(W)$. It is routine to check that the three conditions in Lemma C-2.126 hold.

(iii) Since $k = \mathrm{Mat}_1(k)$, we have $A \cong A \otimes_k k \cong A \otimes_k \mathrm{Mat}_1(k)$, so that $\sim$ is reflexive. Symmetry is obvious; for transitivity, suppose that $A \sim B$ and $B \sim C$; that is,

$$A \otimes_k \mathrm{Mat}_n(k) \cong B \otimes_k \mathrm{Mat}_m(k) \quad \text{and} \quad B \otimes_k \mathrm{Mat}_r(k) \cong C \otimes_k \mathrm{Mat}_s(k).$$

Then $A \otimes_k \mathrm{Mat}_n(k) \otimes_k \mathrm{Mat}_r(k) \cong A \otimes_k \mathrm{Mat}_{nr}(k)$, by part (ii). On the other hand,

$$\begin{aligned} A \otimes_k \mathrm{Mat}_n(k) \otimes_k \mathrm{Mat}_r(k) &\cong B \otimes_k \mathrm{Mat}_m(k) \otimes_k \mathrm{Mat}_r(k) \\ &\cong C \otimes_k \mathrm{Mat}_m(k) \otimes_k \mathrm{Mat}_s(k) \\ &\cong C \otimes_k \mathrm{Mat}_{ms}(k). \end{aligned}$$

Therefore, $A \sim C$, and so $\sim$ is an equivalence relation.

(iv) Define $f\colon A \times A^{\mathrm{op}} \to \mathrm{End}_k(A)$ by $f(a, c) = \lambda_a \circ \rho_c$, where $\lambda_a\colon x \mapsto ax$ and $\rho_c\colon x \mapsto xc$; it is routine to check that $\lambda_a$ and $\rho_c$ are $k$-maps (so their composite is also a $k$-map) and that $f$ is $k$-biadditive. Hence, there is a $k$-map $\widehat{f}\colon A \otimes_k A^{\mathrm{op}} \to \mathrm{End}_k(A)$ with $\widehat{f}(a \otimes c) = \lambda_a \circ \rho_c$. Associativity $a(xc) = (ax)c$

in $A$ says that $\lambda_a \circ \rho_c = \rho_c \circ \lambda_a$, from which it easily follows that $\widehat{f}$ is a $k$-algebra map. As $A \otimes_k A^{\mathrm{op}}$ is a simple $k$-algebra and $\ker \widehat{f}$ is a proper two-sided ideal, we have $\widehat{f}$ injective. Now $\dim_k(\mathrm{End}_k(A)) = \dim_k(\mathrm{Hom}_k(A,A)) = n^2$, where $n = [A:k]$. Since $\dim_k(\mathrm{im}\,\widehat{f}) = \dim_k(A \otimes_k A^{\mathrm{op}}) = n^2$, it follows that $\widehat{f}$ is a $k$-algebra isomorphism: $A \otimes_k A^{\mathrm{op}} \cong \mathrm{End}_k(A)$. •

We now extend Theorem C-2.119 from division algebras to central simple algebras.

**Theorem C-2.128.** *Let $A$ be a central simple $k$-algebra over a field $k$, so that $A \cong \mathrm{Mat}_r(D)$, where $D$ is a division algebra over $k$. If $E$ is a maximal subfield of $D$, then $E$ splits $A$; that is, there is an integer $n$ and an isomorphism*

$$E \otimes_k A \cong \mathrm{Mat}_n(E).$$

*More precisely, if $[D:E] = s$, then $n = rs$ and $[A:k] = (rs)^2$.*

**Proof.** By Theorem C-2.119, $D$ is split by a maximal subfield $E$ (which is, of course, a finite extension of $k$): $E \otimes_k D \cong \mathrm{Mat}_s(E)$, where $s = [D:E] = [E:k]$. Hence,

$$\begin{aligned} E \otimes_k A &\cong E \otimes_k \mathrm{Mat}_r(D) \cong E \otimes_k (D \otimes_k \mathrm{Mat}_r(k)) \\ &\cong (E \otimes_k D) \otimes_k \mathrm{Mat}_r(k) \cong \mathrm{Mat}_s(E) \otimes_k \mathrm{Mat}_r(k) \cong \mathrm{Mat}_{rs}(E). \end{aligned}$$

Therefore, $A \cong \mathrm{Mat}_r(D)$ gives $[A:k] = r^2[D:k] = r^2 s^2$. •

**Definition.** If $[A]$ denotes the equivalence class of a central simple $k$-algebra $A$ under similarity, define the ***Brauer group*** $\mathrm{Br}(k)$ to be the set

$$\mathrm{Br}(k) = \big\{[A] : A \text{ is a central simple } k\text{-algebra}\big\}$$

with binary operation

$$[A][B] = [A \otimes_k B].$$

**Theorem C-2.129.** $\mathrm{Br}(k)$ *is an abelian group for every field $k$. Moreover, if $A \cong \mathrm{Mat}_n(D)$ for a division algebra $D$, then $D$ is a central simple $k$-algebra and $[A] = [D]$ in $\mathrm{Br}(k)$.*

**Proof.** We show that the operation is well-defined: if $A, A', B, B'$ are $k$-algebras with $A \sim A'$ and $B \sim B'$, then $A \otimes_k B \sim A' \otimes_k B'$. The isomorphisms

$$A \otimes_k \mathrm{Mat}_n(k) \cong A' \otimes_k \mathrm{Mat}_m(k) \quad \text{and} \quad B \otimes_k \mathrm{Mat}_r(k) \cong B' \otimes_k \mathrm{Mat}_s(k)$$

give $A \otimes_k B \otimes_k \mathrm{Mat}_n(k) \otimes_k \mathrm{Mat}_r(k) \cong A' \otimes_k B' \otimes_k \mathrm{Mat}_m(k) \otimes_k \mathrm{Mat}_s(k)$ (we are using commutativity and associativity of tensor product), so that Lemma C-2.127(ii) gives $A \otimes_k B \otimes_k \mathrm{Mat}_{nr}(k) \cong A' \otimes_k B' \otimes_k \mathrm{Mat}_{ms}(k)$. Therefore, $A \otimes_k B \sim A' \otimes_k B'$.

That $[k]$ is the identity follows from $k \otimes_k A \cong A$, associativity and commutativity follow from associativity and commutativity of tensor product, and Lemma C-2.127(iv) shows that $[A]^{-1} = [A^{\mathrm{op}}]$. Therefore, $\mathrm{Br}(k)$ is an abelian group.

If $A$ is a central simple $k$-algebra, then $A \cong \mathrm{Mat}_r(D)$ for some finite-dimensional division algebra $D$ over $k$. Hence, $k = Z(A) \cong Z(\mathrm{Mat}_r(D)) \cong Z(D)$, by Theorem C-2.113. Thus, $D$ is a central simple $k$-algebra, $[D] \in \mathrm{Br}(k)$, and $[D] = [A]$ (because $D \otimes_k \mathrm{Mat}_r(k) \cong \mathrm{Mat}_r(D) \cong A \cong A \otimes_k k \cong A \otimes_k \mathrm{Mat}_1(k)$). •

The next proposition shows the significance of the Brauer group.

**Proposition C-2.130.** *If $k$ is a field, then there is a bijection from $\mathrm{Br}(k)$ to the family $\mathcal{D}$ of all isomorphism classes of finite-dimensional division algebras over $k$, and so $|\mathrm{Br}(k)| = |\mathcal{D}|$. Therefore, there exists a noncommutative division ring, finite-dimensional over its center $k$, if and only if $\mathrm{Br}(k) \neq \{0\}$.*

**Proof.** Define a function $\varphi\colon \mathrm{Br}(k) \to \mathcal{D}$ by setting $\varphi([A])$ to be the isomorphism class of $k$ if $A \cong \mathrm{Mat}_n(k)$. Note that Theorem C-2.129 shows that $[A] = [k]$ in $\mathrm{Br}(k)$. Let us see that $\varphi$ is well-defined. If $[k] = [k']$, then $k \sim k'$, so there are integers $n$ and $m$ with $k \otimes_k \mathrm{Mat}_n(k) \cong k' \otimes_k \mathrm{Mat}_m(k)$. Hence, $\mathrm{Mat}_n(k) \cong \mathrm{Mat}_m(k')$. By the uniqueness in the Wedderburn–Artin Theorems, $k \cong k'$ (and $n = m$). Therefore, $\varphi([k]) = \varphi([k'])$.

Clearly, $\varphi$ is surjective, for if $k$ is a finite-dimensional division algebra over $k$, then the isomorphism class of $k$ is equal to $\varphi([k])$. To see that $\varphi$ is injective, suppose that $\varphi([k]) = \varphi([k'])$. Then, $k \cong k'$, which implies $k \sim k'$. •

**Example C-2.131.**

(i) If $k$ is an algebraically closed field, then $\mathrm{Br}(k) = \{0\}$, by Theorem C-2.115.

(ii) If $k$ is a finite field, then Wedderburn's Theorem C-2.124 (= Theorem C-2.31) shows that $\mathrm{Br}(k) = \{0\}$.

(iii) If $k = \mathbb{R}$, then Frobenius's Theorem C-2.125 shows that $\mathrm{Br}(\mathbb{R}) \cong \mathbb{Z}_2$.

(iv) It is proved, using class field theory, that $\mathrm{Br}(\mathbb{Q}_p) \cong \mathbb{Q}/\mathbb{Z}$, where $\mathbb{Q}_p$ is the field of $p$-adic numbers. Moreover, there is an exact sequence

$$0 \to \mathrm{Br}(\mathbb{Q}) \to \mathrm{Br}(\mathbb{R}) \oplus \bigoplus_p \mathrm{Br}(\mathbb{Q}_p) \to \mathbb{Q}/\mathbb{Z} \to 0. \quad \blacktriangleleft$$

In a series of deep papers, $\mathrm{Br}(k)$ was computed for the most interesting fields $k$ arising in algebraic number theory (*local fields*, one of which is $\mathbb{Q}_p$, and *global fields*) by Albert, Brauer, Hasse, and Noether.

**Proposition C-2.132.** *If $E/k$ is a field extension, then there is a homomorphism*

$$f_{E/k}\colon \mathrm{Br}(k) \to \mathrm{Br}(E)$$

*given by $[A] \mapsto [E \otimes_k A]$.*

**Proof.** If $A$ and $B$ are central simple $k$-algebras, then $E \otimes_k A$ and $E \otimes_k B$ are central simple $E$-algebras, by Theorem C-2.113. If $A \sim B$, then $E \otimes_k A \sim E \otimes_k B$ as $E$-algebras, by Exercise C-2.61 on page 221. It follows that the function $f_{E/k}$ is well-defined. Finally, $f_{E/k}$ is a homomorphism, because

$$(E \otimes_k A) \otimes_E (E \otimes_k B) \cong (E \otimes_E E) \otimes_k (A \otimes_k B) \cong E \otimes_k (A \otimes_k B),$$

by Proposition B-5.3 in Part 1, associativity of tensor product. •

**Definition.** If $E/k$ is a field extension, then the ***relative Brauer group***, $\mathrm{Br}(E/k)$, is the kernel of the homomorphism $f_{E/k}\colon \mathrm{Br}(k) \to \mathrm{Br}(E)$:

$$\mathrm{Br}(E/k) = \ker f_{E/k} = \big\{[A] \in \mathrm{Br}(k) : A \text{ is split by } E\big\}.$$

**Corollary C-2.133.** *For every field $k$, we have*

$$\mathrm{Br}(k) = \bigcup_{E/k \text{ finite Galois}} \mathrm{Br}(E/k).$$

**Proof.** This follows from Theorem C-2.128 after showing that we may assume that $E/k$ is Galois. •

In a word, the Brauer group arose as a way to study division rings. It is an interesting object, but we have not really used it seriously. For example, we have not yet seen any noncommutative division rings other than the real division algebra $\mathbb{H}$ (and its variants for subfields $k$ of $\mathbb{R}$). We will remedy this when we introduce *crossed product algebras* in Chapter C-3 on homology. For example, we will see there that division rings exist whose center is a field of characteristic $p > 0$. For further developments, we refer the reader to Jacobson [**112**] and Reiner [**180**].

We introduce homology groups in the next chapter, and we will see Theorem C-3.140: if $E/k$ is a finite Galois extension, then the relative Brauer group $B(E/k) \cong H^2(G, E^\times)$, where $G = \mathrm{Gal}(E/k)$ and $E^\times$ is the multiplicative group of nonzero elements of the field $E$. This will imply that Brauer groups $\mathrm{Br}(k)$ are torsion groups.

## Exercises

**C-2.50.** (i) If $k$ is a subfield of a field $K$, prove that the ring $K \otimes_k k[x]$ is isomorphic to $K[x]$.

(ii) Suppose that $k$ is a field, $p(x) \in k[x]$ is irreducible, and $K = k(\alpha)$, where $\alpha$ is a root of $p(x)$. Prove that, as rings, $K \otimes_k K \cong K[x]/(p(x))$, where $(p(x))$ is the principal ideal in $K[x]$ generated by $p(x)$.

(iii) The polynomial $p(x)$, though irreducible in $k[x]$, may factor in $K[x]$. Give an example showing that the ring $K \otimes_k K$ need not be semisimple.

(iv) Prove that if $K/k$ is a finite separable extension, then $K \otimes_k K$ is semisimple. (The converse is also true.)

**C-2.51.** If $A \cong A'$ and $B \cong B'$ are $k$-algebras, where $k$ is a commutative ring, prove that $A \otimes_k B \cong A' \otimes_k B'$ as $k$-algebras.

∗ **C-2.52.** If $k$ is a commutative ring and $A$ and $B$ are $k$-algebras, prove that

$$(A \otimes_k B)^{\mathrm{op}} \cong A^{\mathrm{op}} \otimes_k B^{\mathrm{op}}.$$

**C-2.53.** If $R$ is a commutative $k$-algebra, where $k$ is a field and $G$ is a group, prove that $R \otimes_k kG \cong RG$.

∗ **C-2.54.** (i) If $k$ is a subring of a commutative ring $R$, prove that $R \otimes_k k[x] \cong R[x]$ as $R$-algebras.

(ii) If $f(x) \in k[x]$ and $(f)$ is the principal ideal in $k[x]$ generated by $f(x)$, prove that $R \otimes_k (f)$ is the principal ideal in $R[x]$ generated by $f(x)$. More precisely, there is a commutative diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & R \otimes_k (f) & \longrightarrow & R \otimes_k k[x] \\ & & \downarrow & & \downarrow \\ 0 & \longrightarrow & (f) & \longrightarrow & R[x]. \end{array}$$

(iii) Let $k$ be a field and $E \cong k[x]/(f)$, where $f(x) \in k[x]$ is irreducible. Prove that $E \otimes_k E \cong E[x]/(f)_E$, where $(f)_E$ is the principal ideal in $E[x]$ generated by $f(x)$.

(iv) Give an example of a field extension $E/k$ with $E \otimes_k E$ not a field.
**Hint.** If $f(x) \in k[x]$ factors into $g(x)h(x)$ in $E[x]$, where $(g, h) = 1$, then the Chinese Remainder Theorem applies.

**C-2.55.** Let $k$ be a field and let $f(x) \in k[x]$ be irreducible. If $K/k$ is a field extension, then $f(x) = p_1(x)^{e_1} \cdots p_n(x)^{e_n} \in K[x]$, where the $p_i(x)$ are distinct irreducible polynomials in $K[x]$ and $e_i \geq 1$.

(i) Prove that $f(x)$ is separable if and only if all $e_i = 1$.

(ii) Prove that a finite field extension $K/k$ is separable if and only if $K \otimes_k K$ is a semisimple ring.
**Hint.** Observe that $K/k$ is a simple extension, so there is an exact sequence $0 \to (f) \to k[x] \to K \to 0$, and then use the Chinese Remainder Theorem.

**C-2.56.** Prove that $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathrm{Mat}_4(\mathbb{R})$ as $\mathbb{R}$-algebras.

**Hint.** Use Corollary C-2.39 for the central simple $\mathbb{R}$-algebra $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}$.

**C-2.57.** We have given one isomorphism $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathrm{Mat}_2(\mathbb{C})$ in Example C-2.114. Describe all possible isomorphisms between these two algebras.

**Hint.** Use the Skolem–Noether Theorem.

**C-2.58.** Prove that $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \times \mathbb{C}$ as $\mathbb{R}$-algebras.

**C-2.59.** (i) Let $\mathbb{C}(x)$ and $\mathbb{C}(y)$ be function fields. Prove that $R = \mathbb{C}(x) \otimes_{\mathbb{C}} \mathbb{C}(y)$ is isomorphic to a subring of $\mathbb{C}(x, y)$. Conclude that $R$ has no zero-divisors.

(ii) Prove that $\mathbb{C}(x) \otimes_{\mathbb{C}} \mathbb{C}(y)$ is not a field.
**Hint.** Show that $R$ is isomorphic to the subring of $\mathbb{C}(x, y)$ consisting of polynomials of the form $f(x, y)/g(x)h(y)$.

(iii) Use Exercise B-1.29 on page 288 in Part 1 to prove that the tensor product of artinian algebras need not be artinian.

∗ **C-2.60.** Let $A$ be a central simple $k$-algebra. If $A$ is split by a field $E$, prove that $A$ is split by any field extension $E'$ of $E$.

∗ **C-2.61.** Let $E/k$ be a field extension. If $A$ and $B$ are central simple $k$-algebras with $A \sim B$, prove that $E \otimes_k A \sim E \otimes_k B$ as central simple $E$-algebras.

**C-2.62.** If $D$ is a finite-dimensional division algebra over $\mathbb{R}$, prove that $D$ is isomorphic to either $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$.

**C-2.63.** Prove that $\mathrm{Mat}_2(\mathbb{H}) \cong \mathbb{H} \otimes_{\mathbb{R}} \mathrm{Mat}_2(\mathbb{R})$ as $\mathbb{R}$-algebras.

∗ **C-2.64.** (i) Let $A$ be a four-dimensional vector space over $\mathbb{Q}$, and let $1, i, j, k$ be a basis. Show that $A$ is a division algebra if we define 1 to be the identity and

$$\begin{aligned} i^2 &= -1, & j^2 &= -2, & k^2 &= -2, \\ ij &= \ k, & jk &= \ 2i, & ki &= \ j, \\ ji &= -k, & kj &= -2i, & ik &= -j. \end{aligned}$$

Prove that $A$ is a division algebra over $\mathbb{Q}$.

(ii) Prove that $\mathbb{Q}(i)$ and $\mathbb{Q}(j)$ are nonisomorphic maximal subfields of $A$.

**C-2.65.** Let $D$ be the $\mathbb{Q}$-subalgebra of $\mathbb{H}$ having basis $1, i, j, k$.

(i) Prove that $D$ is a division algebra over $\mathbb{Q}$.
**Hint.** Compute the center $Z(D)$.

(ii) For any pair of nonzero rationals $p$ and $q$, prove that $D$ has a maximal subfield isomorphic to $\mathbb{Q}(\sqrt{-p^2 - q^2})$.
**Hint.** Compute $(pi + qj)^2$.

**C-2.66. (Dickson)** If $D$ is a division algebra over a field $k$, then each $d \in D$ is algebraic over $k$. Prove that $d, d' \in D$ are conjugate in $D$ if and only if $\operatorname{irr}(d, k) = \operatorname{irr}(d', k)$.

**Hint.** Use the Skolem–Noether Theorem.

**C-2.67.** Prove that if $A$ is a central simple $k$-algebra with $A \sim \operatorname{Mat}_n(k)$, then $A \cong \operatorname{Mat}_m(k)$ for some integer $m$.

**C-2.68.** Prove that if $A$ is a central simple $k$-algebra with $[A]$ of finite order $m$ in $\operatorname{Br}(k)$, then there is an integer $r$ with

$$A \otimes_k \cdots \otimes_k A \cong \operatorname{Mat}_r(k)$$

(there are $m$ factors equal to $A$). In Chapter C-3, we shall see that every element in $\operatorname{Br}(k)$ has finite order.

# Homology

## C-3.1. Introduction

[1] When I was a graduate student, homological algebra was an unpopular subject. The general attitude was that it was a grotesque formalism, boring to learn, and not very useful once one had learned it. Perhaps an algebraic topologist was forced to know this stuff, but surely no one else should waste time on it. The few true believers were viewed as workers at the fringe of mathematics who kept tinkering with their elaborate machine, smoothing out rough patches here and there.

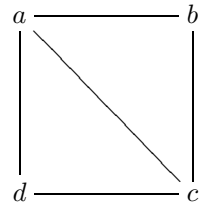
This attitude changed dramatically when Serre characterized regular local rings using homological algebra (they are the commutative noetherian local rings of "finite global dimension"), for this enabled him to prove that any localization of a regular local ring is itself regular (until then, only special cases of this were known). At the same time, M. Auslander and Buchsbaum completed work of Nagata by using global dimension to prove that every regular local ring is a UFD.

In spite of its newfound popularity, homological algebra still "got no respect". For example, the two theorems just mentioned used the notion of *global dimension* of a ring which, in turn, is defined in terms of the *homological dimension* of a module. At that time, in 1957, Kaplansky offered a course in homological algebra. One of his students, Schanuel, noticed that there is an elegant relation between different projective resolutions of the same module (see Proposition B-4.48 in Part 1). Kaplansky seized this result, nowadays called Schanuel's Lemma, for it allowed him to define the homological dimension of a module without having first to develop the fundamental constructs Ext and Tor of homological algebra, and he was then able to prove the theorems of Serre and of Auslander–Buchsbaum (Kaplansky's account of this course can be found in his book [**118**]). However, as more applications

[1] This introduction is adapted from a review I wrote that appeared in *Bulletin of the American Mathematical Society* **33** (1996), 473–475; it is reproduced by permission of the American Mathematical Society.

were found and as more homology and cohomology theories were invented to solve outstanding problems, resistance to homological algebra waned. Today, it is just another standard tool in a mathematician's kit.

The basic idea of homology comes from Green's Theorem, where a double integral over a region $R$ with holes in it is equal to a line integral on the boundary of $R$. Poincaré recognized that whether a topological space $X$ has different kinds of holes is a kind of connectivity. To illustrate, let us assume that $X$ can be "triangulated"; that is, $X$ can be partitioned into finitely many *$n$-simplexes*, where $n \geq 0$: points are 0-simplexes, edges are 1-simplexes, triangles are 2-simplexes, tetrahedra are 3-simplexes, and there are higher-dimensional analogs. The question to ask is whether a union of $n$-simplexes in $X$ that "ought" to be the boundary of some $(n+1)$-simplex actually is such a boundary. For example, when $n = 0$, two points $a$ and $b$ in $X$ ought to be the boundary (endpoints) of a path in $X$; if there is a path in $X$ joining all points $a$ and $b$, then $X$ is called *path connected*; if there is no such path, then $X$ has a 0-dimensional hole. For an example of a one-dimensional hole, let $X$ be the *punctured plane*; that is, the plane with the origin deleted. The perimeter of a triangle $\Delta$ ought to be the boundary of a 2-simplex, but this is not so if $\Delta$ contains the origin in its interior; thus, $X$ has a one-dimensional hole. If $X$ were missing a line segment containing the origin, or even a small disk containing the origin, this hole would still be one-dimensional; we are not considering the size of the hole, but the size of the possible boundary. We must keep our eye on the doughnut and not upon the hole!




For example, in the rectangle drawn above, consider the triangle $[a, b, c]$ with vertices $a$, $b$, $c$ and edges $[a, b]$, $[b, c]$, $[a, c]$. Its boundary $\partial[a, b, c]$ should be $[a, b] + [b, c] + [c, a]$. But edges are oriented (think of $[a, c]$ as a path from $a$ to $c$ and $[c, a]$ as the reverse path from $c$ to $a$), so let us write $[c, a] = -[a, c]$. Thus, the boundary is

$$\partial[a, b, c] = [a, b] - [a, c] + [b, c].$$

Similarly, let us define the boundary of $[a, b]$ to be its endpoints:

$$\partial[a, b] = b - a.$$

We note that

$$\begin{aligned} \partial(\partial[a, b, c]) &= \partial([a, b] - [a, c] + [b, c]) \\ &= b - a - (c - a) + c - b \\ &= 0. \end{aligned}$$

The rectangle with vertices $a, b, c, d$ is the union of two triangles $[a, b, c] + [a, c, d]$, and we check that its boundary is $\partial[a, b, c] + \partial[a, c, d]$ (note that the diagonal $[a, c]$ occurs twice, with different signs, and so it cancels, as it should). We see that

the formalism suggests the use of signs to describe boundaries as certain linear combinations $u$ of edges or points for which $\partial(u) = 0$.

Such ideas lead to the following construction. For each $n \geq 0$, consider all formal linear combinations of $n$-simplexes; that is, form the free abelian group $C_n(X)$ with basis all $n$-simplexes, and call such linear combinations *$n$-chains*. Some of these $n$-chains ought to be boundaries of some union of $(n+1)$-simplexes; call them *$n$-cycles* (for example, adding the three edges of a triangle, with appropriate choice of signs, is a 1-cycle). Certain $n$-chains actually are boundaries, and these are called *$n$-boundaries* (if $\Delta$ is a triangle in the punctured plane $X$, not having the origin in its interior, then the alternating sum of the edges of $\Delta$ is a 1-boundary; on the other hand, if the origin does lie in the interior of $\Delta$, then the alternating sum is a 1-cycle but not a 1-boundary). The family of all the $n$-cycles, $Z_n(X)$, and the family of all the $n$-boundaries, $B_n(X)$, are subgroups of $C_n(X)$. A key ingredient in the construction of homology groups is that the subgroups $Z_n$ and $B_n$ can be defined in terms of homomorphisms: there are *boundary homomorphisms* $\partial_n \colon C_n(X) \to C_{n-1}(X)$ with $Z_n = \ker \partial_n$ and $B_n = \operatorname{im} \partial_{n+1}$, and so there is a sequence of abelian groups and homomorphisms

$$\cdots \to C_3(X) \xrightarrow{\partial_3} C_2(X) \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X).$$

It turns out, for all $n \geq 1$, that $\partial_n \partial_{n+1} = 0$, from which it follows that

$$B_n(X) \subseteq Z_n(X).$$

The interesting group is the quotient group $Z_n(X)/B_n(X)$, denoted by $H_n(X)$ and called the $n$th *homology*[2] *group* of $X$. What survives in this quotient group are the $n$-dimensional holes; that is, those $n$-cycles that are not $n$-boundaries. For example, $H_0(X) = 0$ means that $X$ is path connected: if there are two points $a, b \in X$ that are not connected by a path, then $a - b$ is a cycle that is not a boundary, and so the coset $a - b + B_0(X)$ is a nonzero element of $H_0(X)$. For $n \geq 1$, these groups measure more subtle kinds of connectivity. Topologists modify this construction in two ways. They introduce homology with *coefficients* in an abelian group $G$ by tensoring the sequence of chain groups by $G$ and then taking homology groups; they also consider *cohomology with coefficients* in $G$ by applying the contravariant functor $\operatorname{Hom}(\ , G)$ to the sequence of chain groups and then taking homology groups. Homological algebra arose in trying to compute and to find relations between homology groups of spaces.

[2] I have not been able to discover the etymology of the mathematical term *homology* as used in this context. The word "homology" comes from *homo* + *logos*, and it means "corresponding". Its first usage as a mathematical term occurred in projective geometry in the early nineteenth century, as the name of a specific type of collineation. The earliest occurrence I have found for its usage in the sense of cycles and boundaries is in an article of H. Poincaré: *Analysis Situs*, Journal de l'École Polytechnique, Series II, first issue, 1895 (and Oeuvres, vol. 5), but he does not explain why he chose the term. Emili Bifet has written, in a private communication, "Consider the projective homology, between two distinct (hyper)planes, given by projection from an exterior point. This homology suggests (and provides) a natural way of deforming the boundary of a simplex contained in one plane into the boundary of the corresponding simplex on the other one. Moreover, it suggests a natural way of deforming a boundary into a point. This could be what Poincaré had in mind."

We have already seen, in Proposition B-2.25 in Part 1, that every left $R$-module $M$, where $R$ is a ring, has a description by generators and relations. There is an exact sequence

$$0 \to \ker \varphi \xrightarrow{\iota} F_0 \xrightarrow{\varphi} M \to 0,$$

where $F_0$ is a free left $R$-module and $\iota$ is the inclusion. If $R$ is a PID, then $\ker \varphi$ is free, because every submodule of a free module is itself free; if $R$ is not a PID, then $\ker \varphi$ may not be free. Now take generators and relations of $\ker \varphi$: There is a free module $F_1$ and an exact sequence

$$0 \to \ker \psi \xrightarrow{\kappa} F_1 \xrightarrow{\psi} \ker \varphi \to 0.$$

If we define $F_1 \to F_0$ to be the composite $\iota\psi$, then there is a second exact sequence

$$F_1 \xrightarrow{\iota\psi} F_0 \xrightarrow{\varphi} M \to 0,$$

and, iterating this construction, there is a long exact sequence

$$\cdots \to F_3 \to F_2 \to F_1 \to F_0 \to M \to 0.$$

We can view the submodules $\ker(F_n \to F_{n-1})$ as "relations on relations" (nineteenth-century algebraists called these higher relations *syzygies*). This long exact sequence resembles the sequence of chain groups in topology. There are other contexts in which such exact sequences exist; many algebraic structures give rise to a sequence of homology groups, and these can be used to translate older theorems into the language of homology. Examples of such theorems are Hilbert's Theorem 90 about algebras (see Corollary C-3.138), Whitehead's lemmas about Lie algebras (see Jacobson [**113**], pp. 77 and 89), and Theorem C-3.22, the Schur–Zassenhaus lemma, about groups. There are methods to compute homology and cohomology groups, and this is the most important contribution of homological algebra to this circle of ideas. Although we can calculate many things without them, the most powerful method of computing homology groups uses *spectral sequences*. When I was a graduate student, I always wanted to be able to say, nonchalantly, that such and such is true "by the usual spectral sequence argument", but I never had the nerve. We will introduce spectral sequences at the end of this chapter.

## C-3.2. Semidirect Products

We begin by investigating a basic problem in group theory. A group $G$ having a normal subgroup $K$ can be "factored" into $K$ and $G/K$; the study of extensions involves the inverse question: how much of $G$ can be recovered from a normal subgroup $K$ and the quotient $Q = G/K$? For example, we know that $|G| = |K||Q|$ if $K$ and $Q$ are finite.

Exactness of a sequence of nonabelian groups,

$$\cdots \to G_{n+1} \xrightarrow{d_{n+1}} G_n \xrightarrow{d_n} G_{n-1} \to \cdots,$$

is defined just as it is for abelian groups: $\operatorname{im} d_{n+1} = \ker d_n$ for all $n$. Of course, each $\ker d_n$ is a normal subgroup of $G_n$.

**Definition.** If $K$ and $Q$ are groups, then an ***extension*** of $K$ by $Q$ is a short exact sequence

$$1 \to K \xrightarrow{i} G \xrightarrow{p} Q \to 1.$$

The notations $K$ and $Q$ remind us of kernel and quotient.

There is an alternative usage of the term *extension*, which calls the (middle) group $G$ (not the short exact sequence) an *extension* if it contains a normal subgroup $K_1$ with $K_1 \cong K$ and $G/K_1 \cong Q$. As do most people, we will use the term in both senses.

**Example C-3.1.**

(i) The direct product $K \times Q$ is an extension of $K$ by $Q$; it is also an extension of $Q$ by $K$.

(ii) Both $S_3$ and $\mathbb{Z}_6$ are extensions of $\mathbb{Z}_3$ by $\mathbb{Z}_2$. On the other hand, $\mathbb{Z}_6$ is an extension of $\mathbb{Z}_2$ by $\mathbb{Z}_3$, but $S_3$ is not, for $S_3$ contains no *normal* subgroup of order 2. ◀

We have just seen, for any given pair of groups $K$ and $Q$, that an extension of $K$ by $Q$ always exists (the direct product), but there may be nonisomorphic such extensions. Hence, an extension of $K$ by $Q$ may be viewed as a "product" of $K$ and $Q$, but this product is not single-valued. The ***extension problem*** is to classify all possible extensions of a given pair of groups $K$ and $Q$.

Suppose that a group $G$ has a normal series

$$G = K_0 \supseteq K_1 \supseteq \cdots \supseteq K_{n-1} \supseteq K_n = \{1\}$$

with factor groups $Q_1, \ldots, Q_n$, where

$$Q_i = K_{i-1}/K_i$$

for all $i \geq 1$. Now $K_n = \{1\}$, so that $K_{n-1} = Q_n$, but something more interesting occurs next: $K_{n-2}/K_{n-1} = Q_{n-1}$, so that $K_{n-2}$ is an extension of $K_{n-1}$ by $Q_{n-1}$. If we could solve the extension problem, then we could recapture $K_{n-2}$ from $K_{n-1}$ and $Q_{n-1}$—that is, from $Q_n$ and $Q_{n-1}$. Next, observe that $K_{n-3}/K_{n-2} = Q_{n-2}$, so that $K_{n-3}$ is an extension of $K_{n-2}$ by $Q_{n-2}$. If we could solve the extension problem, then we could recapture $K_{n-3}$ from $K_{n-2}$ and $Q_{n-2}$; that is, we could recapture $K_{n-3}$ from $Q_n, Q_{n-1}$, and $Q_{n-2}$. Climbing up the composition series in this way, we could recapture $G = K_0$ from $Q_n, Q_{n-1}, \ldots, Q_1$; that is, $G$ is a "product" of the factor groups. If the normal series is a composition series, then the Jordan–Hölder Theorem is a unique factorization theorem: the factors in this product, namely, the composition factors of $G$, are uniquely determined by $G$. Therefore, we could survey all finite groups if we knew the finite simple groups and if we could solve the extension problem. All the finite simple groups are now known; the proof of the ***Classification Theorem of Finite Simple Groups*** was completed in the first decade of the twenty-first century. This theorem, one of the deepest theorems in mathematics (its proof needs several thousand pages!), gives a complete list of all the finite simple groups, along with interesting properties of them.

Let us begin by recalling the partition of a group into the cosets of a subgroup. We have already defined a *transversal* of a subgroup $K$ of a group $G$ as a subset $T$ of $G$ consisting of exactly one element from each coset[3] $Kt$ of $K$.

**Definition.** If

$$1 \to K \to G \xrightarrow{p} Q \to 1$$

is an extension, then a ***lifting*** is a function $\ell\colon Q \to G$, not necessarily a homomorphism, with $p\ell = 1_Q$.

Given a transversal, we can construct a lifting. For each $x \in Q$, surjectivity of $p$ provides $\ell(x) \in G$ with $p\ell(x) = x$; thus, the function $x \mapsto \ell(x)$ is a lifting. Conversely, given a lifting, we claim that $\operatorname{im}\ell$ is a transversal of $K$. If $Kg$ is a coset, then $p(g) \in Q$; say, $p(g) = x$. Then $p(g\ell(x)^{-1}) = 1$, so that $a = g\ell(x)^{-1} \in K$ and $Kg = K\ell(x)$. Thus, every coset has a representative in $\ell(Q)$. Finally, we must show that $\ell(Q)$ does not contain two elements in the same coset. If $K\ell(x) = K\ell(y)$, then there is $a \in K$ with $a\ell(x) = \ell(y)$. Apply $p$ to this equation; since $p(a) = 1$, we have $x = y$ and so $\ell(x) = \ell(y)$.

Recall that an ***automorphism*** of a group $K$ is an isomorphism $K \to K$. The ***automorphism group***, denoted by $\operatorname{Aut}(K)$, is the group of all the automorphisms of $K$ with composition as operation.

Of course, extensions are defined for arbitrary groups $K$, but we are going to restrict our attention to the special case when $K$ is abelian. If $G$ is an extension of $K$ by $Q$, it would be confusing to write $G$ multiplicatively and its subgroup $K$ additively. Hence, we shall use the following notational convention: even though $G$ may not be abelian, additive notation will be used for the operation in $G$. Corollary C-3.4 below gives the main reason for this decision.

**Proposition C-3.2.** *Let*

$$0 \to K \xrightarrow{i} G \xrightarrow{p} Q \to 1$$

*be an extension of an abelian group $K$ by a group $Q$, and let $\ell\colon Q \to G$ be a lifting.*

(i) *For every $x \in Q$, conjugation $\theta_x\colon K \to K$, defined by*

$$\theta_x\colon a \mapsto \ell(x) + a - \ell(x),$$

*is independent of the choice of lifting $\ell(x)$ of $x$. (For convenience, we have assumed that $i$ is an inclusion; this merely allows us to write $a$ instead of $i(a)$.)*

(ii) *The function $\theta\colon Q \to \operatorname{Aut}(K)$, defined by $x \mapsto \theta_x$, is a homomorphism.*

[3] We have been working with *left* cosets $tK$, but, in this chapter, the subgroup $K$ will be a normal subgroup, in which case $tK = Kt$ for all $t \in G$. Thus, using right cosets or left cosets is only a matter of convenience.

**Proof.**

(i) Let us now show that $\theta_x$ is independent of the choice of lifting $\ell(x)$ of $x$. Suppose that $\ell'(x) \in G$ and $p\ell'(x) = x$. There is $b \in K$ with $\ell'(x) = \ell(x) + b$ (for $-\ell(x) + \ell'(x) \in \ker p = \operatorname{im} i = K$). Therefore,

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

because $K$ is abelian.

(ii) Now $\theta_x(a) \in K$ because $K \lhd G$, so that each $\theta_x\colon K \to K$; also, $\theta_x$ is an automorphism of $K$, because conjugations are automorphisms.

It remains to prove that $\theta\colon Q \to \operatorname{Aut}(K)$ is a homomorphism. If $x, y \in Q$ and $a \in K$, then

$$\theta_x(\theta_y(a)) = \theta_x(\ell(y) + a - \ell(y)) = \ell(x) + \ell(y) + a - \ell(y) - \ell(x),$$

while

$$\theta_{xy}(a) = \ell(xy) + a - \ell(xy).$$

But $\ell(x) + \ell(y)$ and $\ell(xy)$ are both liftings of $xy$, so that equality $\theta_x\theta_y = \theta_{xy}$ follows from (i). •

Informally, the homomorphism $\theta$ tells "how" $K$ is normal in $G$, for isomorphic copies of a group can sit as normal subgroups of $G$ in different ways. For example, let $K$ be a cyclic group of order 3 and let $Q = \langle x\rangle$ be cyclic of order 2. If $G = K \times Q$, then $G$ is abelian and $K$ lies in the center of $G$. In this case, $\ell(x) + a - \ell(x) = a$ for all $a \in K$ and $\theta_x = 1_K$. On the other hand, if $G = S_3$, then $K = A_3$ which does not lie in the center; if $\ell(x) = (1\ 2)$, then $(1\ 2)(1\ 2\ 3)(1\ 2) = (1\ 3\ 2)$ and $\theta_x$ is not $1_K$.

The existence of a homomorphism $\theta$ equips $K$ with a scalar multiplication making $K$ a left $\mathbb{Z}Q$-module, where $\mathbb{Z}Q$ is the group ring whose elements are all $\sum_{x\in Q} m_x x$ for $m_x \in \mathbb{Z}$.

**Proposition C-3.3.** *Let $K$ and $Q$ be groups with $K$ abelian. Then a homomorphism $\theta\colon Q \to \operatorname{Aut}(K)$ makes $K$ into a left $\mathbb{Z}Q$-module if scalar multiplication is defined by*

$$xa = \theta_x(a)$$

*for all $a \in K$ and $x \in Q$. Conversely, if $K$ is a left $\mathbb{Z}Q$-module, then $x \mapsto \theta_x$ defines a homomorphism $\theta\colon Q \to \operatorname{Aut}(K)$, where $\theta_x\colon a \mapsto xa$.*

**Proof.** Define scalar multiplication as follows. Each $u \in \mathbb{Z}Q$ has a unique expression of the form $u = \sum_{x\in Q} m_x x$, where $m_x \in \mathbb{Z}$ and $m_x = 0$ for all but finitely many $x \in Q$; define

$$\Big(\sum_x m_x x\Big)a = \sum_x m_x\theta_x(a) = \sum_x m_x(xa).$$

We verify the module axioms. Since $\theta$ is a homomorphism, $\theta(1) = 1_K$, and so $1a = \theta_1(a)$ for all $a \in K$. That $\theta_x \in \operatorname{Aut}(K)$ implies $x(a + b) = xa + xb$, from which it follows that $u(a + b) = ua + ub$ for all $u \in \mathbb{Z}Q$. Similarly, we check easily

that $(u+v)a = ua + va$ for $u, v \in \mathbb{Z}Q$. Finally, $(uv)a = u(va)$ will follow from $(xy)a = x(ya)$ for all $x, y \in Q$; but

$$(xy)a = \theta_{xy}(a) = \theta_x(\theta_y(a)) = \theta_x(ya) = x(ya).$$

The proof of the converse is also routine. •

**Corollary C-3.4.** *If*

$$0 \to K \xrightarrow{i} G \xrightarrow{p} Q \to 1$$

*is an extension of an abelian group $K$ by a group $Q$, then $K$ is a left $\mathbb{Z}Q$-module if we define*

$$xa = \ell(x) + a - \ell(x),$$

*where $\ell\colon Q \to G$ is a lifting, $x \in Q$, and $a \in K$; moreover, the scalar multiplication is independent of the choice of lifting $\ell$.*

**Proof.** Propositions C-3.2 and C-3.3. •

From now on, we will abbreviate the term

"left $\mathbb{Z}Q$-module" as "$Q$-module".

Recall that a short exact sequence of left $R$-modules

$$0 \to A \xrightarrow{i} B \xrightarrow{p} C \to 0$$

is *split* if there exists a homomorphism $j\colon C \to B$ with $pj = 1_C$; in this case, the middle module is isomorphic to the direct sum $A \oplus C$. Here is the analogous definition for groups.

**Definition.** An extension of groups

$$0 \to K \xrightarrow{i} G \xrightarrow{p} Q \to 1$$

is ***split*** if there is a homomorphism $j\colon Q \to G$ with $pj = 1_Q$. The middle group $G$ in a split extension is called a ***semidirect product*** of $K$ by $Q$.

Thus, an extension is split if and only if there is a lifting, namely, $j$, that is also a homomorphism. We shall use the following notation: the elements of $K$ shall be denoted by $a, b, c, \ldots$, and the elements of $Q$ shall be denoted by $x, y, z, \ldots$.

**Proposition C-3.5.** *Let $G$ be an additive group having a normal subgroup $K$.*

(i) *If $0 \to K \xrightarrow{i} G \xrightarrow{p} Q \to 1$ is a split extension, where $j\colon Q \to G$ satisfies $pj = 1_Q$, then $i(K) \cap j(Q) = \{0\}$ and $i(K) + j(Q) = G$.*

(ii) *Each $g \in G$ has a unique expression $g = i(a) + j(x)$, where $a \in K$ and $x \in Q$.*

(iii) *Let $K$ and $Q$ be subgroups of a group $G$ with $K \lhd G$. Then $G$ is a semidirect product of $K$ by $Q$ if and only if $K \cap Q = \{0\}$, $K + Q = G$, and each $g \in G$ has a unique expression $g = a + x$, where $a \in K$ and $x \in Q$.*

**Proof.**

(i) If $g \in i(K) \cap j(Q)$, then $g = i(a) = j(x)$ for $a \in K$ and $x \in Q$. Now $g = j(x)$ implies $p(g) = pj(x) = x$, while $g = i(a)$ implies $p(g) = pi(a) = 0$. Therefore, $x = 0$ and $g = j(x) = 0$.

If $g \in G$, then $p(g) = pjp(g)$ (because $pj = 1_Q$), and so $g - (jp(g)) \in \ker p = \operatorname{im} i$; hence, there is $a \in K$ with $g - (jp(g)) = i(a)$, and so $g = i(a) + j(pg) \in i(K) + j(Q)$.

(ii) Every element $g \in G$ has a factorization $g = i(a) + j(pg)$ because $G = i(K) + j(Q)$. To prove uniqueness, suppose that $i(a) + j(x) = i(b) + j(y)$, where $b \in K$ and $y \in Q$. Then $-i(b)+i(a) = j(y)-j(x) \in i(K)\cap j(Q) = \{0\}$, so that $i(a) = i(b)$ and $j(x) = j(y)$.

(iii) Necessity is the special case of (ii) when both $i$ and $j$ are inclusions. Conversely, each $g \in G$ has a unique factorization $g = ax$ for $a \in K$ and $x \in Q$; define $p\colon G \to Q$ by $p(ax) = x$. It is easy to check that $p$ is a surjective homomorphism with $\ker p = K$. •

Compare Proposition C-3.5 to Proposition A-4.83 in Part 1. A semidirect product is so called because a direct product $G$ of $K$ and $Q$ requires, in addition to $KQ = G$ and $K \cap Q = \{1\}$, that both subgroups $K$ and $Q$ be normal; here, only one subgroup must be normal.

**Definition.** If $K \leq G$ and $C \leq G$ satisfies $C \cap K = \{1\}$ and $KC = G$, then $C$ is called a ***complement*** of $K$.

In a semidirect product $G$, the subgroup $K$ is normal; on the other hand, the image $j(Q)$, which Proposition C-3.5 shows to be a complement of $K$, may not be normal. For example, if $G = S_3$ and $K = A_3 = \langle (1\ 2\ 3) \rangle$, we may take $C = \langle \tau \rangle$, where $\tau$ is any transposition in $S_3$; this example also shows that complements need not be unique. However, any two complements of $K$ are isomorphic, for any complement of $K$ is isomorphic to $G/K$.

The definition of semidirect product allows the kernel $K$ to be nonabelian, and such groups arise naturally. For example, the symmetric group $S_n$ is a semidirect product of the alternating group $A_n$ by $\mathbb{Z}_2$. In order to keep hypotheses uniform, however, let us assume in the text (except in some exercises) that $K$ is abelian, even though this assumption is not always needed.

**Example C-3.6.**

(i) A direct product $K \times Q$ is a semidirect product of $K$ by $Q$ (and also of $Q$ by $K$).

(ii) An abelian group $G$ is a semidirect product if and only if it is a direct product (usually called a direct sum), for every subgroup of an abelian group is normal.

(iii) The dihedral group $D_{2n}$ is a semidirect product of $\mathbb{Z}_n$ by $\mathbb{Z}_2$. If $D_{2n} = \langle a, b \rangle$, where $a^n = 1$, $b^2 = 1$, and $bab = a^{-1}$, then $\langle a \rangle$ is a normal subgroup having $\langle b \rangle$ as a complement.

(iv) Theorem C-2.111 says that every Frobenius group is a semidirect product of its Frobenius kernel by its Frobenius complement.

(v) Let $G = \mathbb{H}^\times$, the multiplicative group of nonzero quaternions. It is easy to see that if $\mathbb{R}^+$ is the multiplicative group of positive reals, then the *norm* $N\colon G \to \mathbb{R}^+$, given by

$$N(a + bi + cj + dk) = a^2 + b^2 + c^2 + d^2,$$

is a homomorphism and $G$ is a semidirect product of $\ker N$ by $\mathbb{R}^+$. In Exercise C-3.4 below, we will see that $\ker N \cong SU(2, \mathbb{C})$, the special unitary group.

(vi) Cyclic groups of prime power order are *not* semidirect products, for they cannot be a direct sum of two proper subgroups. ◀

**Definition.** Let $K$ be a $Q$-module. An extension $G$ of $K$ by $Q$ ***realizes the operators*** if, for all $x \in Q$ and $a \in K$, we have

$$xa = \ell(x) + a - \ell(x);$$

that is, the given scalar multiplication of $\mathbb{Z}Q$ on $K$ coincides with the scalar multiplication of Corollary C-3.4 arising from conjugation.

Here is the construction.

**Definition.** Let $Q$ be a group and let $K$ be a $Q$-module. Define

$$G = K \rtimes Q$$

to be the set of all ordered pairs $(a, x) \in K \times Q$ with the operation

$$(a, x) + (b, y) = (a + xb, xy).$$

Notice that $(a, 1) + (0, x) = (a, x)$ in $K \rtimes Q$.

**Proposition C-3.7.** *Given a group $Q$ and a $Q$-module $K$, then $G = K \rtimes Q$ is a semidirect product of $K$ by $Q$ that realizes the operators.*

**Proof.** We begin by proving that $G$ is a group. For associativity,

$$\begin{aligned}[(a, x) + (b, y)] + (c, z) &= (a + xb, xy) + (c, z)\\ &= (a + xb + (xy)c, (xy)z).\end{aligned}$$

On the other hand,

$$\begin{aligned}(a, x) + [(b, y) + (c, z)] &= (a, x) + (b + yc, yz)\\ &= (a + x(b + yc), x(yz)).\end{aligned}$$

Of course, $(xy)z = x(yz)$, because of associativity in $Q$. The first coordinates are also equal: since $K$ is a $Q$-module, we have

$$x(b + yc) = xb + x(yc) = xb + (xy)c.$$

Thus, the operation is associative. The identity element of $G$ is $(0, 1)$, for

$$(0, 1) + (a, x) = (0 + 1a, 1x) = (a, x),$$

and the inverse of $(a, x)$ is $(-x^{-1}a, x^{-1})$, for

$$(-x^{-1}a, x^{-1}) + (a, x) = (-x^{-1}a + x^{-1}a, x^{-1}x) = (0, 1).$$

Therefore, $G$ is a group, by Exercise A-4.27 on page 138 in Part 1.

Define a function $p\colon G \to Q$ by $p\colon (a, x) \mapsto x$. Since the only "twist" occurs in the first coordinate, $p$ is a surjective homomorphism with $\ker p = \{(a, 1)\colon a \in K\}$. If we define $i\colon K \to G$ by $i\colon a \mapsto (a, 1)$, then

$$0 \to K \xrightarrow{i} G \xrightarrow{p} Q \to 1$$

is an extension. Define $j\colon Q \to G$ by $j\colon x \mapsto (0, x)$. It is easy to see that $j$ is a homomorphism, for $(0, x) + (0, y) = (0, xy)$. Now $pjx = p(0, x) = x$, so that $pj = 1_Q$, and the extension splits; that is, $G$ is a semidirect product of $K$ by $Q$. Finally, $G$ realizes the operators: if $x \in Q$, then every lifting of $x$ has the form $\ell(x) = (b, x)$ for some $b \in K$, and

$$\begin{aligned}(b, x) + (a, 1) - (b, x) &= (b + xa, x) + (-x^{-1}b, x^{-1}) \\ &= (b + xa + x(-x^{-1}b), xx^{-1}) \\ &= (b + xa - b, 1) \\ &= (xa, 1). \quad \bullet\end{aligned}$$

We return to the multiplicative notation for a moment. In the next proof, the reader will see that the operation in $K \rtimes Q$ arises from the identity

$$(ax)(by) = a(xbx^{-1})xy.$$

**Theorem C-3.8.** *Let $K$ be an abelian group. If a group $G$ is a semidirect product of $K$ by a group $Q$, then there is a $Q$-module structure on $K$ so that $G \cong K \rtimes Q$.*

**Proof.** Regard $G$ as a group with normal subgroup $K$ that has $Q$ as a complement. We continue writing $G$ additively (even though it may not be abelian), and so we will now write its subgroup $Q$ additively as well. If $a \in K$ and $x \in Q$, define

$$xa = x + a - x;$$

that is, $xa$ is the conjugate of $a$ by $x$. By Proposition C-3.5, each $g \in G$ has a unique expression as $g = a + x$, where $a \in K$ and $x \in Q$. It follows that $\varphi\colon G \to K \rtimes Q$, defined by $\varphi\colon a + x \mapsto (a, x)$, is a bijection. We now show that $\varphi$ is an isomorphism.

$$\begin{aligned}\varphi((a + x) + (b + y)) &= \varphi(a + x + b + (-x + x) + y) \\ &= \varphi(a + (x + b - x) + x + y) \\ &= (a + xb, x + y).\end{aligned}$$

The definition of addition in $K \rtimes Q$ now gives

$$\begin{aligned}(a + xb, x + y) &= (a, x) + (b, y) \\ &= \varphi(a + x) + \varphi(b + y). \quad \bullet\end{aligned}$$

We now use semidirect products to construct some groups.

**Example C-3.9.** If $K = \langle a \rangle \cong \mathbb{Z}_3$, then an automorphism of $K$ is completely determined by the image of the generator $a$; either $a \mapsto a$ and the automorphism is $1_K$, or $a \mapsto 2a$. Therefore, $\mathrm{Aut}(K) \cong \mathbb{Z}_2$; let us denote its generator by $\varphi$, so that $\varphi(a) = 2a$ and $\varphi(2a) = a$; that is, $\varphi$ multiplies by 2. Let $Q = \langle x \rangle \cong \mathbb{Z}_4$, and define $\theta\colon Q \to \mathrm{Aut}(K)$ by $\theta_x = \varphi$; hence

$$xa = 2a \quad \text{and} \quad x2a = a.$$

The group

$$T = \mathbb{Z}_3 \rtimes \mathbb{Z}_4$$

is a group of order 12. If we define $s = (2a, x^2)$ and $t = (0, x)$, then the reader may check that

$$6s = 0 \quad \text{and} \quad 2t = 3s = 2(s + t).$$

There are four other groups of order 12. The Fundamental Theorem of Finite Abelian Groups says there are two abelian groups of this order: $\mathbb{Z}_{12} \cong \mathbb{Z}_3 \times \mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_6 \cong \mathbf{V} \times \mathbb{Z}_3$. Two nonabelian groups of order 12 are $A_4$ and $S_3 \times \mathbb{Z}_2$ (Exercise C-3.7 below says that $A_4 \ncong S_3 \times \mathbb{Z}_2$). The group $T$ just constructed is a new example, and Exercise C-3.17 on page 255 says that every group of order 12 is isomorphic to one of these. (Note that Exercise C-1.7 on page 14 states that $D_{12} \cong S_3 \times \mathbb{Z}_2$.) ◀

**Example C-3.10.** Let $p$ be a prime and let $K = \mathbb{Z}_p \oplus \mathbb{Z}_p$. Hence, $K$ is a vector space over $\mathbb{F}_p$, and so $\mathrm{Aut}(K) \cong \mathrm{GL}(K)$. We choose a basis $a, b$ of $K$, and this gives an isomorphism $\mathrm{Aut}(K) \cong \mathrm{GL}(2, p)$. Let $Q = \langle x \rangle$ be a cyclic group of order $p$.

Define $\theta\colon Q \to \mathrm{GL}(2, p)$ by

$$\theta\colon x^n \mapsto \begin{bmatrix} 1 & 0 \\ n & 1 \end{bmatrix}$$

for all $n \in \mathbb{Z}$. Thus,

$$xa = a + b \quad \text{and} \quad xb = b.$$

It is easy to check that the commutator $x+a-x-a = xa-a = b$, and so $G = K \rtimes Q$ is a group of order $p^3$ with $G = \langle a, b, x \rangle$; these generators satisfy relations

$$pa = pb = px = 0, \quad b = [x, a], \quad \text{and} \quad [b, a] = 0 = [b, x].$$

If $p$ is odd, then we have the nonabelian group of order $p^3$ and exponent $p$ in Proposition C-1.44. If $p = 2$, then $|G| = 8$, and the reader is asked to prove, in Exercise C-3.8 below, that $G \cong D_8$; that is, $D_8 \cong \mathbf{V} \rtimes \mathbb{Z}_2$. In Example C-3.6(iii), we saw that $D_8$ is a semidirect product of $\mathbb{Z}_4$ by $\mathbb{Z}_2$. Thus, $\mathbf{V} \rtimes \mathbb{Z}_2 \cong \mathbb{Z}_4 \rtimes \mathbb{Z}_2$, and so a group can have different factorizations as a semidirect product. ◀

**Example C-3.11.** Let $k$ be a field and let $k^\times$ be its multiplicative group. Now $k^\times$ acts on $k$ by multiplication (if $a \in k$ and $a \neq 0$, then the additive homomorphism $x \mapsto ax$ is an automorphism whose inverse is $x \mapsto a^{-1}x$). Therefore, the semidirect product $k \rtimes k^\times$ is defined. In particular, if $(b, a), (d, c) \in k \rtimes k^\times$, then

$$(b, a) + (d, c) = (ad + b, ac).$$

Recall that an *affine map* is a function $f\colon k \to k$ of the form $f\colon x \mapsto ax + b$, where $a, b \in k$ and $a \neq 0$, and the collection of all affine maps under composition

is the group $\mathrm{Aff}(1,k)$ (see Exercise A-4.36 on page 139 in Part 1). Note that if $g(x) = cx + d$, then

$$\begin{aligned}(f \circ g)(x) &= f(g(x)) \\ &= f(cx+d) \\ &= a(cx+d)+b \\ &= (ac)x + (ad+b).\end{aligned}$$

It is now easy to see that the function $\varphi\colon (b,a) \mapsto f$, where $f(x) = ax + b$, is an isomomorphism $k \rtimes k^\times \to \mathrm{Aff}(1,k)$. ◀

## Exercises

**In the first three exercises, the group $K$ need not be abelian; however, in all other exercises, it is assumed to be abelian.**

**C-3.1.** (i) Prove that $\mathrm{SL}(2,\mathbb{F}_5)$ is an extension of $\mathbb{Z}_2$ by $A_5$ which is not a semidirect product.

(ii) If $k$ is a field, prove that $\mathrm{GL}(n,k)$ is a semidirect product of $\mathrm{SL}(n,k)$ by $k^\times$.
**Hint.** A complement consists of all matrices $\mathrm{diag}\{1,\dots,1,a\}$ with $a \in k^\times$.

∗ **C-3.2.** Let $G$ be a group of order $mn$, where $\gcd(m,n) = 1$. Prove that a normal subgroup $K$ of order $m$ has a complement in $G$ if and only if there exists a subgroup $C \le G$ of order $n$.

∗ **C-3.3.** (**Baer**) (i) Prove that a group $G$ is injective[4] in the category **Groups** if and only if $G = \{1\}$.

**Hint.** Let $A$ be free with basis $\{x,y\}$, and let $B$ be the semidirect product $B = A \rtimes \langle z\rangle$, where $z$ is an element of order 2 that acts on $A$ by $zxz = y$ and $zyz = x$.

(ii) Use (i) and Exercise C-1.123 on page 123 to prove that **Groups** has no injective objects other than $\{1\}$.

∗ **C-3.4.** Let $SU(2)$ be the ***special unitary group*** consisting of all complex matrices $\left[\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right]$ of determinant 1 such that

$$a\overline{b} + c\overline{d} = 0, \quad a\overline{a} + b\overline{b} = 1, \quad c\overline{c} + d\overline{d} = 1.$$

If $S$ is the subgroup of $\mathbb{H}^\times$ in Example C-3.6(v), prove that $S \cong SU(2)$.

**Hint.** Use Exercise B-1.13 on page 281 in Part 1. There is a "polar decomposition" $h = rs$, where $r > 0$ and $s \in \ker N$, where $N\colon G \to \mathbb{R}^+$ is the norm.

**C-3.5.** Give an example of a split extension of groups

$$1 \to K \xrightarrow{i} G \xrightarrow{p} Q \to 1$$

for which there does not exist a homomorphism $q\colon G \to K$ with $qi = 1_K$. Compare with Exercise B-1.55 on page 310 in Part 1.

[4]The term *injective* had not yet been coined when R. Baer, who introduced the notion of injective module, proved this result. After recognizing that injective groups are duals of free groups, he jokingly called such groups ***fascist***, and he was pleased to note that they are trivial.

**C-3.6.** Prove that $\mathbf{Q}$, the group of quaternions, is not a semidirect product.

**Hint.** Recall that $\mathbf{Q}$ has a unique element of order 2.

∗ **C-3.7.** (i) Prove that $A_4 \not\cong S_3 \times \mathbb{Z}_2$.

**Hint.** Use Proposition A-4.67 in Part 1 saying that $A_4$ has no subgroup of order 6.

(ii) Prove that no two of the nonabelian groups of order 12, $A_4$, $S_3 \times \mathbb{Z}_2$, and $T$, are isomorphic. (See Example C-3.9.)

(iii) The affine group $\mathrm{Aff}(1, \mathbb{F}_4)$ (see Example C-3.11) is a nonabelian group of order 12. Is it isomorphic to $A_4$, $S_3 \times \mathbb{Z}_2$, or $T = \mathbb{Z}_3 \rtimes \mathbb{Z}_4$?

∗ **C-3.8.** Prove that the group $G$ of order 8 constructed in Example C-3.10 is isomorphic to $D_8$.

**C-3.9.** If $K$ and $Q$ are solvable groups, prove that a semidirect product of $K$ by $Q$ is also solvable.

**C-3.10.** Let $K$ be an abelian group, let $Q$ be a group, and let $\theta\colon Q \to \mathrm{Aut}(K)$ be a homomorphism. Prove that $K \rtimes Q \cong K \times Q$ if and only if $\theta$ is the trivial map ($\theta_x = 1_K$ for all $x \in Q$).

∗ **C-3.11.**

(i) If $K$ is cyclic of prime order $p$, prove that $\mathrm{Aut}(K)$ is cyclic of order $p - 1$.

(ii) Let $G$ be a group of order $pq$, where $p > q$ are primes. If $q \nmid (p - 1)$, prove that $G$ is cyclic. Conclude, for example, that every group of order 15 is cyclic.

∗ **C-3.12.** Let $G$ be an additive abelian $p$-group, where $p$ is prime.

(i) If $(m, p) = 1$, prove that the function $a \mapsto ma$ is an automorphism of $G$.

(ii) If $p$ is an odd prime and $G = \langle g \rangle$ is a cyclic group of order $p^2$, prove that $\varphi\colon G \to G$, given by $\varphi\colon a \mapsto 2a$, is the unique automorphism with $\varphi(pg) = 2pg$.

## C-3.3. General Extensions and Cohomology

We now proceed to the general extension problem: given a group $Q$ and an abelian group $K$, find all (not necessarily split) extensions

$$1 \to K \to G \to Q \to 1.$$

In light of our discussion of semidirect products, that is, of split extensions, it is reasonable to refine the problem by assuming that $K$ is a $Q$-module and then to seek all those extensions realizing the operators.

One way to describe a group $G$ is to give a multiplication table for it, that is, to list all its elements $a_1, a_2, \ldots$ and all products $a_i a_j$. Indeed, this is how we constructed semidirect products: the elements are all ordered pairs $(a, x)$ with $a \in K$ and $x \in Q$, and multiplication (really addition, because we have chosen to write $G$ additively) is

$$(a, x) + (b, y) = (a + xb, xy).$$

Schreier, in 1926, solved the extension problem in this way, and we present his solution in this section. The proof is not deep; rather, it merely involves manipulating and organizing a long series of elementary calculations.

**Remark.** We must point out, however, that Schreier's solution does not allow us to determine the number of nonisomorphic middle groups $G$. Of course, this last question has no easy answer. If a group $G$ has order $n$, then there are $n!$ different lists of its elements and hence an enormous number of different multiplication tables for $G$. Suppose now that $H$ is another group of order $n$. The problem of determining whether or not $G$ and $H$ are isomorphic is essentially the problem of comparing the families of multiplication tables of each to see if there is one for $G$ and one for $H$ that coincide. Using their sophisticated programs, Besche–Eick–O'Brien [**19**] show that there are exactly 49,487,365,422 nonisomorphic groups of order 1,024. ◀

Our strategy is to extract enough properties of a given extension $G$ that will suffice to reconstruct $G$. Thus, we may assume that $K$ is a $Q$-module, that $G$ is an extension of $K$ by $Q$ that realizes the operators, and that a transversal $\ell\colon Q \to G$ has been chosen. With this initial data, we see that each $g \in G$ has a unique expression of the form

$$g = a + \ell(x), \quad a \in K \quad \text{and} \quad x \in Q;$$

this follows from $G$ being the disjoint union of the cosets $K + \ell(x)$. Furthermore, if $x, y \in Q$, then $\ell(x) + \ell(y)$ and $\ell(xy)$ are both representatives of the same coset (we do not say these representatives are the same!), and so there is an element $f(x, y) \in K$ such that

$$\ell(x) + \ell(y) = f(x, y) + \ell(xy).$$

**Definition.** Given a lifting $\ell\colon Q \to G$, with $\ell(1) = 0$, of an extension $G$ of $K$ by $Q$, then a ***factor set***[5] (or *cocycle*) is a function $f\colon Q \times Q \to K$ such that

$$\ell(x) + \ell(y) = f(x, y) + \ell(xy)$$

for all $x, y \in Q$.

It is natural to choose liftings with $\ell(1) = 0$, and so we have incorporated this condition into the definition of factor set; our factor sets are often called ***normalized factor sets***.

Of course, a factor set depends on the choice of lifting $\ell$. When $G$ is a split extension, then there exists a lifting that is a homomorphism; the corresponding factor set is identically 0. Therefore, we can regard a factor set as the obstruction to a lifting being a homomorphism; that is, factor sets describe how an extension differs from being a split extension.

[5] If we switch to multiplicative notation, we see that a factor set occurs in the factorization $\ell(x)\ell(y) = f(x, y)\ell(xy)$.

**Proposition C-3.12.** *Let $Q$ be a group, $K$ a $Q$-module, and $0 \to K \to G \to Q \to 1$ an extension realizing the operators. If $\ell\colon Q \to G$ is a lifting with $\ell(1) = 0$ and $f\colon Q \times Q \to K$ is the corresponding factor set, then*

(i) *for all $x, y \in Q$,*
$$f(1, y) = 0 = f(x, 1);$$

(ii) *the **cocycle identity** holds: for all $x, y, z \in Q$, we have*
$$f(x, y) + f(xy, z) = xf(y, z) + f(x, yz).$$

**Proof.** Set $x = 1$ in the equation that defines $f(x, y)$,
$$\ell(x) + \ell(y) = f(x, y) + \ell(xy),$$
to see that $\ell(y) = f(1, y) + \ell(y)$ (since $\ell(1) = 0$, by our new assumption), and hence $f(1, y) = 0$. Setting $y = 1$ gives the other equation of (i).

The cocycle identity follows from associativity in $G$. For all $x, y, z \in Q$, we have
$$\begin{aligned}[\ell(x) + \ell(y)] + \ell(z) &= f(x, y) + \ell(xy) + \ell(z)\\ &= f(x, y) + f(xy, z) + \ell(xyz).\end{aligned}$$
On the other hand,
$$\begin{aligned}\ell(x) + [\ell(y) + \ell(z)] &= \ell(x) + f(y, z) + \ell(yz)\\ &= xf(y, z) + \ell(x) + \ell(yz)\\ &= xf(y, z) + f(x, yz) + \ell(xyz). \quad \bullet\end{aligned}$$

It is more interesting that the converse is true. The next result generalizes the construction of $K \rtimes Q$ in Proposition C-3.7.

**Theorem C-3.13.** *Given a group $Q$ and a $Q$-module $K$, a function $f\colon Q \times Q \to K$ is a factor set if and only if it satisfies the cocycle identity*[6]
$$xf(y, z) - f(xy, z) + f(x, yz) - f(x, y) = 0$$
*and $f(1, y) = 0 = f(x, 1)$ for all $x, y, z \in Q$.*

*More precisely, there is an extension $G$ of $K$ by $Q$ realizing the operators, and there is a transversal $\ell\colon Q \to G$ whose corresponding factor set is $f$.*

**Proof.** Necessity is Proposition C-3.12. For the converse, define $G$ to be the set of all ordered pairs $(a, x)$ in $K \times Q$ equipped with the operation
$$(a, x) + (b, y) = (a + xb + f(x, y), xy).$$
(Thus, if $f$ is identically 0, then $G = K \rtimes Q$.) The proof that $G$ is a group is similar to the proof of Proposition C-3.7. To prove associativity, consider
$$\begin{aligned}\big((a, x) + (b, y)\big) + (c, z) &= (a + xb + f(x, y), xy) + (c, z)\\ &= (a + xb + f(x, y) + xyc + f(xy, z), xyz)\end{aligned}$$

[6] Written as an alternating sum, this identity is reminiscent of the formulas describing geometric cycles as described in Section C-3.1.

and

$$\begin{aligned}(a,x)+\big((b,y)+(c,z)\big) &= (a,x)+(b+yc+f(y,z),yz)\\ &= (a+xb+xyc+xf(y,z)+f(x,yz),xyz).\end{aligned}$$

The cocycle identity shows that these elements are equal.

We let the reader prove that the identity is $(0,1)$ and the inverse of $(a,x)$ is

$$-(a,x)=(-x^{-1}a-x^{-1}f(x,x^{-1}),x^{-1}).$$

Define $p\colon G\to Q$ by $p\colon (a,x)\mapsto x$. Because the only "twist" occurs in the first coordinate, it is easy to see that $p$ is a surjective homomorphism with $\ker p = \{(a,1): a\in K\}$. If we define $i\colon K\to G$ by $i\colon a\mapsto (a,1)$, then we have an extension $0\to K\xrightarrow{i} G\xrightarrow{p} Q\to 1$.

To see that this extension realizes the operators, we must show, for every lifting $\ell$, that $xa=\ell(x)+a-\ell(x)$ for all $a\in K$ and $x\in Q$. Now $\ell(x)=(b,x)$ for some $b\in K$ and

$$\begin{aligned}\ell(x)+(a,1)-\ell(x) &= (b,x)+(a,1)-(b,x)\\ &= (b+xa,x)+(-x^{-1}b-x^{-1}f(x,x^{-1}),x^{-1})\\ &= (b+xa+x[-x^{-1}b-x^{-1}f(x,x^{-1})]+f(x,x^{-1}),1)\\ &= (xa,1).\end{aligned}$$

Finally, we must show that $f$ is the factor set determined by $\ell$. Choose the lifting $\ell(x)=(0,x)$ for all $x\in Q$. The factor set $F$ determined by $\ell$ is defined by

$$\begin{aligned}F(x,y) &= \ell(x)+\ell(y)-\ell(xy)\\ &= (0,x)+(0,y)-(0,xy)\\ &= (f(x,y),xy)+(-(xy)^{-1}f(xy,(xy)^{-1}),(xy)^{-1})\\ &= (f(x,y)+xy[-(xy)^{-1}f(xy,(xy)^{-1})]+f(xy,(xy)^{-1}),xy(xy)^{-1})\\ &= (f(x,y),1). \quad \bullet\end{aligned}$$

The next result shows that we have found all the extensions of a $Q$-module $K$ by a group $Q$.

**Definition.** Given a group $Q$, a $Q$-module $K$, and a factor set $f$, let $G(K,Q,f)$ denote the middle group of the extension of $K$ by $Q$ constructed in Theorem C-3.13.

**Theorem C-3.14.** *Let $Q$ be a group, let $K$ be a $Q$-module, and let $G$ be an extension of $K$ by $Q$ realizing the operators. Then there exists a factor set $f\colon Q\times Q\to K$ with*

$$G\cong G(K,Q,f).$$

**Proof.** Let $\ell\colon Q\to G$ be a lifting, and let $f\colon Q\times Q\to K$ be the corresponding factor set: that is, for all $x,y\in Q$, we have

$$\ell(x)+\ell(y)=f(x,y)+\ell(xy).$$

Since $G$ is the disjoint union of the cosets, $G = \bigcup_{x\in Q} K + \ell(x)$, each $g \in G$ has a unique expression $g = a + \ell(x)$ for $a \in K$ and $x \in Q$. Uniqueness implies that the function $\varphi\colon G \to G(K, Q, f)$, given by

$$\varphi\colon g = a + \ell(x) \mapsto (a, x),$$

is a well-defined bijection. We now show that $\varphi$ is an isomorphism:

$$\begin{aligned}\varphi(a + \ell(x) + b + \ell(y)) &= \varphi(a + \ell(x) + b - \ell(x) + \ell(x) + \ell(y))\\ &= \varphi(a + xb + \ell(x) + \ell(y))\\ &= \varphi(a + xb + f(x, y) + \ell(xy))\\ &= (a + xb + f(x, y), xy)\\ &= (a, x) + (b, y)\\ &= \varphi(a + \ell(x)) + \varphi(b + \ell(y)). \quad \bullet\end{aligned}$$

**Remark.** For later use, note that if $a \in K$, then $\varphi(a) = \varphi(a + \ell(1)) = (a, 1)$ and, if $x \in Q$, then $\varphi(\ell(x)) = (0, x)$. This would not be so, had we chosen a lifting $\ell$ with $\ell(1) \neq 0$. ◀

We have now described all extensions in terms of factor sets, but factor sets are determined by liftings. Any extension has many different liftings, and so our description, which depends on a choice of lifting, must have repetitions.

**Lemma C-3.15.** *Given a group $Q$ and a $Q$-module $K$, let $G$ be an extension of $K$ by $Q$ realizing the operators. Let $\ell$ and $\ell'$ be liftings that give rise to factor sets $f$ and $f'$, respectively. Then there exists a function $h\colon Q \to K$ with $h(1) = 0$ and, for all $x, y \in Q$,*

$$f'(x, y) - f(x, y) = xh(y) - h(xy) + h(x).$$

**Proof.** For each $x \in Q$, both $\ell(x)$ and $\ell'(x)$ lie in the same coset of $K$ in $G$, and so there exists an element $h(x) \in K$ with

$$\ell'(x) = h(x) + \ell(x).$$

Since $\ell(1) = 0 = \ell'(1)$, we have $h(1) = 0$. The main formula is derived as follows:

$$\begin{aligned}\ell'(x) + \ell'(y) &= [h(x) + \ell(x)] + [h(y) + \ell(y)]\\ &= h(x) + xh(y) + \ell(x) + \ell(y),\end{aligned}$$

because $G$ realizes the operators. The equations continue,

$$\begin{aligned}\ell'(x) + \ell'(y) &= h(x) + xh(y) + f(x, y) + \ell(xy)\\ &= h(x) + xh(y) + f(x, y) - h(xy) + \ell'(xy).\end{aligned}$$

By definition, $f'$ satisfies $\ell'(x) + \ell'(y) = f'(x, y) + \ell'(xy)$. Therefore,

$$f'(x, y) = h(x) + xh(y) + f(x, y) - h(xy),$$

and so

$$f'(x, y) - f(x, y) = xh(y) - h(xy) + h(x). \quad \bullet$$

**Definition.** Given a group $Q$ and a $Q$-module $K$, a function $g\colon Q \times Q \to K$ is called a ***coboundary*** if there exists a function $h\colon Q \to K$ with $h(1) = 0$ such that, for all $x, y \in Q$,

$$g(x, y) = xh(y) - h(xy) + h(x).$$

The term *coboundary* arises because its formula is an alternating sum analogous to the formula for geometric boundaries that we described in Section C-3.1.

We have just shown that if $f$ and $f'$ are factor sets of an extension $G$ that arise from different liftings, then $f' - f$ is a coboundary.

**Definition.** Given a group $Q$ and a $Q$-module $K$, define

$$Z^2(Q, K) = \{\text{all factor sets } f\colon Q \times Q \to K\}$$

and

$$B^2(Q, K) = \{\text{all coboundaries } g\colon Q \times Q \to K\}.$$

**Proposition C-3.16.** *Given a group $Q$ and a $Q$-module $K$, then $Z^2(Q, K)$ is an abelian group with operation pointwise addition,*

$$f + f'\colon (x, y) \mapsto f(x, y) + f'(x, y),$$

*and $B^2(Q, K)$ is a subgroup of $Z^2(Q, K)$.*

**Proof.** To see that $Z^2$ is a group, it suffices to prove that $f - f'$ satisfies the two identities in Proposition C-3.12. This is obvious: just subtract the equations for $f$ and $f'$.

To see that $B^2$ is a subgroup of $Z^2$, we must first show that every coboundary $g$ is a factor set, that is, that $g$ satisfies the two identities in Proposition C-3.12. This, too, is routine and is left to the reader. Next, we must show that $B^2$ is a nonempty subset; but the zero function, $g(x, y) = 0$ for all $x, y \in Q$, is clearly a coboundary. Finally, we show that $B^2$ is closed under subtraction. If $h, h'\colon Q \to K$ show that $g$ and $g'$ are coboundaries, that is, $g(x, y) = xh(y) - h(xy) + h(x)$ and $g'(x, y) = xh'(y) - h'(xy) + h'(x)$, then

$$(g - g')(x, y) = x(h - h')(y) - (h - h')(xy) + (h - h')(x). \quad \bullet$$

A given extension has many liftings and, hence, many factor sets, but the difference of any two of these factor sets is a coboundary. Therefore, the following quotient group suggests itself.

**Definition.** Given a group $Q$ and a $Q$-module $K$, their ***second cohomology group*** is defined by

$$H^2(Q, K) = Z^2(Q, K)/B^2(Q, K).$$

**Definition.** Given a group $Q$ and a $Q$-module $K$, two extensions $G$ and $G'$ of $K$ by $Q$ that realize the operators are called ***equivalent*** if there is a factor set $f$ of $G$ and a factor set $f'$ of $G'$ so that $f' - f$ is a coboundary.

**Proposition C-3.17.** *Given a group $Q$ and a $Q$-module $K$, two extensions $G$ and $G'$ of $K$ by $Q$ that realize the operators are equivalent if and only if there exists an isomorphism $\gamma\colon G \to G'$ making the following diagram commute*:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & K & \xrightarrow{i} & G & \xrightarrow{p} & Q & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle 1_K} & & \downarrow{\scriptstyle \gamma} & & \downarrow{\scriptstyle 1_Q} & & \\ 0 & \longrightarrow & K & \xrightarrow{i'} & G' & \xrightarrow{p'} & Q & \longrightarrow & 1. \end{array}$$

**Remark.** A diagram chase shows that any homomorphism $\gamma$ making the diagram commute is necessarily an isomorphism. ◀

**Proof.** Assume that the two extensions are equivalent. We begin by setting up notation. Let $\ell\colon Q \to G$ and $\ell'\colon Q \to G'$ be liftings, and let $f, f'$ be the corresponding factor sets; that is, for all $x, y \in Q$, we have

$$\ell(x) + \ell(y) = f(x, y) + \ell(xy),$$

with a similar equation for $f'$ and $\ell'$. Equivalence means that there is a function $h\colon Q \to K$ with $h(1) = 0$ and

$$f(x, y) - f'(x, y) = xh(y) - h(xy) + h(x)$$

for all $x, y \in Q$. Since $G = \bigcup_{x\in Q} K + \ell(x)$ is a disjoint union, each $g \in G$ has a unique expression $g = a + \ell(x)$ for $a \in K$ and $x \in Q$; similarly, each $g' \in G'$ has a unique expression $g' = a + \ell'(x)$.

This part of the proof generalizes that of Theorem C-3.14. Define $\gamma\colon G \to G'$ by

$$\gamma(a + \ell(x)) = a + h(x) + \ell'(x).$$

This function makes the diagram commute. If $a \in K$, then

$$\gamma(a) = \gamma(a + \ell(1)) = a + h(1) + \ell'(1) = a;$$

furthermore,

$$p'\gamma(a + \ell(x)) = p'(a + h(x) + \ell'(x)) = x = p(a + \ell(x)).$$

Finally, $\gamma$ is a homomorphism:

$$\begin{aligned} \gamma\big((a + \ell(x)) + (b + \ell(y))\big) &= \gamma(a + xb + f(x, y) + \ell(xy)) \\ &= a + xb + f(x, y) + h(xy) + \ell'(xy), \end{aligned}$$

while

$$\begin{aligned} \gamma(a + \ell(x)) + \gamma(b + \ell(y)) &= \big(a + h(x) + \ell'(x)\big) + \big(b + h(y) + \ell'(y)\big) \\ &= a + h(x) + xb + xh(y) + f'(x, y) + \ell'(xy) \\ &= a + xb + \big(h(x) + xh(y) + f'(x, y)\big) + \ell'(xy) \\ &= a + xb + f(x, y) + h(xy) + \ell'(xy). \end{aligned}$$

We have used the given equation for $f - f'$ (remember that the terms other than $\ell'(xy)$ all lie in the abelian group $K$, and so they may be rearranged).

Conversely, assume that there exists an isomorphism $\gamma$ making the diagram commute, so that $\gamma(a) = a$ for all $a \in K$ and

$$x = p(\ell(x)) = p'\gamma(\ell(x))$$

for all $x \in Q$. It follows that $\gamma\ell\colon Q \to G'$ is a lifting. Applying $\gamma$ to the equation $\ell(x)+\ell(y) = f(x,y)+\ell(xy)$ that defines the factor set $f$, we see that $\gamma f$ is the factor set determined by the lifting $\gamma\ell$. But $\gamma f(x,y) = f(x,y)$ for all $x, y \in Q$ because $f(x,y) \in K$. Therefore, $f$ is also a factor set of the second extension. On the other hand, if $f'$ is any other factor set for the second extension, then Lemma C-3.15 shows that $f - f' \in B^2$; that is, the extensions are equivalent. •

We say that the isomorphism $\gamma$ in Proposition C-3.17 ***implements*** the equivalence. The remark after Theorem C-3.14 shows that the isomorphism $\gamma\colon G \to G(K,Q,f)$ implements an equivalence of extensions.

**Example C-3.18.** If two extensions of $K$ by $Q$ realizing the operators are equivalent, then their middle groups are isomorphic. However, the converse is false: we give an example of two inequivalent extensions with isomorphic middle groups. Let $p$ be an odd prime, and consider the following diagram:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & K & \xrightarrow{i} & G & \xrightarrow{\pi} & Q & \longrightarrow & 1 \\
 & & \downarrow{\scriptstyle 1_K} & & \vdots & & \downarrow{\scriptstyle 1_Q} & & \\
0 & \longrightarrow & K & \xrightarrow{i'} & G' & \xrightarrow{\pi'} & Q & \longrightarrow & 1.
\end{array}$$

Define $K = \langle a \rangle$, a cyclic group of order $p$, $G = \langle g \rangle = G'$, a cyclic group of order $p^2$, and $Q = \langle x \rangle$, where $x = g + K$. In the top row, define $i(a) = pg$ and $\pi$ to be the natural map; in the bottom row define $i'(a) = 2pg$ and $\pi'$ to be the natural map. Note that $i'$ is injective because $p$ is odd.

Suppose there is an isomorphism $\gamma\colon G \to G'$ making the diagram commute. Commutativity of the first square implies $\gamma(pa) = 2pa$, and this forces $\gamma(g) = 2g$, by Exercise C-3.12 on page 236; commutativity of the second square gives $g + K = 2g + K$; that is, $g \in K$. We conclude that the two extensions are not equivalent. ◀

The next theorem summarizes the calculations in this section.

**Theorem C-3.19 (Schreier).** *Let $Q$ be a group, let $K$ be a $Q$-module, and let $e(Q,K)$ denote the family of all the equivalence classes of extensions of $K$ by $Q$ realizing the operators. There is a bijection*

$$\varphi\colon H^2(Q,K) \to e(Q,K)$$

*that takes* 0 *to the class of the split extension.*

**Proof.** Denote the equivalence class of an extension

$$0 \to K \to G \to Q \to 1$$

by $[G]$. Define $\varphi\colon H^2(Q,K) \to e(Q,K)$ by

$$\varphi\colon f + B^2 \mapsto [G(K,Q,f)],$$

where $f$ is a factor set of the extension and the target extension is that constructed in Theorem C-3.13.

First, $\varphi$ is a well-defined injection: $f$ and $g$ are factor sets with $f+B^2 = g+B^2$ if and only if $[G(K,Q,f)] = [G(K,Q,g)]$, by Proposition C-3.17. To see that $\varphi$ is a surjection, let $[G] \in e(Q,K)$. By Theorem C-3.14 and the remark following it, $[G] = [G(K,Q,f)]$ for some factor set $f$, and so $[G] = \varphi(f+B^2)$. Finally, the zero factor set corresponds to the semidirect product. •

If $H$ is a group and if there is a bijection $\varphi\colon H \to X$, where $X$ is a set, then there is a unique operation defined on $X$ making $X$ a group and $\varphi$ an isomorphism: given $x, y \in X$, there are $g, h \in H$ with $x = \varphi(g)$ and $y = \varphi(h)$, and we define $xy = \varphi(gh)$. In particular, there is a way to add two equivalence classes of extensions; it is called ***Baer sum*** (see Section C-3.8).

**Corollary C-3.20.** *If $Q$ is a group, $K$ is a $Q$-module, and $H^2(Q,K) = \{0\}$, then every extension of $K$ by $Q$ realizing the operators is a semidirect product.*

**Proof.** By the theorem, $e(Q,K)$ has only one element; since the split extension always exists, this one element must be the equivalence class of the split extension. Therefore, every extension of $K$ by $Q$ realizing the operators is split, and so its middle group is a semidirect product. •

We now apply Schreier's Theorem.

**Theorem C-3.21.** *Let $G$ be a finite group of order $mn$, where $\gcd(m,n) = 1$. If $K$ is an abelian normal subgroup of order $m$, then $K$ has a complement and $G$ is a semidirect product.*

**Proof.** Define $Q = G/K$. By Corollary C-3.20, it suffices to prove that every factor set $f\colon Q \times Q \to K$ is a coboundary. Define $\sigma\colon Q \to K$ by

$$\sigma(x) = \sum_{y \in Q} f(x,y);$$

$\sigma$ is well-defined because $Q$ is finite and $K$ is abelian. Now sum the cocycle identity

$$xf(y,z) - f(xy,z) + f(x,yz) - f(x,y) = 0$$

over all $z \in Q$ to obtain

$$x\sigma(y) - \sigma(xy) + \sigma(x) = nf(x,y)$$

(as $z$ varies over all of $Q$, so does $yz$). Since $\gcd(m,n) = 1$, there are integers $s$ and $t$ with $sm + tn = 1$. Define $h\colon Q \to K$ by

$$h(x) = t\sigma(x).$$

Note that $h(1) = 0$ and

$$xh(y) - h(xy) + h(x) = f(x,y) - msf(x,y).$$

But $sf(x,y) \in K$, and so $msf(x,y) = 0$. Therefore, $f$ is a coboundary. •

**Remark.** Recall P. Hall's Theorem C-1.57: if $G$ is a finite solvable group of order $ab$, where $\gcd(a,b) = 1$, then $G$ has a subgroup of order $a$ and any two such subgroups are conjugate. In particular, in a solvable group, every (not necessarily normal) Sylow subgroup has a complement. Because of this theorem, a (not necessarily normal) subgroup $H$ of a finite group $G$ is called a ***Hall subgroup*** if $\gcd(|H|, [G:H]) = 1$. Thus, Theorem C-3.21 is often stated as every normal Hall subgroup of an arbitrary finite group has a complement. ◀

We now remove the hypothesis that $K$ be abelian.

**Theorem C-3.22 (Schur–Zassenhaus[7] Lemma).** *Let $G$ be a finite group of order $mn$, where $\gcd(m,n) = 1$. If $K$ is a normal subgroup of order $m$, then $K$ has a complement and $G$ is a semidirect product.*

**Proof.** By Exercise C-3.2 on page 235, it suffices to prove that $G$ contains a subgroup of order $n$; we prove the existence of such a subgroup by induction on $m \geq 1$. Of course, the base step $m = 1$ is true.

Suppose that there is a proper subgroup $T$ of $K$ with $\{1\} \subseteq T \lhd G$. Then $K/T \lhd G/T$ and $(G/T)/(K/T) \cong G/K$ has order $n$. Since $T \subseteq K$, we have $|K/T| < |K| = m$, and so the inductive hypothesis provides a subgroup $N/T \subseteq G/T$ with $|N/T| = n$. Now $|N| = n|T|$, where $\gcd(|T|, n) = 1$ (because $|T|$ is a divisor of $|K| = m$), so that $T$ is a normal subgroup of $N$ whose order and index are relatively prime. Since $|T| < |K| = m$, the inductive hypothesis provides a subgroup $C$ of $N$ (which is obviously a subgroup of $G$) of order $n$.

We may now assume that $K$ is a minimal normal subgroup of $G$; that is, there is no normal subgroup $T$ of $G$ with $\{1\} \subsetneq T \subsetneq K$. Let $p$ be a prime divisor of $|K|$ and let $P$ be a Sylow $p$-subgroup of $K$. By the Frattini Argument (Lemma C-1.56), we have $G = KN_G(P)$. Therefore,

$$\begin{aligned} G/K &= KN_G(P)/K \\ &\cong N_G(P)/(K \cap N_G(P)) \\ &= N_G(P)/N_K(P). \end{aligned}$$

Hence, $|N_K(P)|n = |N_K(P)||G/K| = |N_G(P)|$. If $N_G(P)$ is a proper subgroup of $G$, then $|N_K(P)| < m$, and induction provides a subgroup of $N_G(P) \subseteq G$ of order $n$. Therefore, we may assume that $N_G(P) = G$; that is, $P \lhd G$.

Since $\{1\} \subsetneq P \subseteq K$ and $P$ is normal in $G$, we must have $P = K$, because $K$ is a minimal normal subgroup. But $P$ is a $p$-group, and so its center, $Z(P)$, is nontrivial. By Exercise C-1.36 on page 32, we have $Z(P) \lhd G$, and so $Z(P) = P$, again because $P = K$ is a minimal normal subgroup of $G$. It follows that $P$ is abelian, and we have reduced the problem to Theorem C-3.21. •

**Corollary C-3.23.** *If a finite group $G$ has a normal Sylow $p$-subgroup $P$, for some prime divisor $p$ of $|G|$, then $G$ is a semidirect product; more precisely, $P$ has a complement.*

[7] Schur proved this theorem, in 1904, for the special case $Q$ cyclic. Zassenhaus, in 1938, proved the theorem for arbitrary finite $Q$.

**Proof.** The order and index of a Sylow subgroup are relatively prime. •

There is another part of the Schur–Zassenhaus Lemma that we have not stated: if $K$ is a normal subgroup of $G$ whose order and index are relatively prime, then any two complements of $K$ are conjugate subgroups. We are now going to see that there is an analog of $H^2(K, Q)$ whose vanishing implies conjugacy of complements when $K$ is abelian. This group, $H^1(K, Q)$, arises, as did $H^2(K, Q)$, from a series of elementary calculations.

We begin with a computational lemma. Let $Q$ be a group, let $K$ be a $Q$-module, and let $0 \to K \to G \to Q \to 1$ be a split extension. Choose a lifting $\ell\colon Q \to G$, so that every element $g \in G$ has a unique expression of the form

$$g = a + \ell x,$$

where $a \in K$ and $x \in Q$.

**Definition.** An automorphism $\varphi$ of a group $G$ ***stabilizes*** an extension $0 \to K \to G \to Q \to 1$ if the following diagram commutes:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & K & \xrightarrow{i} & G & \xrightarrow{p} & Q & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle 1_K} & & \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle 1_Q} & & \\ 0 & \longrightarrow & K & \xrightarrow{i} & G & \xrightarrow{p} & Q & \longrightarrow & 1. \end{array}$$

The set of all stabilizing automorphisms of an extension of $K$ by $Q$, where $K$ is a $Q$-module, forms a group under composition, denoted by

$$\mathrm{Stab}(Q, K).$$

Note that a stabilizing automorphism is an isomorphism that implements an equivalence of an extension with itself. We shall see, in Proposition C-3.26, that $\mathrm{Stab}(Q, K)$ does not depend on the extension.

**Proposition C-3.24.** *Let $Q$ be a group, let $K$ be a $Q$-module, and let*

$$0 \to K \xrightarrow{i} G \xrightarrow{p} Q \to 1$$

*be a split extension, where $i$ is the inclusion.*

(i) *If $\ell\colon Q \to G$ is a lifting, then every stabilizing automorphism $\varphi\colon G \to G$ has the unique form*

$$\varphi(a + \ell x) = a + d(x) + \ell x, \tag{1}$$

*where $d(x) \in K$ is independent of the choice of lifting $\ell$.*

(ii) Eq. (1) *defines a stabilizing automorphism if and only if, for all $x, y \in Q$, the function $d\colon Q \to K$ satisfies*

$$d(xy) = d(x) + xd(y). \tag{2}$$

**Proof.**

(i) Remember, since $\ell$ is a lifting of $Q$, that $\operatorname{im}\ell$ is a transversal of $K$ in $G$, and so every $g \in G$ has a unique expression $g = a + \ell x$, where $a \in K$ and $x \in Q$. Now $\varphi(a + \ell x) = \varphi(a) + \varphi(\ell x)$. As $\varphi$ is stabilizing, $\varphi i = i$ and $p\varphi = p$. Since $i\colon K \to G$ is the inclusion (which is merely a convenience to allow us to write $a$ instead of $i(a)$), we have $\varphi(a) = a$ for all $a \in K$. To use the second constraint $p\varphi = p$, suppose that $\varphi(\ell x) = d(x) + \ell y$ for some $d(x) \in K = \ker p$ and $y \in Q$. Now

$$x = p(\ell x) = p\varphi(\ell x) = p(d(x) + \ell y) = y,$$

because $p(\ell y) = y$, for this is the definition of lifting. Thus, $x = y$, $\ell x = \ell y$, and Eq. (1) holds. Uniqueness of Eq. (1) follows from uniqueness of the expression writing each element of $G$ in the form $a + \ell x$ (which merely says that $G$ is the disjoint union of the cosets of $K$).

We now show that $d$ is independent of the choice of lifting. Suppose that $\ell'\colon Q \to G$ is another lifting and that $\varphi(\ell' x) = d'(x) + \ell' x$ for some $d'(x) \in K$. There is $k \in K$ with $\ell' x = k + \ell x$, for $p\ell' x = x = p\ell x$. Therefore, $k + \ell x - \ell' x = 0$, and

$$\begin{aligned} d'(x) &= \varphi(\ell' x) - \ell' x \\ &= \varphi(k + \ell x) - \ell' x \\ &= k + d(x) + \ell x - \ell' x \\ &= d(x). \end{aligned}$$

(ii) We now show that Eq. (2) holds for $d$. Since $d$ is independent of the choice of lifting $\ell$ and since the extension splits, we may assume that $\ell$ is a homomorphism: $\ell x + \ell y = \ell(xy)$. We compute $\varphi(\ell x + \ell y)$ in two ways.

On the one hand,

$$\varphi(\ell x + \ell y) = \varphi(\ell(xy)) = d(xy) + \ell(xy).$$

On the other hand, since the extension realizes the operators,

$$\begin{aligned} \varphi(\ell x + \ell y) &= \varphi(\ell x) + \varphi(\ell y) \\ &= d(x) + \ell x + d(y) + \ell y \\ &= d(x) + xd(y) + \ell(xy). \end{aligned}$$

Canceling $\ell(xy)$ gives the result.

Conversely, we must show that the function $\varphi\colon G \to G$ defined in Eq. (1) is a stabilizing automorphism. As in Proposition C-3.7, we view the semidirect product $G$ as all ordered pairs $(a, x) \in K \rtimes Q$ with operation $(a, x) + (b, y) = (a + xb, xy)$. With this notation,

$$\varphi((a, x)) = (a + d(x), x), \quad ia = (a, 1), \text{ and } p(a, x) = x.$$

The Five Lemma shows, if $\varphi$ is a homomorphism, that it is an automorphism. Now

$$\begin{aligned} \varphi((a, x) + (b, y)) &= \varphi((a + xb, xy)) \\ &= (a + xb + d(xy), xy). \end{aligned}$$

On the other hand, since elements in $K$ commute,

$$\begin{aligned}\varphi((a,x))+\varphi((b,y)) &= (a+d(x),x)+(b+d(y),y)\\ &= (a+d(x)+x(b+d(y)),xy)\\ &= (a+d(x)+xb+xd(y),xy)\\ &= (a+xb+d(x)+xd(y),xy).\end{aligned}$$

It follows that Eq. (2) holds.

To see that $\varphi$ is stabilizing, observe that $\varphi((a,1)) = (a+d(1),1) = (a,1)$, because $d(1) = 0$ (for $d(1) = d(1\cdot 1) = d(1) + 1d(1)$). Also, $p\varphi((a,x)) = p(a+d(x),x) = x = p(a,x)$ •

We give a name to functions like $d$.

**Definition.** Let $Q$ be a group and let $K$ be a $Q$-module. A ***derivation***[8] (or ***crossed homomorphism***) is a function $d\colon Q \to K$ such that

$$d(xy) = xd(y) + d(x).$$

The set of all derivations, $\mathrm{Der}(Q,K)$, is an abelian group under pointwise addition (if $K$ is a trivial $Q$-module, then $\mathrm{Der}(Q,K) = \mathrm{Hom}(Q,K)$).

We saw, in the proof of Proposition C-3.24(ii), that $d(1) = 0$.

**Example C-3.25.**

(i) If $Q$ is a group and $K$ is a $Q$-module, then a function $u\colon Q \to K$ of the form $u(x) = xa_0 - a_0$, where $a_0 \in K$, is a derivation:

$$\begin{aligned}u(x)+xu(y) &= xa_0 - a_0 + x(ya_0 - a_0)\\ &= xa_0 - a_0 + xya_0 - xa_0\\ &= xya_0 - a_0\\ &= u(xy).\end{aligned}$$

A derivation $u$ of the form $u(x) = xa_0 - a_0$ is called a ***principal derivation***.

If the action of $Q$ on $K$ is conjugation, $xa = x + a - x$, then

$$xa_0 - a_0 = x + a_0 - x - a_0;$$

that is, $xa_0 - a_0$ is the commutator of $x$ and $a_0$.

(ii) It is easy to check that the set $\mathrm{PDer}(Q,K)$ of all the principal derivations is a subgroup of $\mathrm{Der}(Q,K)$. ◀

Recall that $\mathrm{Stab}(Q,K)$ denotes the group of all the stabilizing automorphisms of an extension of $K$ by $Q$.

**Proposition C-3.26.** *If $Q$ is a group, $K$ is a $Q$-module, and $0 \to K \to G \to Q \to 1$ is a split extension, then there is an isomorphism $\mathrm{Stab}(Q,K) \to \mathrm{Der}(Q,K)$.*

[8] In Section C-2.6 we defined a derivation of a (not-necessarily-associative) ring $R$ as a function $d\colon R \to R$ with $d(xy) = d(x)y + xd(y)$. Derivations here are defined on modules, not on rings.

**Proof.** Let $\varphi$ be a stabilizing automorphism. If $\ell\colon Q \to G$ is a lifting, then Proposition C-3.24(ii) says that $\varphi(a+\ell x) = a+d(x)+\ell x$, where $d$ is a unique derivation; hence, the function $\mathrm{Stab}(Q,K) \to \mathrm{Der}(Q,K)$, given by $\varphi \mapsto d$, is well-defined, and it is easily seen to be a homomorphism.

To see that this map is an isomorphism, we construct its inverse. If $d \in \mathrm{Der}(Q,K)$, define $\varphi\colon G \to G$ by $\varphi(a+\ell x) = a + d(x) + \ell x$. Now $\varphi$ is stabilizing, by Proposition C-3.24, and $d \mapsto \varphi$ is the desired inverse function. •

It is not obvious from its definition that $\mathrm{Stab}(Q,K)$ is abelian, for its binary operation is composition. However, $\mathrm{Stab}(Q,K)$ is abelian, for $\mathrm{Der}(Q,K)$ is.

Recall that an automorphism $\varphi$ of a group $G$ is called an *inner automorphism* if it is a conjugation; that is, there is $c \in G$ with $\varphi(g) = c + g - c$ for all $g \in G$ (if $G$ is written additively).

**Lemma C-3.27.** *Let $0 \to K \to G \to Q \to 1$ be a split extension, and let $\ell\colon Q \to G$ be a lifting. Then a function $\varphi\colon G \to G$ is an inner stabilizing automorphism by some $a_0 \in K$ if and only if*

$$\varphi(a+\ell x) = a + xa_0 - a_0 + \ell x.$$

**Proof.** If we write $d(x) = xa_0 - a_0$, then $\varphi(a+\ell x) = a+d(x)+\ell x$. But $d$ is a (principal) derivation, and so $\varphi$ is a stabilizing automorphism, by Proposition C-3.24. Finally, $\varphi$ is conjugation by $-a_0$, for

$$-a_0 + (a+\ell x) + a_0 = -a_0 + a + xa_0 + \ell x = \varphi(a+\ell x).$$

Conversely, assume that $\varphi$ is a stabilizing conjugation. That $\varphi$ is stabilizing says that $\varphi(a+\ell x) = a + d(x) + \ell x$; that $\varphi$ is conjugation by $a_0 \in K$ says that $\varphi(a+\ell x) = a_0 + a + \ell x - a_0$. But $a_0 + a + \ell x - a_0 = a_0 + a - xa_0 + \ell x$, so that $d(x) = a_0 - xa_0$, as desired. •

**Definition.** If $Q$ is a group and $K$ is a $Q$-module, define

$$H^1(Q,K) = \mathrm{Der}(Q,K)/\mathrm{PDer}(Q,K),$$

where $\mathrm{PDer}(Q,K)$ is the subgroup of $\mathrm{Der}(Q,K)$ consisting of all the principal derivations.

**Proposition C-3.28.** *Let $0 \to K \to G \to Q \to 1$ be a split extension, and let $C$ and $C'$ be complements of $K$ in $G$. If $H^1(Q,K) = \{0\}$, then $C$ and $C'$ are conjugate.*

**Proof.** Since $G$ is a semidirect product, there are liftings $\ell\colon Q \to G$, with image $C$, and $\ell'\colon Q \to G$, with image $C'$, which are homomorphisms. Thus, the factor sets $f$ and $f'$ determined by each of these liftings is identically zero, and so $f' - f = 0$. But Lemma C-3.15 says that there exists $h\colon Q \to K$, namely, $h(x) = \ell' x - \ell x$, with

$$0 = f'(x,y) - f(x,y) = xh(y) - h(xy) + h(x);$$

thus, $h$ is a derivation. Since $H^1(Q,K) = \{0\}$, $h$ is a principal derivation: there is $a_0 \in K$ with

$$\ell' x - \ell x = h(x) = xa_0 - a_0$$

for all $x \in Q$. Since addition in $G$ satisfies $\ell' x - a_0 = -xa_0 + \ell' x$, we have

$$\ell x = a_0 - xa_0 + \ell' x = a_0 + \ell' x - a_0.$$

But $\operatorname{im} \ell = C$ and $\operatorname{im} \ell' = C'$, and so $C$ and $C'$ are conjugate via $a_0$. •

We can now supplement the Schur–Zassenhaus Theorem.

**Theorem C-3.29.** *Let $G$ be a finite group of order $mn$, where $\gcd(m, n) = 1$. If $K$ is an abelian normal subgroup of order $m$, then $G$ is a semidirect product of $K$ by $G/K$, and any two complements of $K$ are conjugate.*

**Proof.** By Proposition C-3.28, it suffices to prove that $H^1(Q, K) = \{0\}$, where $Q = G/K$. Note, first, that $|Q| = |G|/|K| = mn/m = n$.

Let $d\colon Q \to K$ be a derivation: for all $x, y \in Q$, we have

$$d(xy) = xd(y) + d(x).$$

Sum this equation over all $y \in Q$ to obtain

$$\Delta = x\Delta + nd(x),$$

where $\Delta = \sum_{y \in Q} d(y)$ (as $y$ varies over $Q$, so does $xy$). Since $\gcd(m, n) = 1$, there are integers $s$ and $t$ with $sn + tm = 1$. Hence,

$$d(x) = snd(x) + tmd(x) = snd(x),$$

because $d(x) \in K$ and so $md(x) = 0$. Therefore,

$$d(x) = s\Delta - xs\Delta.$$

Setting $a_0 = -s\Delta$, we see that $d$ is a principal derivation. •

Removing the assumption in Theorem C-3.29 that $K$ is abelian is much more difficult than removing this assumption in Theorem C-3.21. One first proves that complements are conjugate if either $K$ or $Q$ is a solvable group (see Robinson [**181**], p. 255). Since $|Q|$ and $|K|$ are relatively prime, at least one of them has odd order. The Feit–Thompson Theorem (which says that every group of odd order is solvable) now completes the proof.

Let us contemplate the formulas that have arisen:

$$\begin{aligned} \text{factor set}: &\quad 0 = xf(y, z) - f(xy, z) + f(x, yz) - f(x, y), \\ \text{coboundary}: &\quad f(x, y) = xh(y) - h(xy) + h(x), \\ \text{derivation}: &\quad 0 = xd(y) - d(xy) + d(x), \\ \text{principal derivation}: &\quad d(x) = xa_0 - a_0. \end{aligned}$$

All these formulas involve alternating sums; factor sets and derivations seem to be in kernels, and coboundaries and principal derivations seem to be in images. Let us make this more precise.

Denote the cartesian product of $n$ copies of $Q$ by $Q^n$; it is customary to denote an element of $Q^n$ by $[x_1 \mid \cdots \mid x_n]$, using bars, instead of by $(x_1, \ldots, x_n)$, using commas. When $n = 0$, we denote $Q^0$ by the one-point set $\{[\,]\}$ whose only member is denoted by $[\,]$. Factor sets and coboundaries are certain functions $Q^2 \to K$, and derivations are certain functions $Q^1 \to K$. Let $B_n$ be the free left $\mathbb{Z}Q$-module

with basis $Q^n$; in particular, $B_0$ is free with basis $[\,]$ (so that $B_0 \cong \mathbb{Z}Q$). By the definition of basis, every function $f\colon Q^n \to K$ gives a unique $Q$-homomorphism $\widetilde{f}\colon B_n \to K$ extending $f$, for $K$ is a $Q$-module; that is, if

$$\mathbf{Fun}(Q^n, K)$$

denotes the family of all functions $Q^n \to K$ in the category **Sets**, then $f \mapsto \widetilde{f}$ gives a bijection

$$\mathbf{Fun}(Q^n, K) \to \mathrm{Hom}_{\mathbb{Z}Q}(B_n, K).$$

The inverse of this function is restriction

$$\mathrm{res}\colon\ \mathrm{Hom}_{\mathbb{Z}Q}(B_n, K) \to \mathbf{Fun}(Q^n, K),$$

defined by $\mathrm{res}\colon g \mapsto g|Q^n$.

We now define maps suggested by these various formulas:

$$\begin{aligned}
&d_3\colon B_3 \to B_2\colon && d_3[\,x \mid y \mid z\,] = x[\,y \mid z\,] - [\,xy \mid z\,] + [\,x \mid yz\,] - [\,x \mid y\,],\\
&d_2\colon B_2 \to B_1\colon && d_2[\,x \mid y\,] = x[\,y\,] - [\,xy\,] + [\,x\,],\\
&d_1\colon B_1 \to B_0\colon && d_1[\,x\,] = x[\,\,] - [\,\,].
\end{aligned}$$

We have defined each of $d_3$, $d_2$, and $d_1$ on bases of free modules, and so each extends by linearity to a $Q$-map.

Exercise C-2.12 on page 145 leads us to consider $\mathbb{Z}$ as a $Q$-module, and this will help us organize these formulas.

**Proposition C-3.30.** *For any group $Q$, there is an isomorphism $\mathbb{Z}Q/\ker\varepsilon' \cong \mathbb{Z}$, where $\mathbb{Z}$ is regarded as a trivial $Q$-module and $\varepsilon'\colon \mathbb{Z}Q \to \mathbb{Z}$ is defined by*

$$\varepsilon'\colon \sum_{x\in Q} m_x x \mapsto \sum_{x\in Q} m_x.$$

**Proof.** Now $\varepsilon'$ is a $Q$-map, for if $x \in Q$, then $\varepsilon'(x) = 1$; on the other hand, $\varepsilon'(x) = \varepsilon'(x \cdot 1) = x\varepsilon'(1) = 1$, because $\mathbb{Z}$ is a trivial $Q$-module. •

**Proposition C-3.31.** *If $\mathbb{Z}$ is a trivial $Q$-module, then the sequence*

$$B_3 \xrightarrow{d_3} B_2 \xrightarrow{d_2} B_1 \xrightarrow{d_1} B_0 \xrightarrow{\varepsilon} \mathbb{Z} \longrightarrow 0 \tag{3}$$

*is an exact sequence of $Q$-modules, where $\varepsilon\colon B_0 \to \mathbb{Z}$ is defined by*

$$\varepsilon\colon \sum_{x\in Q} m_x x[\,] \mapsto \sum_{x\in Q} m_x.$$

.

**Sketch of Proof.** It is clear that $\varepsilon$ is surjective, for $\varepsilon = \varepsilon'\alpha$, where $\alpha\colon u[\,] \mapsto u$ for all $u \in \mathbb{Z}Q$ is an isomorphism $B_0 \to \mathbb{Z}Q$. We will check only that $\varepsilon d_1 = 0$, $d_1 d_2 = 0$, and $d_2 d_3 = 0$; that is, $\mathrm{im}\, d_1 \subseteq \ker\varepsilon$, $\mathrm{im}\, d_2 \subseteq \ker d_1$, and $\mathrm{im}\, d_3 \subseteq \ker d_2$; the (trickier) reverse inclusions will be proved in Theorem C-3.121 when we consider

the *bar resolution*. We have

$$\varepsilon d_1[\,x] = \varepsilon(x[\,] - [\,]) = 1 - 1 = 0,$$

$$\begin{aligned} d_1 d_2[\,x \mid y] &= d_1(x[\,y] - [\,xy] + [\,x]) \\ &= x(d_1[\,y] - d_1[\,xy] + d_1[\,x]) \\ &= x(y[\,] - [\,]) - (xy[\,] - [\,]) + (x[\,] - [\,]) \\ &= xy[\,] - x[\,] - xy[\,] + [\,] + x[\,] - [\,] \\ &= 0. \end{aligned}$$

(the equation $d_1 x[\,y] = x d_1[\,y]$ holds because $d_1$ is a $Q$-map). The reader should note that this is the same calculation as in Proposition C-3.16,

$$\begin{aligned} d_2 d_3[\,x \mid y \mid z] &= d_2(x[\,y \mid z] - [\,xy \mid z] + [\,x \mid yz] - [\,x \mid y]) \\ &= x d_2[\,y \mid z] - d_2[\,xy \mid z] + d_2[\,x \mid yz] - d_2[\,x \mid y] \\ &= x(\,y[z] - [\,yz] + [\,y]) - (xy[\,z] - [\,xyz] + [\,xy]) \\ &\qquad + (x[\,yz] - [\,xyz] + [\,x]) - (x[\,y] - [\,xy] + [\,x]) \\ &= 0. \quad \bullet \end{aligned}$$

The map $\varepsilon\colon B_0 \to \mathbb{Z}$, usually written as a map $\mathbb{Z}Q \to \mathbb{Z}$, is called the ***augmentation***. See Exercise C-2.12 on page 145.

If $X$ is a set and $K$ is a module, then functions $X \to K$ are the same as homomorphisms $B \to K$, where $B$ is the free module having basis $X$: formally, $\mathbf{Fun}(X, \quad)\colon {}_{\mathbb{Z}Q}\mathbf{Mod} \to \mathbf{Sets}$ is a functor which is naturally equivalent to $\mathrm{Hom}_{\mathbb{Z}Q}(B, \quad)$ (see Exercise C-3.18 on page 255 which says that there is an adjoint pair of functors lurking). Applying the contravariant functor $\mathrm{Hom}_{\mathbb{Z}Q}(\ , K)$ to the sequence in Proposition C-3.31, we obtain a (not necessarily exact) sequence

$$\mathrm{Hom}(B_3, K) \xleftarrow{d_3^*} \mathrm{Hom}(B_2, K) \xleftarrow{d_2^*} \mathrm{Hom}(B_1, K) \xleftarrow{d_1^*} \mathrm{Hom}(B_0, K);$$

inserting the bijections $\mathrm{res}\colon g \mapsto g|Q^n$ gives a commutative diagram of sets:

$$\begin{array}{ccccccc} \mathbf{Fun}(Q^3, K) & \longleftarrow & \mathbf{Fun}(Q^2, K) & \longleftarrow & \mathbf{Fun}(Q, K) & \longleftarrow & \mathbf{Fun}(\{1\}, K) \\ \uparrow{\scriptstyle \mathrm{res}} & & \uparrow{\scriptstyle \mathrm{res}} & & \uparrow{\scriptstyle \mathrm{res}} & & \uparrow{\scriptstyle \mathrm{res}} \\ \mathrm{Hom}(B_3, K) & \xleftarrow{d_3^*} & \mathrm{Hom}(B_2, K) & \xleftarrow{d_2^*} & \mathrm{Hom}(B_1, K) & \xleftarrow{d_1^*} & \mathrm{Hom}(B_0, K). \end{array}$$

We regard a function $f\colon Q^n \to K$ as the restriction of the $Q$-map $\widetilde{f}\colon B_n \to K$ which extends it. Suppose that $f\colon Q^2 \to K$ lies in $\ker d_3^*$. Then $0 = d_3^*(f) = f d_3$. Hence, for all $x, y, z \in Q$, we have

$$\begin{aligned} 0 &= f d_3[\,x \mid y \mid z] \\ &= f(x[\,y \mid z] - [\,xy \mid z] + [\,x \mid yz] - [\,x \mid y]) \\ &= x f[\,y \mid z] - f[\,xy \mid z] + f[\,x \mid yz] - f[\,x \mid y]; \end{aligned}$$

the equation $f(x[\,y \mid z]) = x f[\,y \mid z]$ holds because $f$ is the restriction of a $Q$-map. Thus, $f$ is a factor set.

If $f$ lies in $\operatorname{im} d_2^*$, then there is some $h\colon Q \to K$ with $f = d_2^*(h) = hd_2$. Thus,

$$\begin{aligned} f[\,x \mid y] &= hd_2[\,x \mid y] \\ &= h(x[y] - [xy] + [x]) \\ &= xh[y] - h[xy] + h[x]; \end{aligned}$$

the equation $h(x[y]) = xh[y]$ holds because $h$ is the restriction of a $Q$-map. Thus, $f$ is a coboundary.

If $g\colon Q \to K$, then

$$d_2^*(g)[\,x \mid y] = gd_2[\,x \mid y] = g(x[\,y] - [\,y] + [\,x]).$$

Hence, if $g$ is in $\ker d_2^*$, then $0 = d_2^*(g)[\,x \mid y]$, so that $g(xy) = xg(y) + g(x)$ and $g$ is a derivation.

We now compute $\operatorname{im} d_1^*$. Let $k\colon \{[\,]\} \to K$ and $k([\,]) = a_0$ (this is just a fancy way of specifying an element of $K$). For all $x \in Q$,

$$d_1^*(k)[\,x] = kd_1[\,x] = k(x[\,] - [\,]) = xk([\,]) - [\,] = xa_0 - a_0.$$

Thus, $d_1^*(k)$ is a principal derivation.

Observe that $d_2d_3 = 0$ implies $d_3^*d_2^* = 0$, which is equivalent to $\operatorname{im} d_2^* \subseteq \ker d_3^*$; that is, every coboundary is a factor set, which is Proposition C-3.16. Similarly, $d_1d_2 = 0$ implies $\operatorname{im} d_1^* \subseteq \ker d_2^*$; that is, every principal derivation is a derivation, which is Example C-3.25(ii).

As long as we are computing kernels and images, what is $\ker d_1^*$? When we computed $\operatorname{im} d_1^*$, we saw, for all $x \in Q$, that $kd_1(x) = xa_0 - a_0$, where $k([\,]) = a_0$. Hence, if $kd_1(x) = 0$, then $xa_0 = a_0$ for all $x \in Q$. We have been led to the following definition.

**Definition.** If $Q$ is a group and $K$ is a $Q$-module, define

$$H^0(Q, K) = \ker d_1^*.$$

The submodule $K^Q$ of ***fixed points*** is defined by

$$K^Q = \{a \in K : xa = a \text{ for all } x \in Q\}.$$

The reader may show that $H^0(Q, K) = \ker d_1^* \cong K^Q$. It is easy to see that $K^Q$ is a trivial $Q$-module; indeed, it is the maximal trivial $Q$-submodule of $K$.

We have seen that the ***cohomology groups*** $H^2(Q, K)$, $H^1(Q, K)$, $H^0(Q, K)$ are obtained after applying the contravariant functor $\operatorname{Hom}_{\mathbb{Z}Q}(\ \ , K)$ to exact sequence (3),

$$\begin{aligned} H^2(Q, K) &= \ker d_3^* / \operatorname{im} d_2^*, \\ H^1(Q, K) &= \ker d_2^* / \operatorname{im} d_1^*, \\ H^0(Q, K) &= \ker d_1^*. \end{aligned}$$

The functor $- \otimes_{\mathbb{Z}Q} K$ can also be applied to exact sequence (3); the tensor product is defined because we may view the free $Q$-modules $B_n$ as right $Q$-modules,

as in Example B-1.20(v) in Part 1. We obtain ***homology groups***:

$$H_2(Q,K) = \ker(d_2 \otimes 1)/\operatorname{im}(d_3 \otimes 1),$$
$$H_1(Q,K) = \ker(d_1 \otimes 1)/\operatorname{im}(d_2 \otimes 1),$$
$$H_0(Q,K) = \ker(\varepsilon \otimes 1).$$

The notation for cohomology groups uses superscripts, while the notation for homology uses subscripts.

We will see that $H_0(Q,K)$ is the maximal trivial $Q$-quotient of $K$ and, in the special case $K = \mathbb{Z}$ viewed as a trivial $Q$-module, that $H_1(Q,\mathbb{Z}) \cong Q/Q'$, where $Q'$ is the commutator subgroup of $Q$.

There are other applications of cohomology in group theory besides the Schur–Zassenhaus Lemma. For example, if $G$ is a group, $a \in G$, and $\gamma_a\colon g \mapsto aga^{-1}$ is conjugation by $a$, then $(\gamma_a)^n\colon g \mapsto a^n g a^{-n}$ for all $n$. Hence, if $a$ has prime order $p$ and $a \notin Z(G)$, then $\gamma_a$ is an inner automorphism of order $p$. A theorem of Gaschütz (see Rotman [**187**], p. 589) uses cohomology to prove that every finite nonabelian $p$-group $G$ has an outer automorphism $\alpha$ of order $p$.

In the next sections, we discuss homological algebra; *cohomology of groups* is the proper context in which to understand the constructions in this section.

## Exercises

∗ **C-3.13.** Let $Q$ be a group and let $K$ be a $Q$-module. Prove that any two split extensions of $K$ by $Q$ realizing the operators are equivalent.

**C-3.14.** Let $Q$ be a group and let $K$ be a $Q$-module.

(i) If $K$ and $Q$ are finite groups, prove that $H^2(Q,K)$ is also finite.

(ii) Let $\tau(K,Q)$ denote the number of nonisomorphic middle groups $G$ that occur in extensions of $K$ by $Q$ realizing the operators. Prove that

$$\tau(K,Q) \le |H^2(Q,K)|.$$

(iii) Give an example showing that the inequality in (ii) can be strict.
**Hint.** Observe that $\tau(\mathbb{Z}_p,\mathbb{Z}_p) = 2$ (note that the kernel is the trivial module because every group of order $p^2$ is abelian).

**C-3.15.** Recall that the ***generalized quaternion group*** $\mathbf{Q}_n$ is a group of order $2^n$, where $n \ge 3$, which is generated by two elements $a$ and $b$ such that

$$a^{2^{n-1}} = 1, \quad bab^{-1} = a^{-1}, \quad \text{and} \quad b^2 = a^{2^{n-2}}.$$

(i) Prove that $\mathbf{Q}_n$ has a unique element $z$ of order 2 and that $Z(\mathbf{Q}_n) = \langle z\rangle$. Conclude that $\mathbf{Q}_n$ is not a semidirect product.

(ii) Prove that $\mathbf{Q}_n$ is a ***central extension*** (i.e., $\theta$ is trivial) of $\mathbb{Z}_2$ by $D_{2^{n-1}}$.

(iii) Using factor sets, give another proof of the existence of $\mathbf{Q}_n$.

∗ **C-3.16.** If $p$ is an odd prime, prove that every group $G$ of order $2p$ is a semidirect product of $\mathbb{Z}_p$ by $\mathbb{Z}_2$, and conclude that either $G$ is cyclic or $G \cong D_{2p}$.

∗ **C-3.17.** Show that every group $G$ of order 12 is isomorphic to one of the following five groups:

$$\mathbb{Z}_{12}, \quad \mathbf{V} \times \mathbb{Z}_3, \quad A_4, \quad S_3 \times \mathbb{Z}_2, \quad T,$$

where $T$ is the group in Example C-3.9.

∗ **C-3.18.** Let $U\colon {}_{\mathbb{Z}Q}\mathbf{Mod} \to \mathbf{Sets}$ be the forgetful functor which assigns to each module its set of elements. By Exercise B-4.23 on page 474 in Part 1, there exists a ***free functor*** $\Phi\colon \mathbf{Fun} \to {}_{\mathbb{Z}Q}\mathbf{Mod}$ which assigns to each set $X$ the free $Q$-module $\Phi(X)$ with basis $X$.

(i) Prove that $\mathbf{Fun}(X, \ )$, $\mathrm{Hom}(\Phi, \ )\colon {}_{\mathbb{Z}Q}\mathbf{Mod} \to \mathbf{Sets}$ are naturally equivalent.

(ii) Prove that the ordered pair $(\Phi, U)$ is an adjoint pair of functors. (Compare this exercise to Exercise C-1.96 on page 91.)

∗ **C-3.19.** Let $Q$ be a group.

(i) Prove that $\mathrm{Fix}^Q\colon {}_{\mathbb{Z}Q}\mathbf{Mod} \to {}_{\mathbb{Z}Q}\mathbf{Mod}$ is a covariant functor, where $\mathrm{Fix}^Q\colon K \mapsto K^Q$ and, if $\varphi\colon K \to L$, then $\varphi^Q\colon K^Q \to L^Q$ is the restriction $\varphi|K^Q$.

(ii) Prove that the functors $\mathrm{Fix}^Q$ and $\mathrm{Hom}_{\mathbb{Z}Q}(\mathbb{Z}, \ \ )$ are naturally equivalent, where $\mathbb{Z}$ is viewed as a trivial $Q$-module. Conclude that $\mathrm{Fix}^Q$ is left exact.

## C-3.4. Complexes

In this section, the word *module* will always mean "left $R$-module", where $R$ is a ring.

By Proposition B-2.25 in Part 1, for every module $M$, there is a free module $P_0$ and a surjection $\varepsilon\colon P_0 \to M$; that is, there is an exact sequence

$$0 \to \Omega_1 \xrightarrow{i} P_0 \xrightarrow{\varepsilon} M \to 0,$$

where $\Omega_1 = \ker \varepsilon$ and $i\colon \Omega_1 \to P_0$ is the inclusion. This is just another way of describing $M$ by generators and relations. If $X$ is a basis of $P_0$, then we say that $X$ (really, $\varepsilon(X)$) is a set of generators of $M$, that $\Omega_1$ are relations (usually the submodule $\Omega_1$ is replaced by a generating subset of it), and that $(X \mid \Omega_1)$ is a presentation of $M \cong P_0/\ker \varepsilon = P_0/\operatorname{im} i$.

The idea now is to take generators and relations of $\Omega_1$, getting "second-order" relations $\Omega_2$, and then to keep iterating this construction giving a *free resolution* of $M$, an infinitely long exact sequence of free modules, which should be regarded as a glorified description of $M$ by generators and relations. Thus, it makes sense to replace $M$ by such a long exact sequence, and this is the fundamental idea underlying homological algebra. (A similar construction occurs in algebraic topology, where a topological space $X$ is replaced by a sequence of *chain groups* which yields its homology groups $H_n(X)$.)

**Definition.** A ***projective resolution*** of a module $M$ is an exact sequence,

$$\cdots \to P_n \to P_{n-1} \to \cdots \to P_1 \to P_0 \to M \to 0,$$

in which each module $P_n$ is projective. A ***free resolution*** is a projective resolution in which each module $P_n$ is free.

Recall Proposition C-3.31, which states that Eq. (3),

$$B_3 \xrightarrow{d_3} B_2 \xrightarrow{d_2} B_1 \xrightarrow{d_1} B_0 \xrightarrow{\varepsilon} \mathbb{Z} \to 0,$$

is an exact sequence of free left $\mathbb{Z}Q$-modules, where $B_n$ is the free $Q$-module with basis $Q^n$. We have only proved part of this proposition (a complete proof will be given when we treat the *bar resolution*), but we say now that this sequence can be lengthened to be a free resolution of the trivial $Q$-module $\mathbb{Z}$.

**Proposition C-3.32.** *Every module $M$ has a free resolution* (*and hence it has a projective resolution*).

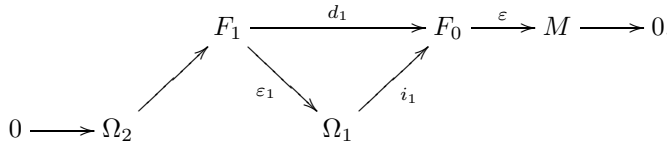
**Proof.** By Proposition B-2.25 in Part 1, there is a free module $F_0$ and an exact sequence

$$0 \to \Omega_1 \xrightarrow{i_1} F_0 \xrightarrow{\varepsilon} M \to 0.$$

Similarly, there is a free module $F_1$, a surjection $\varepsilon_1 \colon F_1 \to \Omega_1$, and an exact sequence

$$0 \to \Omega_2 \xrightarrow{i_2} F_1 \xrightarrow{\varepsilon_1} \Omega_1 \to 0.$$

If we define $d_1 \colon F_1 \to F_0$ to be the composite $i_1\varepsilon_1$, then $\operatorname{im} d_1 = \Omega_1 = \ker \varepsilon$ and $\ker d_1 = \Omega_2$. Therefore, there is an exact sequence




Thus, we have spliced the two short exact sequences to form a longer exact sequence. Plainly, this construction can be iterated for all $n \geq 0$ (so that the ultimate exact sequence is infinitely long). •

There is a dual construction, giving "co-generators" and "co-relations".

**Definition.** An ***injective resolution*** of a module $M$ is an exact sequence,

$$0 \to M \to E^0 \to E^1 \to \cdots \to E^n \to E^{n+1} \to \cdots,$$

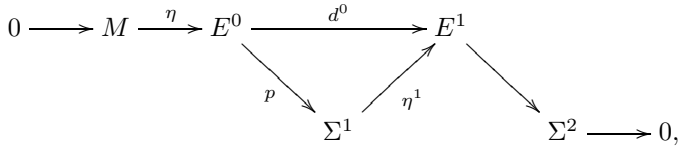
in which each module $E^n$ is injective.

**Proposition C-3.33.** *Every module $M$ has an injective resolution.*

**Proof.** We use Theorem B-4.64 in Part 1, which states that every module can be imbedded as a submodule of an injective module. Thus, there is an injective module $E^0$, an injection $\eta \colon M \to E^0$, and an exact sequence

$$0 \to M \xrightarrow{\eta} E^0 \xrightarrow{p} \Sigma^1 \to 0,$$

where $\Sigma^1 = \operatorname{coker}\eta$ and $p$ is the natural map. Now repeat: there is an injective module $E^1$, an imbedding $\eta^1\colon \Sigma^1 \to E^1$, yielding an exact sequence




where $d^0$ is the composite $d^0 = \eta^1 p$. This splicing can be iterated for all $n \geq 0$. •

We now generalize both of these definitions.

**Definition.** A ***complex***[9] $(\mathbf{C}_\bullet, d_\bullet)$ is a sequence of modules and maps, for every $n \in \mathbb{Z}$,

$$\mathbf{C}_\bullet = \cdots \to C_{n+1} \xrightarrow{d_{n+1}} C_n \xrightarrow{d_n} C_{n-1} \to \cdots,$$

in which $d_n d_{n+1} = 0$ for all $n$. The maps $d_n$ are called ***differentiations***.

Usually, we will shorten the notation $(\mathbf{C}_\bullet, d_\bullet)$ to $\mathbf{C}_\bullet$.

Note that the equation $d_n d_{n+1} = 0$ is equivalent to

$$\operatorname{im} d_{n+1} \subseteq \ker d_n.$$

**Example C-3.34.**

(i) Every exact sequence is a complex, for the required inclusions, $\operatorname{im} d_{n+1} \subseteq \ker d_n$, are now equalities, $\operatorname{im} d_{n+1} = \ker d_n$.

(ii) The sequence of chain groups of a triangulated space $X$,

$$\cdots \to C_3(X) \xrightarrow{\partial_3} C_2(X) \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X),$$

is a complex.

(iii) In Exercise C-2.23 on page 168, we considered the de Rham complex of a connected open subset $X$ of $\mathbb{R}^n$:

$$0 \to \Delta^0(X) \xrightarrow{d^0} \Delta^1(X) \xrightarrow{d^1} \Delta^2(X) \to \cdots \to \Delta^{n-1}(X) \xrightarrow{d^{n-1}} \Delta^n(X) \to 0,$$

where the maps are the exterior derivatives.

(iv) The ***zero complex*** $\mathbf{0}_\bullet$ is the complex $(\mathbf{C}_\bullet, d_\bullet)$ each of whose terms $C_n = \{0\}$ and, necessarily, each of whose differentiations $d_n = 0$.

(v) Every homomorphism $f\colon A \to B$ is a differentiation of some complex. Define a complex $(\mathbf{C}_\bullet, d_\bullet)$ with $C_1 = A$, $C_0 = B$, $d_1 = f$, and all other terms and differentiations zero. In particular, if $A = \{0\} = B$, then $f\colon A \to B$ is the zero map and $(\mathbf{C}_\bullet, d_\bullet)$ is the zero complex.

(vi) If $\{M_n : n \in \mathbb{Z}\}$ is any sequence of modules, then $(\mathbf{M}_\bullet, d_\bullet)$ is a complex with $n$th term $M_n$ if we define $d_n = 0$ for all $n$.

[9] These are also called ***chain complexes*** in the literature.

(vii) A complex must have a module for every $n \in \mathbb{Z}$. We force projective resolutions to be complexes by defining $C_n = \{0\}$ for all negative $n$; there is no problem defining differentiations $d_n \colon C_n \to C_{n-1}$ for $n \leq 0$, for there is only the zero map from any module into $\{0\}$. Thus, every projective resolution of a module $M$,

$$\cdots \to P_1 \to P_0 \to M \to 0,$$

is a complex if we add $\{0\}$'s to the right.

(viii) Every injective resolution of a module $M$,

$$0 \to M \to E^0 \to E^1 \to \cdots,$$

is a complex if we add $\{0\}$'s to the left. We are using a convenient notation, letting indices be superscripts instead of subscripts. According to the definition of complex, however, differentiations must lower indices: $d_n \colon C_n \to C_{n-1}$ for all $n \in \mathbb{Z}$. The simplest way to satisfy the definition here is to use negative indices: define $C_{-n} = E^n$, so that

$$0 \to M \to C_0 \to C_{-1} \to C_{-2} \to \cdots$$

is a complex.

(ix) If $\mathbf{C}_\bullet$ is a complex,

$$\mathbf{C}_\bullet = \cdots \to C_n \xrightarrow{d_n} C_{n-1} \to \cdots,$$

and $F$ is a covariant additive functor, say, $F \colon {}_R\mathbf{Mod} \to \mathbf{Ab}$, then $F(\mathbf{C}_\bullet)$, defined by

$$F(\mathbf{C}_\bullet) = \cdots \to F(C_n) \xrightarrow{Fd_n} F(C_{n-1}) \to \cdots,$$

is also a complex. Since $F$ is an additive functor, the equation $0 = F(0)$ holds, so that

$$0 = F(0) = F(d_n d_{n+1}) = F(d_n)F(d_{n+1}).$$

Note that even if the original complex is exact, the functored complex $F(\mathbf{C}_\bullet)$ may not be exact.

(x) If $F$ is a contravariant additive functor, it is also true that $F(\mathbf{C}_\bullet)$ is a complex, but we have to arrange notation so that differentiations lower indices by 1. In more detail, after applying $F$, we have

$$F(\mathbf{C}_\bullet) = \cdots \leftarrow F(C_n) \xleftarrow{Fd_n} F(C_{n-1}) \leftarrow \cdots;$$

the differentiations $Fd_n$ increase indices by 1. Introducing negative indices almost solves the problem. If we define $X_{-n} = F(C_n)$, then the sequence is rewritten as

$$F(\mathbf{C}_\bullet) = \cdots \to X_{-n+1} \xrightarrow{Fd_n} X_{-n} \to \cdots.$$

However, the index on the map should be $-n+1$, and not $n$. Define

$$\delta_{-n+1} = Fd_n.$$

The relabeled sequence now reads properly:

$$F(\mathbf{C}_\bullet) = \cdots \to X_{-n+1} \xrightarrow{\delta_{-n+1}} X_{-n} \to \cdots.$$

Negative indices are awkward, however, and the following notation is customary: change the sign of the index by raising it to a superscript: write

$$\delta^n = \delta_{-n} = Fd_{n+1}.$$

The final version of the functored sequence now looks like this:

$$F(\mathbf{C}_\bullet) = \cdots \to X^{n-1} \stackrel{\delta^{n-1}}{\longrightarrow} X^n \to \cdots . \quad \blacktriangleleft$$

We now define morphisms of complexes.

**Definition.** If $(\mathbf{C}_\bullet, d_\bullet)$ and $(\mathbf{C}'_\bullet, d'_\bullet)$ are complexes, then a ***chain map***

$$f = f_\bullet \colon (\mathbf{C}_\bullet, d_\bullet) \to (\mathbf{C}'_\bullet, d'_\bullet)$$

is a sequence of maps $f_n \colon C_n \to C'_n$, for all $n \in \mathbb{Z}$, making the following diagram commute:

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & C_{n+1} & \xrightarrow{d_{n+1}} & C_n & \xrightarrow{d_n} & C_{n-1} & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle f_{n+1}} & & \downarrow{\scriptstyle f_n} & & \downarrow{\scriptstyle f_{n-1}} & & \\
\cdots & \longrightarrow & C'_{n+1} & \xrightarrow{d'_{n+1}} & C'_n & \xrightarrow{d'_n} & C'_{n-1} & \longrightarrow & \cdots .
\end{array}$$

It is easy to check that the composite $gf$ of two chain maps

$$f_\bullet \colon (\mathbf{C}_\bullet, d_\bullet) \to (\mathbf{C}'_\bullet, d'_\bullet) \quad \text{and} \quad g_\bullet \colon (\mathbf{C}'_\bullet, d'_\bullet) \to (\mathbf{C}''_\bullet, d''_\bullet)$$

is itself a chain map, where $(gf)_n = g_n f_n$. The ***identity chain map*** $1_{\mathbf{C}_\bullet}$ on $(\mathbf{C}_\bullet, d_\bullet)$ is the sequence of identity maps $1_{C_n} \colon C_n \to C_n$.

**Definition.** If $R$ is a ring, then all complexes of left $R$-modules and chain maps form a category, denoted by ${}_R\mathbf{Comp}$; if the ring $R$ is understood from the context, then we will omit the prescript $R$.

Just as we examined vector spaces before introducing linear transformations, we now examine complexes before defining homology (and cohomology) which are functors whose domain is ${}_R\mathbf{Comp}$.

Many of the constructions in ${}_R\mathbf{Mod}$ can also be done in the category ${}_R\mathbf{Comp}$. We merely list the definitions and state certain properties whose verifications are straightforward exercises for the reader.

(i) The category ${}_R\mathbf{Comp}$ is a pre-additive category (that is, the Hom's are abelian groups and the distributive laws hold whenever possible) if we define

$$(f + g)_n = f_n + g_n \text{ for all } n \in \mathbb{Z}.$$

(ii) An ***isomorphism*** in ${}_R\mathbf{Comp}$ is an equivalence in this category. The reader should check that a chain map $f \colon \mathbf{C}_\bullet \to \mathbf{C}'_\bullet$ is an isomorphism if and only if $f_n \colon C_n \to C'_n$ is an isomorphism in ${}_R\mathbf{Mod}$ for all $n \in \mathbb{Z}$. (It is necessary to check that the sequence of inverses $f_n^{-1}$ is, in fact, a chain map; that is, the appropriate diagram commutes.)

(iii) A complex $(\mathbf{A}_\bullet, \delta_\bullet)$ is a ***subcomplex*** of a complex $(\mathbf{C}_\bullet, d_\bullet)$ if, for every $n \in \mathbb{Z}$, we have $A_n$ a submodule of $C_n$ and $\delta_n = d_n|A_n$.

If $i_n\colon A_n \to C_n$ is the inclusion, then it is easy to see that $\mathbf{A}_\bullet$ is a subcomplex of $\mathbf{C}_\bullet$ if and only if $i\colon \mathbf{A}_\bullet \to \mathbf{C}_\bullet$ is a chain map.

(iv) If $\mathbf{A}_\bullet$ is a subcomplex of $\mathbf{C}_\bullet$, then the ***quotient complex*** is

$$\mathbf{C}_\bullet/\mathbf{A}_\bullet = \cdots \to C_n/A_n \xrightarrow{d''_n} C_{n-1}/A_{n-1} \to \cdots,$$

where $d''_n\colon c_n + A_n \mapsto d_n c_n + A_{n-1}$ (it must be shown that $d''_n$ is well-defined: if $c_n + A_n = b_n + A_n$, then $d_n c_n + A_{n-1} = d_n b_n + A_{n-1}$). If $\pi_n\colon C_n \to C_n/A_n$ is the natural map, then $\pi\colon \mathbf{C}_\bullet \to \mathbf{C}_\bullet/\mathbf{A}_\bullet$ is a chain map.

(v) If $f_\bullet\colon (\mathbf{C}_\bullet, d_\bullet) \to (\mathbf{C}'_\bullet, d'_\bullet)$ is a chain map, define

$$\mathbf{ker}\, f = \cdots \to \ker f_{n+1} \xrightarrow{\delta_{n+1}} \ker f_n \xrightarrow{\delta_n} \ker f_{n-1} \to \cdots,$$

where $\delta_n = d_n|\ker f_n$, and define

$$\mathbf{im}\, f = \cdots \to \operatorname{im} f_{n+1} \xrightarrow{\Delta_{n+1}} \operatorname{im} f_n \xrightarrow{\Delta_n} \operatorname{im} f_{n-1} \to \cdots,$$

where $\Delta_n = d'_n|\operatorname{im} f_n$. It is easy to see that $\mathbf{ker}\, f$ is a subcomplex of $\mathbf{C}_\bullet$, that $\mathbf{im}\, f$ is a subcomplex of $\mathbf{C}'_\bullet$, and that the ***First Isomorphism Theorem*** holds:

$$\mathbf{C}_\bullet/\mathbf{ker}\, f \cong \mathbf{im}\, f.$$

(vi) A sequence of complexes and chain maps

$$\cdots \to {\mathbf{C}_\bullet}^{n+1} \xrightarrow{f^{n+1}} {\mathbf{C}_\bullet}^{n} \xrightarrow{f^n} {\mathbf{C}_\bullet}^{n-1} \to \cdots$$

is an ***exact sequence*** in ${}_R\mathbf{Comp}$ if, for all $n \in \mathbb{Z}$,

$$\mathbf{im}\, f^{n+1} = \mathbf{ker}\, f^n.$$

We may check that if $\mathbf{A}_\bullet$ is a subcomplex of $\mathbf{C}_\bullet$, then there is an exact sequence of complexes

$$\mathbf{0}_\bullet \to \mathbf{A}_\bullet \xrightarrow{i} \mathbf{C}_\bullet,$$

where $\mathbf{0}_\bullet$ is the zero complex and $i$ is the chain map of inclusions. More generally, if $i\colon \mathbf{C}_\bullet \to \mathbf{C}'_\bullet$ is a chain map, then each $i_n$ is injective if and only if there is an exact sequence $\mathbf{0}_\bullet \to \mathbf{C}_\bullet \xrightarrow{i} \mathbf{C}'_\bullet$. Similarly, if $p\colon \mathbf{C}_\bullet \to \mathbf{C}''_\bullet$ is a chain map, then each $p_n$ is surjective if and only if there is an exact sequence

$$\mathbf{C}_\bullet \xrightarrow{p} \mathbf{C}''_\bullet \to \mathbf{0}_\bullet.$$

Thus, a sequence of complexes $\cdots \to {\mathbf{C}_\bullet}^{n+1} \xrightarrow{f^{n+1}} {\mathbf{C}_\bullet}^{n} \xrightarrow{f^n} {\mathbf{C}_\bullet}^{n-1} \to \cdots$ is exact in ${}_R\mathbf{Comp}$ if and only if, for every $m \in \mathbb{Z}$,

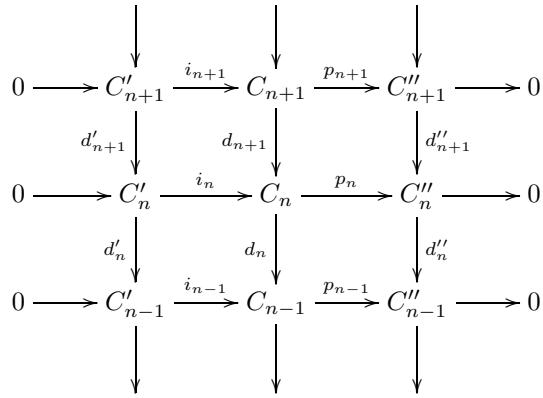
$$\cdots \to C_m^{n+1} \to C_m^n \to C_m^{n-1} \to \cdots$$

is an exact sequence of modules in ${}_R\mathbf{Mod}$.

(vii) The reader should realize that the notation for a short exact sequence of complexes,

$$\mathbf{0}_\bullet \longrightarrow \mathbf{C}'_\bullet \xrightarrow{i} \mathbf{C}_\bullet \xrightarrow{p} \mathbf{C}''_\bullet \longrightarrow \mathbf{0}_\bullet,$$

is very compact. For example, if we write a complex as a column, then a short exact sequence of complexes is really the infinite commutative diagram with exact rows:



More generally, a long exact sequence of complexes is a commutative diagram having a module at every lattice point in the plane.

(viii) If $(\mathbf{C}_\bullet^\alpha, d_\bullet^\alpha)_{\alpha\in I}$ is a family of complexes, then their ***direct sum*** is the complex

$$\sum_\alpha \mathbf{C}_\bullet^\alpha = \cdots \to \sum_\alpha C_{n+1}^\alpha \xrightarrow{\sum_\alpha d_n^\alpha} \sum_\alpha C_n^\alpha \xrightarrow{\sum_\alpha d_{n-1}^\alpha} \sum_\alpha C_{n-1}^\alpha \to \cdots,$$

where $\sum_\alpha d_n^\alpha$ acts coordinatewise; that is, $\sum_\alpha d_n^\alpha \colon (c_n^\alpha) \mapsto (d_n^\alpha c_n^\alpha)$.

To summarize, we can view ${}_R\mathbf{Comp}$ as a category having virtually the same properties as the category ${}_R\mathbf{Mod}$ of modules; indeed, we should view a complex as a generalized module. (Categories such as ${}_R\mathbf{Mod}$ and ${}_R\mathbf{Comp}$ are called ***abelian categories***; we shall discuss them in Chapter C-4.)

## Exercises

**C-3.20.** Regard the map $d\colon \mathbb{Z} \to \mathbb{Z}$, defined by $d\colon m \mapsto 2m$, as a complex, as in Example C-3.34(v). Prove that it is not a projective object in the category ${}_\mathbb{Z}\mathbf{Comp}$ even though each of its terms is a projective $\mathbb{Z}$-module.

∗ **C-3.21.** View $\mathbb{Z}$ as the category $\mathbf{PO}(\mathbb{Z})$ whose objects are integers $n$ and whose morphisms are $n \to m$ whenever $n \le m$, but with no other morphisms. (If we view $\mathbb{Z}$ as a partially ordered set, then this is the associated category defined in Example B-4.1(viii) in Part 1.) Prove that a complex $(\mathbf{C}_\bullet, d_\bullet)$ is a contravariant functor $\mathbf{PO}(\mathbb{Z}) \to {}_R\mathbf{Mod}$ and that a chain map is a natural transformation.

∗ **C-3.22.** (i) Let $0 \to F_n \to F_{n-1} \to \cdots \to F_0 \to 0$ be an exact sequence of finitely generated free $k$-modules, where $k$ is a commutative ring. Prove that

$$\sum_{i=0}^{n} (-1)^i \operatorname{rank}(F_i) = 0.$$

(ii) Let

$$0 \to F_n \to F_{n-1} \to \cdots \to F_0 \to M \to 0$$

and

$$0 \to F'_m \to F'_{m-1} \to \cdots \to F'_0 \to M \to 0$$

be free resolutions of a $k$-module $M$ in which all $F_i$ and $F'_j$ are finitely generated free $k$-modules. Prove that

$$\sum_{i=0}^{n}(-1)^i \operatorname{rank}(F_i) = \sum_{j=0}^{m}(-1)^j \operatorname{rank}(F'_j).$$

Their common value is denoted by $\chi(M)$, and it is called the ***Euler–Poincaré characteristic*** of $M$ (see Exercise C-1.116(iii) on page 110).
**Hint.** Use Schanuel's Lemma.

## C-3.5. Homology Functors

The definition of homology is very simple, but we have delayed presenting it until now because homology cannot be appreciated without seeing how it arose (the extension problem, in the 1920s) as well as an example of what it is good for (the Schur–Zassenhaus Theorem). Its first most prominent appearance occurred in the 1930s, in the construction of homology groups of triangulated spaces that we described in Section C-3.1.

**Definition.** If $\mathbf{C}_\bullet = \cdots \to C_{n+1} \xrightarrow{d_{n+1}} C_n \xrightarrow{d_n} C_{n-1} \to \cdots$ is a complex, define its

$$\textbf{\textit{n-cycles}} = Z_n(\mathbf{C}_\bullet) = \ker d_n,$$
$$\textbf{\textit{n-boundaries}} = B_n(\mathbf{C}_\bullet) = \operatorname{im} d_{n+1}.$$

Since the equation $d_n d_{n+1} = 0$ in a complex is equivalent to the condition

$$\operatorname{im} d_{n+1} \subseteq \ker d_n,$$

we have $B_n(\mathbf{C}_\bullet) \subseteq Z_n(\mathbf{C}_\bullet)$ for every complex $\mathbf{C}_\bullet$.

**Definition.** If $\mathbf{C}_\bullet$ is a complex and $n \in \mathbb{Z}$, its $n$th ***homology*** is

$$H_n(\mathbf{C}_\bullet) = Z_n(\mathbf{C}_\bullet)/B_n(\mathbf{C}_\bullet).$$

**Example C-3.35.** A complex is an exact sequence if and only if all its homology groups are $\{0\}$: that is, $H_n(\mathbf{C}_\bullet) = \{0\}$ for all $n$. Thus, homology measures the deviation of a complex from being an exact sequence. For this reason, an exact sequence is often called an ***acyclic complex***; *acyclic* means "no cycles", that is, no cycles that are not boundaries. ◀

**Example C-3.36.** In Example C-3.34(v), we saw that every homomorphism $f\colon A \to B$ can be viewed as a differentiation in a complex $\mathbf{C}_\bullet$, where $C_1 = A$, $C_0 = B$, $d_1 = f$, and $C_n = \{0\}$ to the left ($n > 1$) and to the right ($n < 0$). Now $d_2\colon \{0\} \to A$

being the zero map implies $\operatorname{im} d_2 = 0$, while $d_0 \colon B \to \{0\}$ being the zero map implies $\ker d_0 = B$; it follows that

$$H_n(\mathbf{C}_\bullet) = \begin{cases} \ker f & \text{if } n = 1, \\ \operatorname{coker} f & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases} \quad \blacktriangleleft$$

**Proposition C-3.37.** *Homology $H_n \colon {}_R\mathbf{Comp} \to {}_R\mathbf{Mod}$ is an additive functor for each $n \in \mathbb{Z}$.*

**Proof.** We have just defined $H_n$ on objects; it remains to define $H_n$ on morphisms. If $f \colon (\mathbf{C}_\bullet, d_\bullet) \to (\mathbf{C}'_\bullet, d'_\bullet)$ is a chain map, define $H_n(f) \colon H_n(\mathbf{C}_\bullet) \to H_n(\mathbf{C}'_\bullet)$ by

$$H_n(f) \colon z_n + B_n(\mathbf{C}_\bullet) \mapsto f_n z_n + B_n(\mathbf{C}'_\bullet).$$

We first show that $f_n z_n$ is a cycle and that $H_n(f)$ is independent of the choice of cycle $z_n$; both of these follow from $f$ being a chain map, that is, from commutativity of the following diagram:

$$\begin{array}{ccccc} C_{n+1} & \xrightarrow{d_{n+1}} & C_n & \xrightarrow{d_n} & C_{n-1} \\ \downarrow{\scriptstyle f_{n+1}} & & \downarrow{\scriptstyle f_n} & & \downarrow{\scriptstyle f_{n-1}} \\ C'_{n+1} & \xrightarrow[d'_{n+1}]{} & C'_n & \xrightarrow[d'_n]{} & C'_{n-1}. \end{array}$$

If $z$ is an $n$-cycle in $Z_n(\mathbf{C}_\bullet)$, so that $d_n z = 0$, then commutativity of the diagram gives

$$d'_n f_n z = f_{n-1} d_n z = 0.$$

Therefore, $f_n z$ is an $n$-cycle.

Next, assume that $z + B_n(\mathbf{C}_\bullet) = y + B_n(\mathbf{C}_\bullet)$; hence, $z - y \in B_n(\mathbf{C}_\bullet)$; that is,

$$z - y = d_{n+1} c$$

for some $c \in C_{n+1}$. Applying $f_n$ gives

$$f_n z - f_n y = f_n d_{n+1} c = d'_{n+1} f_{n+1} c \in B_n(\mathbf{C}'_\bullet).$$

Thus, $f_n z + B_n(\mathbf{C}'_\bullet) = f_n y + B_n(\mathbf{C}'_\bullet)$; that is, $H_n(f)$ is well-defined.

Let us see that $H_n$ is a functor. It is obvious that $H_n(1_{\mathbf{C}_\bullet})$ is the identity. If $f$ and $g$ are chain maps whose composite $gf$ is defined, then for every $n$-cycle $z$, we have

$$\begin{aligned} H_n(gf) \colon z + B &\mapsto (gf)_n(z + B) \\ &= g_n f_n (z + B) \\ &= H_n(g)(f_n z + B) \\ &= H_n(g) H_n(f)(z + B). \end{aligned}$$

Finally, $H_n$ is additive: if $g\colon (\mathbf{C}_\bullet, d_\bullet) \to (\mathbf{C}'_\bullet, d'_\bullet)$ is another chain map, then

$$\begin{aligned} H_n(f+g)\colon z + B_n(\mathbf{C}_\bullet) &\mapsto (f_n + g_n)z + B_n(\mathbf{C}'_\bullet) \\ &= f_n z + g_n z + B_n(\mathbf{C}'_\bullet) \\ &= \big(H_n(f) + H_n(g)\big)(z + B_n(\mathbf{C}'_\bullet)). \quad \bullet \end{aligned}$$

**Definition.** We call $H_n(f)$ the ***induced map***, and we usually denote it by $f_{n*}$, or even by $f_*$.

**Proposition C-3.38.** *Let $R$ and $A$ be rings, and let $T\colon {}_R\mathbf{Mod} \to {}_A\mathbf{Mod}$ be an exact additive functor. Then $T$ commutes with homology; that is, for every complex $(\mathbf{C}_\bullet, d_\bullet) \in {}_R\mathbf{Comp}$ and for every $n \in \mathbb{Z}$, there is an isomorphism*

$$H_n(T\mathbf{C}_\bullet, Td_\bullet) \cong TH_n(\mathbf{C}_\bullet, d_\bullet).$$

**Proof.** Consider the commutative diagram with exact bottom row,

$$\begin{array}{ccccccccc} & & C_{n+1} & \xrightarrow{d_{n+1}} & C_n & \xrightarrow{d_n} & C_{n-1} & & \\ & & \downarrow{\scriptstyle d'_{n+1}} & & \uparrow{\scriptstyle k} & & & & \\ 0 & \longrightarrow & \operatorname{im} d_{n+1} & \xrightarrow{j} & \ker d_n & \longrightarrow & H_n(\mathbf{C}_\bullet) & \longrightarrow & 0, \end{array}$$

where $j$ and $k$ are inclusions and $d'_{n+1}$ is just $d_{n+1}$ with its target changed from $C_n$ to $\operatorname{im} d_{n+1}$. Applying the exact functor $T$ gives the commutative diagram with exact bottom row

$$\begin{array}{ccccccccc} & & TC_{n+1} & \xrightarrow{Td_{n+1}} & TC_n & \xrightarrow{Td_n} & TC_{n-1} & & \\ & & \downarrow{\scriptstyle Td'_{n+1}} & & \uparrow{\scriptstyle Tk} & & & & \\ 0 & \longrightarrow & T(\operatorname{im} d_{n+1}) & \xrightarrow{Tj} & T(\ker d_n) & \longrightarrow & TH_n(\mathbf{C}_\bullet) & \longrightarrow & 0. \end{array}$$

On the other hand, because $T$ is exact, we have $T(\operatorname{im} d_{n+1}) = \operatorname{im} T(d_{n+1})$ and $T(\ker d_n) = \ker(Td_n)$, so that the bottom row is

$$0 \to \operatorname{im}(Td_{n+1}) \to \ker(Td_n) \to TH_n(\mathbf{C}_\bullet) \to 0.$$

By definition, $\ker(Td_n)/\operatorname{im}(Td_{n+1}) = H_n(T\mathbf{C}_\bullet)$, and so $H_n(T\mathbf{C}_\bullet) \cong TH_n(\mathbf{C}_\bullet)$, by Proposition B-1.46 in Part 1. $\bullet$

We now introduce an algebraic version of a notion that arises in topology.

**Definition.** A chain map $f\colon (\mathbf{C}_\bullet, d_\bullet) \to (\mathbf{C}'_\bullet, d'_\bullet)$ is ***nullhomotopic*** if, for all $n$, there are maps $s_n\colon A_n \to A'_{n+1}$ with

$$f_n = d'_{n+1}s_n + s_{n-1}d_n,$$

$$\begin{array}{ccccccccccc} \cdots & \longrightarrow & A_{n+1} & \xrightarrow{d_{n+1}} & A_n & \xrightarrow{d_n} & A_{n-1} & \longrightarrow & \cdots & & \\ & & \downarrow{\scriptstyle f_{n+1}} & \swarrow{\scriptstyle s_n} & \downarrow{\scriptstyle f_n} & \swarrow{\scriptstyle s_{n-1}} & \downarrow{\scriptstyle f_{n-1}} & & & & \\ \cdots & \longrightarrow & A'_{n+1} & \xrightarrow[d'_{n+1}]{} & A'_n & \xrightarrow[d'_n]{} & A'_{n-1} & \longrightarrow & \cdots. & & \end{array}$$

If $f, g\colon (\mathbf{C}_\bullet, d_\bullet) \to (\mathbf{C}'_\bullet, d'_\bullet)$ are chain maps, then $f$ is ***homotopic***[10] to $g$, denoted by $f \backsimeq g$, if $f - g$ is nullhomotopic.

**Proposition C-3.39.** *Homotopic chain maps induce the same homomorphism between homology groups: if $f, g\colon (\mathbf{C}_\bullet, d_\bullet) \to (\mathbf{C}'_\bullet, d'_\bullet)$ are chain maps and $f \backsimeq g$, then*

$$f_{*n} = g_{*n}\colon H_n(\mathbf{C}_\bullet) \to H_n(\mathbf{C}'_\bullet).$$

**Proof.** If $z$ is an $n$-cycle, then $d_n z = 0$ and

$$f_n z - g_n z = d'_{n+1} s_n z + s_{n-1} d_n z = d'_{n+1} s_n z.$$

Therefore, $f_n z - g_n z \in B_n(\mathbf{C}'_\bullet)$, and so $f_{*n} = g_{*n}$. •

**Definition.** A complex $(\mathbf{C}_\bullet, d_\bullet)$ has a ***contracting homotopy***[11] if its identity $1_{\mathbf{C}_\bullet}$ is nullhomotopic.

**Proposition C-3.40.** *A complex $(\mathbf{C}_\bullet, d_\bullet)$ having a contracting homotopy is acyclic; that is, it is an exact sequence.*

**Proof.** We use Example C-3.35. Now $1_{\mathbf{C}_\bullet}\colon H_n(\mathbf{C}_\bullet) \to H_n(\mathbf{C}_\bullet)$ is the identity map, while $0_*\colon H_n(\mathbf{C}_\bullet) \to H_n(\mathbf{C}_\bullet)$ is the zero map. Since $1_{\mathbf{C}_\bullet} \backsimeq 0$, however, these maps are the same. It follows that $H_n(\mathbf{C}_\bullet) = \{0\}$ for all $n$; that is, $\ker d_n = \operatorname{im} d_{n+1}$ for all $n$, and this is the definition of exactness. •

After completing the definition of the free resolution of the trivial $Q$-module $\mathbb{Z}$ whose first few terms were given in Proposition C-3.31, we will prove exactness (in Proposition C-3.30) by showing that it has a contracting homotopy.

The following elementary construction is fundamental; it gives a relation between different homology modules. The proof is a series of diagram chases. Ordinarily, we would just say that the proof is routine, but, because of the importance of the result, we present (perhaps too many) details; as a sign that the proof is routine, we drop subscripts.

**Theorem C-3.41 (Connecting Homomorphism).** *If*

$$\mathbf{0}_\bullet \to \mathbf{C}'_\bullet \xrightarrow{i} \mathbf{C}_\bullet \xrightarrow{p} \mathbf{C}''_\bullet \to \mathbf{0}_\bullet$$

*is an exact sequence of complexes, then, for each $n \in \mathbb{Z}$, there is a homomorphism*

$$\partial_n\colon H_n(\mathbf{C}''_\bullet) \to H_{n-1}(\mathbf{C}'_\bullet)$$

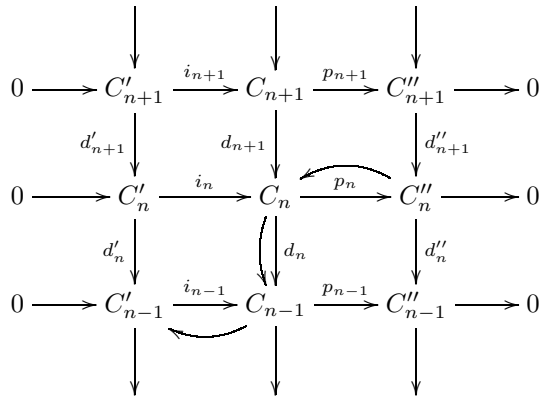
*defined by*

$$\partial_n\colon z''_n + B_n(\mathbf{C}''_\bullet) \mapsto i_{n-1}^{-1} d_n p_n^{-1} z''_n + B_{n-1}(\mathbf{C}'_\bullet).$$

[10] Recall (from page 87) that two continuous functions $f, g\colon X \to Y$ are ***homotopic*** if $f$ can be "deformed" into $g$; that is, there exists a continuous $F\colon X \times \mathbf{I} \to Y$, where $\mathbf{I} = [0, 1]$ is the closed unit interval, with $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$ for all $x \in X$. In topology, one proves that every continuous $f\colon X \to Y$ induces homomorphisms $f_*\colon H_n(X) \to H_n(Y)$, and if $f$ and $g$ are homotopic, then $f_* = g_*$. The algebraic definition of homotopy given here has been distilled from the proof of this topological theorem.

[11] A topological space is called ***contractible*** if its identity map is homotopic to a constant map.

**Proof.** We will make many notational abbreviations in this proof. Consider the commutative diagram having exact rows:




Suppose that $z'' \in C''_n$ and $d''z'' = 0$. Since $p_n$ is surjective, there is $c \in C_n$ with $pc = z''$. Now push $c$ down to $dc \in C_{n-1}$. By commutativity, $p_{n-1}dc = d''p_nc = d''z'' = 0$, so that $dc \in \ker p_{n-1} = \operatorname{im} i_{n-1}$. Therefore, there is a unique $c' \in C'_{n-1}$ with $i_{n-1}c' = dc$, for $i_{n-1}$ is an injection. Thus, $i_{n-1}^{-1}dp_n^{-1}z''$ makes sense; that is, the claim is that

$$\partial_n(z'' + B''_n) = c' + B'_{n-1}$$

is a well-defined homomorphism.

First, let us show independence of the choice of lifting. Suppose that $p_n\check{c} = z''$, where $\check{c} \in C_n$. Then $c - \check{c} \in \ker p_n = \operatorname{im} i_n$, so that there is $u' \in C'_n$ with $i_nu' = c - \check{c}$. By commutativity of the first square, we have

$$i_{n-1}d'u' = di_nu' = dc - d\check{c}.$$

Hence, $i^{-1}dc - i^{-1}d\check{c} = d'u' \in B'_{n-1}$; that is, $i^{-1}dc + B'_{n-1} = i^{-1}d\check{c} + B'_{n-1}$. Thus, the formula gives a well-defined function

$$Z''_n \to C'_{n-1}/B'_{n-1}.$$

Second, the function $Z''_n \to C'_{n-1}/B'_{n-1}$ is a homomorphism. If $z'', z''_1 \in Z''_n$, let $pc = z''$ and $pc_1 = z''_1$. Since the definition of $\partial$ is independent of the choice of lifting, choose $c + c_1$ as a lifting of $z'' + z''_1$. This step may now be completed in a routine way.

Third, we show that if $i_{n-1}c' = dc$, then $c'$ is a cycle: $0 = ddc = dic' = idc'$, and so $d'c' = 0$ because $i$ is an injection. Hence, the formula gives a homomorphism

$$Z'' \to Z'/B' = H_{n-1}.$$

Finally, the subgroup $B''_n$ goes into $B'_{n-1}$. Suppose that $z'' = d''c''$, where $c'' \in C''_{n+1}$, and let $pu = c''$, where $u \in C_{n+1}$. Commutativity gives $pdu = d''pu = d''c'' = z''$. Since $\partial(z'')$ is independent of the choice of lifting, we choose $du$ with $pdu = z''$, and so $\partial(z'' + B'') = i^{-1}d(du) + B' = B'$. Therefore, the formula does give a homomorphism $\partial_n \colon H_n(\mathbf{C}''_\bullet) \to H_{n-1}(\mathbf{C}'_\bullet)$. •

The first question we ask is what homology functors do to a short exact sequence of complexes. The next theorem is also proved by diagram chasing and, again, we give too many details because of the importance of the result. The reader should try to prove the theorem before looking at the proof we give.

**Theorem C-3.42 (Long Exact Sequence).** *If*

$$\mathbf{0}_\bullet \to \mathbf{C}'_\bullet \xrightarrow{i} \mathbf{C}_\bullet \xrightarrow{p} \mathbf{C}''_\bullet \to \mathbf{0}_\bullet$$

*is an exact sequence of complexes, then there is an exact sequence of modules*

$$\cdots \to H_{n+1}(\mathbf{C}''_\bullet) \xrightarrow{\partial_{n+1}} H_n(\mathbf{C}'_\bullet) \xrightarrow{i_*} H_n(\mathbf{C}_\bullet) \xrightarrow{p_*} H_n(\mathbf{C}''_\bullet) \xrightarrow{\partial_n} H_{n-1}(\mathbf{C}'_\bullet) \to \cdots .$$

**Proof.** This proof is also routine. Our notation is abbreviated, and there are six inclusions to verify.

(i) $\operatorname{im} i_* \subseteq \ker p_*$.

$$p_* i_* = (pi)_* = 0_* = 0.$$

(ii) $\ker p_* \subseteq \operatorname{im} i_*$.

If $p_*(z + B) = pz + B'' = B''$, then $pz = d''c''$ for some $c'' \in C''_{n+1}$. But $p$ surjective gives $c'' = pc$ for some $c \in C_{n+1}$, so that $pz = d''pc = pdc$, because $p$ is a chain map, and so $p(z - dc) = 0$. By exactness, there is $c' \in C'_n$ with $ic' = z - dc$. Now $c'$ is a cycle, for $id'c' = dic' = dz - ddc = 0$, because $z$ is a cycle; since $i$ is injective, $d'c' = 0$. Therefore,

$$i_*(c' + B') = ic' + B = z - dc + B = z + B.$$

(iii) $\operatorname{im} p_* \subseteq \ker \partial$.

If $p_*(c + B) = pc + B'' \in \operatorname{im} p_*$, then $\partial(pz + B'') = z' + B'$, where $iz' = dp^{-1}pz$. Since this formula is independent of the choice of lifing of $pz$, let us choose $p^{-1}pz = z$. Now $dp^{-1}pz = dz = 0$, because $z$ is a cycle. Thus, $iz' = 0$, and hence $z' = 0$, because $i$ is injective.

(iv) $\ker \partial \subseteq \operatorname{im} p_*$.

If $\partial(z'' + B'') = B'$, then $z' = i^{-1}dp^{-1}z'' \in B'$; that is, $z' = d'c'$ for some $c' \in C'$. But $iz' = id'c' = dic' = dp^{-1}z''$, so that $d(p^{-1}z'' - ic') = 0$; that is, $p^{-1}z'' - ic'$ is a cycle. Moreover, since $pi = 0$ because of exactness of the original sequence,

$$p_*(p^{-1}z'' - ic' + B) = pp^{-1}z'' - pic' + B'' = z'' + B''.$$

(v) $\operatorname{im} \partial \subseteq \ker i_*$.

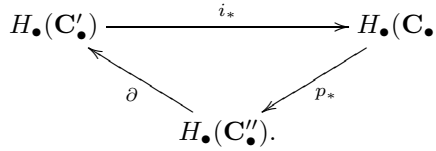
We have $i_*\partial(z'' + B'') = iz' + B'$, where $iz' = dp^{-1}z'' \in B$; that is, $i_*\partial = 0$.

(vi) $\ker i_* \subseteq \operatorname{im} \partial$.

If $i_*(z' + B') = iz' + B = B$, then $iz' = dc$ for some $c \in C$. Since $p$ is a chain map, $d''pc = pdc = piz' = 0$, by exactness of the original sequence, and so $pc$ is a cycle. But

$$\partial(pc + B'') = i^{-1}dp^{-1}pc + B' = i^{-1}dc + B' = i^{-1}iz' + B' = z' + B'. \quad \bullet$$

Theorem C-3.42 is often called the ***exact triangle*** because of the diagram




The following corollary received its name because of the curved path in the diagram occurring in the proof of Theorem C-3.41.

**Corollary C-3.43 (Snake Lemma).** *Given a commutative diagram of modules with exact rows,*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A' & \longrightarrow & A & \longrightarrow & A'' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & B' & \longrightarrow & B & \longrightarrow & B'' & \longrightarrow & 0, \end{array}$$

*there is an exact sequence*

$$0 \to \ker f \to \ker g \to \ker h \to \operatorname{coker} f \to \operatorname{coker} g \to \operatorname{coker} h \to 0.$$

**Proof.** If we view each of the vertical maps $f$, $g$, and $h$ as a complex (as in Example C-3.34(v)), then the given commutative diagram can be viewed as a short exact sequence of complexes. The homology groups of each of these complexes has only two nonzero terms: for example, Example C-3.36 shows that the homology groups of the first column are $H_1 = \ker f$, $H_0 = \operatorname{coker} f$, and all other $H_n = \{0\}$. The Snake Lemma now follows at once from the long exact sequence. •

**Theorem C-3.44 (Naturality of $\partial$).** *Given a commutative diagram of complexes with exact rows,*

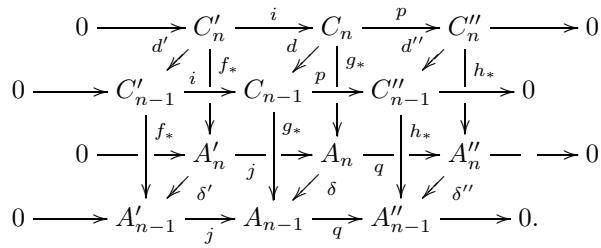
$$\begin{array}{ccccccccc} \mathbf{0}_\bullet & \longrightarrow & \mathbf{C}'_\bullet & \xrightarrow{i} & \mathbf{C}_\bullet & \xrightarrow{p} & \mathbf{C}''_\bullet & \longrightarrow & \mathbf{0}_\bullet \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ \mathbf{0}_\bullet & \longrightarrow & \mathbf{A}'_\bullet & \xrightarrow{j} & \mathbf{A}_\bullet & \xrightarrow{q} & \mathbf{A}''_\bullet & \longrightarrow & \mathbf{0}_\bullet, \end{array}$$

*there is a commutative diagram of modules with exact rows,*

$$\begin{array}{ccccccccccc} \cdots & \longrightarrow & H_n(\mathbf{C}'_\bullet) & \xrightarrow{i_*} & H_n(\mathbf{C}_\bullet) & \xrightarrow{p_*} & H_n(\mathbf{C}''_\bullet) & \xrightarrow{\partial} & H_{n-1}(\mathbf{C}'_\bullet) & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle f_*} & & \downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle h_*} & & \downarrow{\scriptstyle f_*} & & \\ \cdots & \longrightarrow & H_n(\mathbf{A}'_\bullet) & \xrightarrow{j_*} & H_n(\mathbf{A}_\bullet) & \xrightarrow{q_*} & H_n(\mathbf{A}''_\bullet) & \xrightarrow{\partial'} & H_{n-1}(\mathbf{A}'_\bullet) & \longrightarrow & \cdots. \end{array}$$

**Proof.** Exactness of the rows is Theorem C-3.42, while commutativity of the first two squares follows from $H_n$ being a functor. To prove commutativity of the square involving the connecting homomorphism, let us first display the chain maps and

differentiations in one (three-dimensional!) diagram:



If $z'' + B(\mathbf{C}''_\bullet) \in H_n(\mathbf{C}''_\bullet)$, we must show that

$$f_*\partial(z'' + B(\mathbf{C}''_\bullet)) = \partial' h_*(z'' + B(\mathbf{C}''_\bullet)).$$

Let $c \in C_n$ be a lifting of $z''$; that is, $pc = z''$. Now $\partial(z'' + B(\mathbf{C}''_\bullet)) = z' + B(\mathbf{C}'_\bullet)$, where $iz' = dc$. Hence, $f_*\partial(z'' + B(\mathbf{A}''_\bullet)) = fz' + B(\mathbf{A}'_\bullet)$. On the other hand, since $h$ is a chain map, we have $qgc = hpc = hz''$. In computing $\partial'(hz'' + B(\mathbf{A}''_\bullet))$, we choose $gc$ as the lifting of $hz''$. Hence, $\partial'(hz'' + B(\mathbf{A}''_\bullet)) = u' + B(\mathbf{A}'_\bullet)$, where $ju' = \delta gc$. But

$$jfz' = giz' = gdc = \delta gc = ju',$$

and so $fz' = u'$, because $j$ is injective. •

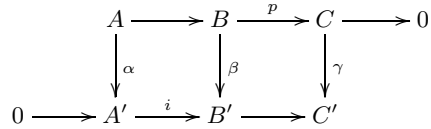
We shall apply these general results in the next section.

## Exercises

∗ **C-3.23.** If $\mathbf{C}_\bullet$ is a complex with $C_n = \{0\}$ for some $n$, prove that $H_n(\mathbf{C}_\bullet) = \{0\}$.

∗ **C-3.24.** Prove that isomorphic complexes have the same homology: if $\mathbf{C}_\bullet$ and $\mathbf{D}_\bullet$ are isomorphic, then $H_n(\mathbf{C}_\bullet) \cong H_n(\mathbf{D}_\bullet)$ for all $n$.

∗ **C-3.25.** In this exercise, we prove that the Snake Lemma implies the Long Exact Sequence (the converse is Corollary C-3.43). Consider a commutative diagram with exact rows (note that two zeros are "missing" from this diagram):



(i) Prove that $\Delta\colon \ker\gamma \to \operatorname{coker}\alpha$, defined by

$$\Delta\colon z \mapsto i^{-1}\beta p^{-1}z + \operatorname{im}\alpha,$$

is a well-defined homomorphism.

(ii) Prove that there is an exact sequence

$$\ker\alpha \to \ker\beta \to \ker\gamma \xrightarrow{\Delta} \operatorname{coker}\alpha \to \operatorname{coker}\beta \to \operatorname{coker}\gamma.$$

(iii) Given a commutative diagram with exact rows,

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A'_n & \longrightarrow & A_n & \longrightarrow & A''_n & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle d'_n} & & \downarrow{\scriptstyle d} & & \downarrow{\scriptstyle d''_n} & & \\ 0 & \longrightarrow & A'_{n-1} & \longrightarrow & A_{n-1} & \longrightarrow & A''_{n-1} & \longrightarrow & 0, \end{array}$$

prove that the following diagram is commutative and has exact rows:

$$\begin{array}{ccccccccc} & & A'_n/\operatorname{im} d'_{n+1} & \longrightarrow & A_n/\operatorname{im} d_{n+1} & \longrightarrow & A''_n/\operatorname{im} d''_{n+1} & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle d'} & & \downarrow{\scriptstyle d} & & \downarrow{\scriptstyle d''} & & \\ 0 & \longrightarrow & \ker d'_{n-1} & \longrightarrow & \ker d_{n-1} & \longrightarrow & \ker d''_{n-1} & & \end{array}$$

(iv) Use (ii) and this last diagram to give another proof of the Long Exact Sequence.

∗ **C-3.26.** Let $f, g\colon \mathbf{C}_\bullet \to \mathbf{C}'_\bullet$ be chain maps, and let $F\colon \mathbf{C}_\bullet \to \mathbf{C}'_\bullet$ be an additive functor. If $f \simeq g$, prove that $Ff \simeq Fg$; that is, if $f$ and $g$ are homotopic, then $Ff$ and $Fg$ are homotopic.

**C-3.27.** Let $\mathbf{0}_\bullet \to \mathbf{C}'_\bullet \xrightarrow{i} \mathbf{C}_\bullet \xrightarrow{p} \mathbf{C}''_\bullet \to \mathbf{0}_\bullet$ be an exact sequence of complexes in which $\mathbf{C}'_\bullet$ and $\mathbf{C}''_\bullet$ are acyclic; prove that $\mathbf{C}_\bullet$ is also acyclic.

**C-3.28.** Let $(\mathbf{C}_\bullet, d_\bullet)$ be a complex each of whose differentiations $d_n$ is the zero map. Prove that $H_n(\mathbf{C}_\bullet) \cong C_n$ for all $n$.

∗ **C-3.29.** Prove that homology commutes with direct sums: for all $n$, there are natural isomorphisms

$$H_n\Big(\sum_\alpha \mathbf{C}^\alpha_\bullet\Big) \cong \sum_\alpha H_n(\mathbf{C}^\alpha_\bullet).$$

∗ **C-3.30.** (i) Define a direct system of complexes $\{\mathbf{C}^i_\bullet, \varphi^i_j\}$, and prove that $\varinjlim \mathbf{C}^i_\bullet$ exists.

(ii) If $\{\mathbf{C}^i_\bullet, \varphi^i_j\}$ is a direct system of complexes over a directed index set, prove, for all $n \geq 0$, that

$$H_n(\varinjlim \mathbf{C}^i_\bullet) \cong \varinjlim H_n(\mathbf{C}^i_\bullet).$$

∗ **C-3.31.** Suppose that a complex $(\mathbf{C}_\bullet, d_\bullet)$ of $R$-modules has a contracting homotopy in which the maps $s_n\colon C_n \to C_{n+1}$ satisfying

$$1_{C_n} = d_{n+1}s_n + s_{n-1}d_n$$

are only $\mathbb{Z}$-maps. Prove that $(\mathbf{C}_\bullet, d_\bullet)$ is an exact sequence.

**C-3.32.** (i) Let $\mathbf{C}_\bullet : 0 \to C_n \to C_{n-1} \to \cdots \to C_0 \to 0$ be a complex of finitely generated free $k$-modules over a commutative ring $k$. Prove that

$$\sum_{i=0}^{n}(-1)^i \operatorname{rank}(C_i) = \sum_{i=0}^{n}(-1)^i \operatorname{rank}(H_i(\mathbf{C}_\bullet)).$$

**Hint.** See Exercise C-3.22 on page 261.

(ii) Let $0 \to M' \to M \to M'' \to 0$ be an exact sequence of $k$-modules. If two of the modules have an Euler–Poincaré characteristic, prove that the third module does, too, and that

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

∗ **C-3.33.** (i) **(Barratt–Whitehead).** Consider the commutative diagram with exact rows:

$$\begin{array}{ccccccccccc}
A_n & \xrightarrow{i_n} & B_n & \xrightarrow{p_n} & C_n & \xrightarrow{\partial_n} & A_{n-1} & \longrightarrow & B_{n-1} & \longrightarrow & C_{n-1} \\
\downarrow{\scriptstyle f_n} & & \downarrow{\scriptstyle g_n} & & \downarrow{\scriptstyle h_n} & & \downarrow{\scriptstyle f_{n-1}} & & \downarrow{\scriptstyle g_{n-1}} & & \downarrow{\scriptstyle h_{n-1}} \\
A'_n & \xrightarrow[j_n]{} & B'_n & \xrightarrow[q_n]{} & C'_n & \longrightarrow & A'_{n-1} & \longrightarrow & B'_{n-1} & \longrightarrow & C'_{n-1}.
\end{array}$$

If each $h_n$ is an isomorphism, prove that there is an exact sequence

$$A_n \xrightarrow{(f_n, i_n)} A'_n \oplus B_n \xrightarrow{j_n - q_n} B'_n \xrightarrow{\partial_n h_n^{-1} q_n} A_{n-1} \to A'_{n-1} \oplus B_{n-1} \to B'_{n-1},$$

where $(f_n, i_n)\colon a_n \mapsto (f_n a_n, i_n a_n)$ and $j_n - g_n\colon (a'_n, b_n) \mapsto j_n a'_n - g_n b_n$.

(ii) **(Mayer–Vietoris).** Assume, in the second diagram of Theorem C-3.44, that every third vertical map $h_*$ is an isomorphism. Prove that there is an exact sequence

$$\cdots \to H_n(\mathbf{C}'_\bullet) \to H_n(\mathbf{A}'_\bullet) \oplus H_n(\mathbf{C}_\bullet) \to H_n(\mathbf{A}_\bullet) \to H_{n-1}(\mathbf{C}'_\bullet) \to \cdots .$$

**Remark.** If $A$ is a subspace of a (triangulated)[12] topological space $X$, then the chain complex $C_\bullet(A)$ is a subcomplex of $C_\bullet(X)$, and one defines the $n$th ***relative homology group*** $H_n(X, A)$ as the $n$th homology of the quotient complex:

$$H_n(X, A) = H_n(\mathbf{C}_\bullet(X)/\mathbf{C}_\bullet(A)).$$

Note that if $A = \varnothing$, then $H_n(X, \varnothing) = H_n(X)$.

The ***Eilenberg–Steenrod axioms*** (see Spanier [**210**], pp. 199–205) characterize homology functors on the category $\mathbf{Top}^2$ having objects all pairs $(X, A)$ of topological spaces with $A$ a subspace of $X$ and morphisms $f\colon (X, A) \to (X', A')$, where $f\colon X \to X'$ is a continuous function with $f(A) \subseteq A'$. Assume that $h_n\colon \mathbf{Top^2} \to \mathbf{Ab}$ is a sequence of functors, for $n \geq 0$, satisfying the long exact sequence, naturality of connecting homomorphisms, $h_n(f) = h_n(g)$ whenever $f$ and $g$ are homotopic, $h_0(X) = \mathbb{Z}$ and $h_n(X) = \{0\}$ for all $n > 0$ when $X$ is a 1-point space, and ***excision***: given a pair $(X, A)$ in $\mathbf{Top}^2$ and an open set $U \subseteq X$ whose closure is contained in the interior of $A$, then the inclusion $(X - U, A - U) \to (X, A)$ induces isomorphisms $h_n(X - U, A - U) \to h_n(X, A)$ for all $n \geq 0$. Then there are natural isomorphisms $h_n \to H_n$ for all $n$. In the presence of the other axioms, excision can be replaced by exactness of the Mayer–Vietoris sequence. ◀

## C-3.6. Derived Functors

In order to apply the general results about homology, we need a source of short exact sequences of complexes, as well as commutative diagrams in which they sit. As we have said earlier, the idea is to replace a module by a resolution of it. We then apply either Hom or $\otimes$, and the resulting homology modules are called Ext or Tor. Given a short exact sequence of modules, we shall see that we may replace each of its modules by a resolution and obtain a short exact sequence of complexes.

This section is fairly dry, but it is necessary to establish the existence of homology functors. The most useful theorems in this section are Theorem C-3.46 (Comparison Theorem), Proposition C-3.50 (the basic construction is well-defined),

[12] That $X$ is triangulated is too strong a hypothesis. Using *singular theory*, homology groups of arbitrary topological spaces can be defined.

Corollary C-3.57 (Long Exact Sequence), and Proposition C-3.58 (Naturality of the Connecting Homomorphism).

For those readers who are interested in using Ext (the right derived functors of Hom) or Tor (the left derived functors of tensor) immediately and who are willing to defer looking at mazes of arrows, the next theorem gives a set of axioms characterizing the functors $\mathrm{Ext}^n$.

**Theorem C-3.45.** *Let* $\mathrm{EXT}^n\colon {}_R\mathbf{Mod} \to \mathbf{Ab}$ *be a sequence of contravariant functors, for* $n \geq 0$*, such that*

(i) *for every short exact sequence* $0 \to A \to B \to C \to 0$*, there is a long exact sequence and natural connecting homomorphisms*

$$\cdots \to \mathrm{EXT}^n(C) \to \mathrm{EXT}^n(B) \to \mathrm{EXT}^n(A) \xrightarrow{\Delta_n} \mathrm{EXT}^{n+1}(C) \to \cdots;$$

(ii) *there is* $M \in \mathrm{obj}_R\mathbf{Mod}$ *with* $\mathrm{EXT}^0$ *and* $\mathrm{Hom}_R(\ ,M)$ *naturally equivalent*;

(iii) $\mathrm{EXT}^n(P) = \{0\}$ *for all projective modules* $P$ *and all* $n \geq 1$.

*If* $\mathrm{Ext}^n(\ ,M)$ *is another sequence of contravariant functors satisfying these same axioms, then* $\mathrm{EXT}^n$ *is naturally equivalent to* $\mathrm{Ext}^n(\ ,M)$ *for all* $n \geq 0$.

**Remark.** There are similar axiomatic descriptions of the covariant Ext functors and of the Tor functors in Exercises C-3.43 and C-3.44 on page 308. ◀

**Proof.** We proceed by induction on $n \geq 0$. The base step is axiom (ii).

For the inductive step, given a module $A$, choose a short exact sequence

$$0 \to L \to P \to A \to 0,$$

where $P$ is projective. By axiom (i), there is a diagram with exact rows:

$$\begin{array}{ccccccc}
\mathrm{EXT}^0(P) & \longrightarrow & \mathrm{EXT}^0(L) & \xrightarrow{\Delta_0} & \mathrm{EXT}^1(A) & \longrightarrow & \mathrm{EXT}^1(P) \\
\downarrow{\scriptstyle \tau_P} & & \downarrow{\scriptstyle \tau_L} & & \| & & \\
\mathrm{Hom}(P,M) & \longrightarrow & \mathrm{Hom}(L,M) & \xrightarrow{\partial_0} & \mathrm{Ext}^1(A,M) & \longrightarrow & \mathrm{Ext}^1(P,M),
\end{array}$$

where the maps $\tau_P$ and $\tau_L$ are the isomorphisms given by axiom (ii). This diagram commutes because of the naturality of the equivalence $\mathrm{EXT}^0 \to \mathrm{Hom}(\ ,M)$. By axiom (iii), $\mathrm{Ext}^1(P,M) = \{0\}$ and $\mathrm{EXT}^1(P) = \{0\}$. It follows that the maps $\Delta_0$ and $\partial_0$ are surjective. This is precisely the sort of diagram in Proposition B-1.46 in Part 1, and so there exists an isomorphism $\mathrm{EXT}^1(A) \to \mathrm{Ext}^1(A,M)$ making the augmented diagram commute.

We may now assume that $n \geq 1$, and we look further out in the long exact sequence. By axiom (i), there is a diagram with exact rows

$$\begin{array}{ccccccc}
\mathrm{EXT}^n(P) & \longrightarrow & \mathrm{EXT}^n(L) & \xrightarrow{\Delta_n} & \mathrm{EXT}^{n+1}(A) & \longrightarrow & \mathrm{EXT}^{n+1}(P) \\
 & & \downarrow{\scriptstyle \sigma} & & \| & & \\
\mathrm{Ext}^n(P,M) & \longrightarrow & \mathrm{Ext}^n(L,M) & \xrightarrow{\partial_n} & \mathrm{Ext}^{n+1}(A,M) & \longrightarrow & \mathrm{Ext}^{n+1}(P,M),
\end{array}$$

where $\sigma\colon \operatorname{EXT}^n(L) \to \operatorname{Ext}^n(L,M)$ is an isomorphism given by the inductive hypothesis. Since $n \geq 1$, all four terms involving the projective $P$ are $\{0\}$; it follows from exactness of the rows that both $\Delta_n$ and $\partial_n$ are isomorphisms. Finally, the composite $\partial_n \sigma \Delta_n^{-1}\colon \operatorname{EXT}^{n+1}(A) \to \operatorname{Ext}^{n+1}(A,M)$ is an isomorphism.

It remains to prove that the isomorphisms $\operatorname{EXT}^n(A) \to \operatorname{Ext}^n(A,M)$ constitute a natural transformation. It is here the assumed naturality in axiom (i) of the connecting homomorphism is used, and this is left for the reader to do. •

Such slow starting induction proofs, proving results for $n = 0$ and $n = 1$ before proving the inductive step, arise frequently, and they are called ***dimension shifting***.

The rest of this section consists of constructing functors that satisfy axioms (i), (ii), and (iii). We prove existence of Ext and Tor using derived functors (there are other proofs as well). As these functors are characterized by a short list of properties, we can usually work with Ext and Tor without being constantly aware of the details of their construction.

We begin with a technical definition.

**Definition.** If $\cdots \to P_2 \to P_1 \xrightarrow{d_1} P_0 \to A \to 0$ is a projective resolution of a module $A$, then its ***deleted projective resolution*** is the complex

$$\mathbf{P}_A = \cdots \to P_2 \to P_1 \to P_0 \to 0.$$

Similarly, if $0 \to A \to E^0 \xrightarrow{d^0} E^1 \to E^2 \to \cdots$ is an injective resolution of a module $A$, then a ***deleted injective resolution*** is the complex

$$\mathbf{E}^A = 0 \to E^0 \to E^1 \to E^2 \to \cdots.$$

In either case, deleting $A$ loses no information: $A \cong \operatorname{coker} d_1$ in the first case, and $A \cong \ker d^0$ in the second case. Of course, a deleted resolution is no longer exact:

$$H_0(\mathbf{P}_A) = \ker(P_0 \to \{0\})/\operatorname{im} d_1 = P_0/\operatorname{im} d_1 \cong A.$$

We know that a module has many presentations, and so the next result is fundamental.

**Theorem C-3.46 (Comparison Theorem).** *Given a map $f\colon A \to A'$, consider the diagram*

$$\begin{array}{ccccccccccc}
\cdots & \longrightarrow & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\varepsilon} & A & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \check{f}_2} & & \downarrow{\scriptstyle \check{f}_1} & & \downarrow{\scriptstyle \check{f}_0} & & \downarrow{\scriptstyle f} & & \\
\cdots & \longrightarrow & P'_2 & \xrightarrow[d'_2]{} & P'_1 & \xrightarrow[d'_1]{} & P'_0 & \xrightarrow[\varepsilon']{} & A' & \longrightarrow & 0,
\end{array}$$

*where the rows are complexes. If each $P_n$ in the top row is projective and if the bottom row is exact, then there exists a chain map $\check{f}\colon \mathbf{P}_A \to \mathbf{P}'_{A'}$ making the completed diagram commute. Moreover, any two such chain maps are homotopic.*

**Remark.** The dual of the Comparison Theorem is also true. Now the complexes go off to the right, the top row is assumed exact, and every term in the bottom row other than $A'$ is injective. ◀

**Proof.**

(i) We prove the existence of $\check{f}_n$ by induction on $n \geq 0$. For the base step $n = 0$, consider the diagram

$$\begin{array}{ccccc} & & P_0 & & \\ & \swarrow & \downarrow{\scriptstyle f\varepsilon} & & \\ P_0' & \xrightarrow{\varepsilon'} & A' & \longrightarrow & 0. \end{array}$$

Since $\varepsilon'$ is surjective and $P_0$ is projective, there exists a map $\check{f}_0 \colon P_0 \to P_0'$ with $\varepsilon' \check{f}_0 = f\varepsilon$.

For the inductive step, consider the diagram

$$\begin{array}{ccccc} P_{n+1} & \xrightarrow{d_{n+1}} & P_n & \xrightarrow{d_n} & P_{n-1} \\ & & \downarrow{\scriptstyle \check{f}_n} & & \downarrow{\scriptstyle \check{f}_{n-1}} \\ P'_{n+1} & \xrightarrow[d'_{n+1}]{} & P'_n & \xrightarrow[d'_n]{} & P'_{n-1}. \end{array}$$

If we can show that $\operatorname{im} \check{f}_n d_{n+1} \subseteq \operatorname{im} d'_{n+1}$, then we will have the diagram

$$\begin{array}{ccccc} & & P_{n+1} & & \\ & \swarrow & \downarrow{\scriptstyle \check{f}_n d_{n+1}} & & \\ P'_{n+1} & \xrightarrow{d'_{n+1}} & \operatorname{im} d'_{n+1} & \longrightarrow & 0 \end{array}$$

and projectivity of $P_{n+1}$ will provide a map $\check{f}_{n+1} \colon P_{n+1} \to P'_{n+1}$ with $d'_{n+1}\check{f}_{n+1} = \check{f}_n d_{n+1}$. To check that the inclusion does hold, note that exactness at $P'_n$ of the bottom row of the original diagram gives $\operatorname{im} d'_{n+1} = \ker d'_n$, and so it suffices to prove that $d'_n \check{f}_n d_{n+1} = 0$. But $d'_n \check{f}_n d_{n+1} = \check{f}_{n-1} d_n d_{n+1} = 0$.

(ii) We now prove uniqueness of $\check{f}$ to homotopy. If $h \colon \mathbf{P}_\bullet \to \mathbf{P}'_\bullet$ is a chain map also satisfying $\varepsilon' h_0 = f\varepsilon$, then we construct the terms $s_n \colon P_n \to P'_{n+1}$ of a homotopy $s$ by induction on $n \geq -1$; that is, we want

$$h_n - \check{f}_n = d'_{n+1} s_n + s_{n-1} d_n.$$

Let us now begin the induction. First, define $\check{f}_{-1} = f = h_{-1}$. If we define $s_{-1} = 0 = s_{-2}$, then

$$h_{-1} - \check{f}_{-1} = f - f = 0 = d'_0 s_{-1} + s_{-2} d_{-1}$$

for any choice of $d'_0$ and $d_{-1}$; define $d'_0 = \varepsilon'$ and $d_{-1} = 0$.

For the inductive step, it suffices to prove, for all $n \geq -1$, that

$$\operatorname{im}(h_{n+1} - \check{f}_{n+1} - s_n d_{n+1}) \subseteq \operatorname{im} d'_{n+2},$$

for we have a diagram with exact row

$$\begin{array}{ccccc} & & P_{n+1} & & \\ & \swarrow^{d'_{n+2}} & \downarrow{\scriptstyle h_{n+1}-\check{f}_{n+1}-s_n d_{n+1}} & & \\ P'_{n+2} & \longrightarrow & \operatorname{im} d'_{n+2} & \longrightarrow & 0 \end{array}$$

and projectivity of $P_{n+1}$ will give a map $s_{n+1}\colon P_{n+1} \to P'_{n+2}$ satisfying the desired equation. As in part (i) of the proof, exactness of the bottom row of the original diagram gives $\operatorname{im} d'_{n+2} = \ker d'_{n+1}$, and so it suffices to prove

$$d'_{n+1}(h_{n+1} - \check{f}_{n+1} - s_n d_{n+1}) = 0.$$

But

$$\begin{aligned} d'_{n+1}(h_{n+1} - \check{f}_{n+1} - s_n d_{n+1}) &= d'_{n+1}(h_{n+1} - \check{f}_{n+1}) - d'_{n+1} s_n d_{n+1} \\ &= d'_{n+1}(h_{n+1} - \check{f}_{n+1}) - (h_n - \check{f}_n - s_{n-1} d_n) d_{n+1} \\ &= d'_{n+1}(h_{n+1} - \check{f}_{n+1}) - (h_n - \check{f}_n) d_{n+1}, \end{aligned}$$

and the last term is 0 because $h$ and $\check{f}$ are chain maps. •

We introduce a term to describe the chain map $\check{f}$ just constructed.

**Definition.** If $f\colon A \to A'$ is a map of modules and if $\mathbf{P}_A$ and $\mathbf{P}'_{A'}$ are deleted projective resolutions of $A$ and $A'$, respectively, then a chain map $\check{f}\colon \mathbf{P}_A \to \mathbf{P}'_{A'}$

$$\begin{array}{ccccccccccc} \cdots & \longrightarrow & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\varepsilon} & A & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \check{f}_2} & & \downarrow{\scriptstyle \check{f}_1} & & \downarrow{\scriptstyle \check{f}_0} & & \downarrow{\scriptstyle f} & & \\ \cdots & \longrightarrow & P'_2 & \xrightarrow[d'_2]{} & P'_1 & \xrightarrow[d'_1]{} & P'_0 & \xrightarrow[\varepsilon']{} & A' & \longrightarrow & 0 \end{array}$$

is said to be ***over*** $f$ if

$$f\varepsilon = \varepsilon' \check{f}_0.$$

Thus, the Comparison Theorem implies, given a homomorphism $f\colon A \to A'$, that a chain map over $f$ always exists between deleted projective resolutions of $A$ and $A'$; moreover, such a chain map is unique to homotopy.

Given a pair of rings $R$ and $S$ and an additive covariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, we are now going to construct, for all $n \in \mathbb{Z}$, its ***left derived functors*** $L_nT\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$.

The definition will be in two parts: first on objects and then on morphisms.

Choose, once for all, a deleted projective resolution $\mathbf{P}_A$ of every module $A$. As in Example C-3.34(ix), form the complex $T\mathbf{P}_A$, and take homology:

$$L_nT(A) = H_n(T\mathbf{P}_A).$$

This definition is suggested by two examples. First, in algebraic topology, we tensor the complex of a triangulated space $X$ to get homology groups $H_n(X; G)$ of $X$ with ***coefficients*** in an abelian group $G$; or, we apply $\operatorname{Hom}(\ , G)$ to get a complex whose homology groups are called ***cohomology groups*** of $X$ with coefficients in $G$

(of course, this last functor is contravariant). Second, when we considered group extensions, the formulas that arose suggested constructing a free resolution of the trivial module $\mathbb{Z}$ and then applying $\mathrm{Hom}(\ ,K)$ or $-\otimes K$ to this resolution.

We now define $L_nT(f)$, where $f\colon A \to A'$ is a homomorphism. By the Comparison Theorem, there is a chain map $\check{f}\colon \mathbf{P}_A \to \mathbf{P}'_{A'}$ over $f$. It follows that $T\check{f}\colon T\mathbf{P}_A \to T\mathbf{P}'_{A'}$ is also a chain map, and we define $L_nT(f)\colon L_nT(A) \to L_nT(A')$ by

$$L_nT(f) = H_n(T\check{f}) = (T\check{f})_*.$$

In more detail, if $z \in \ker Td_n$, then

$$(L_nT)f\colon z + \operatorname{im} Td_{n+1} \mapsto (T\check{f}_n)z + \operatorname{im} Td'_{n+1}.$$

In pictures, look at the chosen projective resolutions:

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & P_1 & \longrightarrow & P_0 & \longrightarrow & A & \longrightarrow & 0 \\
 & & & & & & \downarrow{\scriptstyle f} & & \\
\cdots & \longrightarrow & P'_1 & \longrightarrow & P'_0 & \longrightarrow & A' & \longrightarrow & 0.
\end{array}$$

Fill in a chain map $\check{f}$ over $f$, delete $A$ and $A'$, apply $T$ to this diagram, and then take the map induced by $T\check{f}$ in homology.

**Example C-3.47.** If $r \in Z(R)$ is a central element in a ring $R$ and if $A$ is a left $R$-module, then $\mu_r\colon A \to A$, defined by $\mu_r\colon a \mapsto ra$, is an $R$-map. We call $\mu_r$ ***multiplication by*** $r$.

**Definition.** A functor $T\colon {}_R\mathbf{Mod} \to {}_R\mathbf{Mod}$, of either variance, ***preserves multiplications*** if $T(\mu_r)\colon TA \to TA$ is multiplication by $r$ for all $r \in Z(R)$.

Tensor product and Hom preserve multiplications. We claim that if $T$ preserves multiplications, then $L_nT$ also preserves multiplications; that is,

$$L_nT(\mu_r) = \text{ multiplication by } r.$$

Given a projective resolution $\cdots \to P_1 \xrightarrow{d_1} P_0 \xrightarrow{\varepsilon} A \to 0$, it is easy to see that $\check{\mu}$ is a chain map over $\mu_r$,

$$\begin{array}{ccccccccccc}
\cdots & \longrightarrow & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\varepsilon} & A & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \check{\mu}_2} & & \downarrow{\scriptstyle \check{\mu}_1} & & \downarrow{\scriptstyle \check{\mu}_0} & & \downarrow{\scriptstyle f} & & \\
\cdots & \longrightarrow & P_2 & \xrightarrow[d_2]{} & P_1 & \xrightarrow[d_1]{} & P_0 & \xrightarrow[\varepsilon]{} & A & \longrightarrow & 0,
\end{array}$$

where $\check{\mu}_n\colon P_n \to P_n$ is multiplication by $r$ for every $n \geq 0$. Since $T$ preserves multiplications, the terms of the chain map $T\check{\mu}$ are multiplication by $r$, and so the induced maps in homology are also multiplication by $r$:

$$(T\check{\mu})_*\colon z_n + \operatorname{im} Td_{n+1} \mapsto (T\check{\mu}_n)z_n + \operatorname{im} Td_{n+1} = rz_n + \operatorname{im} Td_{n+1},$$

where $z_n \in \ker Td_n$. ◀

**Proposition C-3.48.** *Given a pair of rings $R$ and $S$ and an additive covariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, then*

$$L_nT\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$$

*is an additive covariant functor for every $n$.*

**Proof.** We will prove that $L_nT$ is well-defined on morphisms; it is then routine to check that it is a covariant additive functor (remember that $H_n$ is a covariant additive functor from complexes to modules).

If $h\colon \mathbf{P}_A \to \mathbf{P}'_{A'}$ is another chain map over $f$, then the Comparison Theorem says that $h \backsimeq \check{f}$; therefore, $Th \backsimeq T\check{f}$, by Exercise C-3.26 on page 270, and so $H_n(Th) = H_n(T\check{f})$, by Proposition C-3.39. •

**Proposition C-3.49.** *If $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is a covariant additive functor, then $L_nTA = \{0\}$ for all negative $n$ and for all $A$.*

**Proof.** By Exercise C-3.23 on page 269, we have $L_nTA = \{0\}$ because, when $n$ is negative, the $n$th term of $\mathbf{P}_A$ is $\{0\}$. •

**Definition.** If $B$ is a left $R$-module and $T = \ - \otimes_R B$, define

$$\operatorname{Tor}_n^R(\quad, B) = L_nT.$$

Thus, if

$$\mathbf{P}_A = \cdots \to P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \to 0$$

is the chosen deleted projective resolution of a module $A$, then

$$\operatorname{Tor}_n^R(A, B) = H_n(\mathbf{P}_A \otimes_R B) = \frac{\ker(d_n \otimes 1_B)}{\operatorname{im}(d_{n+1} \otimes 1_B)}.$$

The domain of $\operatorname{Tor}_n^R(\quad, B)$ is $\mathbf{Mod}_R$, the category of right $R$-modules; its target is $\mathbf{Ab}$, the category of abelian groups. For example, if $R$ is commutative, then $A \otimes_R B$ is an $R$-module, and so the values of $\operatorname{Tor}_n^R(\ , B)$ lie in ${}_R\mathbf{Mod}$.

**Definition.** If $A$ is a right $R$-module and $T = A \otimes_R -$, define $\operatorname{tor}_n^R(A, \quad) = L_nT$. Thus, if

$$\mathbf{Q}_B = \cdots \to Q_2 \xrightarrow{d_2} Q_1 \xrightarrow{d_1} Q_0 \to 0$$

is the chosen deleted projective resolution of a module $B$, then

$$\operatorname{tor}_n^R(A, B) = H_n(A \otimes_R \mathbf{Q}_B) = \frac{\ker(1_A \otimes d_n)}{\operatorname{im}(1_A \otimes d_{n+1})}.$$

The domain of $\operatorname{tor}_n^R(A, \quad)$ is ${}_R\mathbf{Mod}$, the category of left $R$-modules; its target is $\mathbf{Ab}$, the category of abelian groups, but, as before, its target may be smaller (if $R = \mathbb{Q}$, for example) or larger (if $R = \mathbb{Z}G$, for every $\mathbb{Z}$-module can be viewed as a (trivial) $R$-module).

One of the nice theorems of homological algebra is, for all $A$ and $B$ (and for all $R$ and $n$), that

$$\operatorname{Tor}_n^R(A, B) \cong \operatorname{tor}_n^R(A, B).$$

There is a proof using spectral sequences, but there is also an elementary proof due to Zaks (see Rotman [**187**], p. 197).

There are now several points to discuss. The definition of $L_nT$ assumes that a choice of deleted projection resolution of each module has been made. Does $L_nT$ depend on this choice? And, once we dispose of this question (the answer is that $L_nT$ does not depend on the choice), how can we use these functors?

Assume that new choices $\widetilde{\mathbf{P}}_A$ of deleted projective resolutions have been made, and let us denote the left derived functors arising from these new choices by $\widetilde{L}_nT$.

**Proposition C-3.50.** *Given a pair of rings $R$ and $S$ and an additive covariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, then, for each $n$, the functors $L_nT$ and $\widetilde{L}_nT$ are naturally equivalent. In particular, for all $A$,*

$$(L_nT)A \cong (\widetilde{L}_nT)A,$$

*and so these modules are independent of the choice of (deleted) projective resolution of $A$.*

**Proof.** Consider the diagram

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & P_2 & \longrightarrow & P_1 & \longrightarrow & P_0 & \longrightarrow & A \longrightarrow 0 \\
 & & & & & & & & \downarrow{\scriptstyle 1_A} \\
\cdots & \longrightarrow & \widetilde{P}_2 & \longrightarrow & \widetilde{P}_1 & \longrightarrow & \widetilde{P}_0 & \longrightarrow & A \longrightarrow 0,
\end{array}$$

where the top row is the chosen projective resolution of $A$ used to define $L_nT$ and the bottom row is that used to define $\widetilde{L}_nT$. By the Comparison Theorem, there is a chain map $\iota\colon \mathbf{P}_A \to \widetilde{\mathbf{P}}_A$ over $1_A$. Applying $T$ gives a chain map $T\iota\colon T\mathbf{P}_A \to T\widetilde{\mathbf{P}}_A$ over $T1_A = 1_{TA}$. This last chain map induces homomorphisms, one for each $n$,

$$\tau_A = (T\iota)_*\colon (L_nT)A \to (\widetilde{L}_nT)A.$$

We now prove that each $\tau_A$ is an isomorphism (thereby proving the last statement in the theorem) by constructing its inverse. Turn the preceding diagram upside down, so that the chosen projective resolution $\mathbf{P}_A \to A \to 0$ is now the bottom row. Again, the Comparison Theorem gives a chain map, say, $\kappa\colon \widetilde{\mathbf{P}}_A \to \mathbf{P}_A$. Now the composite $\kappa\iota$ is a chain map from $\mathbf{P}_A$ to itself over $1_{\mathbf{P_A}}$. By the uniqueness statement in the Comparison Theorem, $\kappa\iota \backsimeq 1_{\mathbf{P_A}}$; similarly, $\iota\kappa \backsimeq 1_{\widetilde{\mathbf{P}}_A}$. It follows that $T(\iota\kappa) \backsimeq 1_{T\widetilde{\mathbf{P}}_A}$ and $T(\kappa\iota) \backsimeq 1_{T\mathbf{P}_A}$. Hence, $1 = (T\iota\kappa)_* = (T\iota)_*(T\kappa)_*$ and $1 = (T\kappa\iota)_* = (T\kappa)_*(T\iota)_*$. Therefore, $\tau_A = (T\iota)_*$ is an isomorphism.

We now prove that the isomorphisms $\tau_A$ constitute a natural equivalence: that is, if $f\colon A \to B$ is a homomorphism, then the following diagram commutes:

$$\begin{array}{ccc}
(L_nT)A & \xrightarrow{\tau_A} & (\widetilde{L}_nT)A \\
{\scriptstyle L_nT(f)}\downarrow & & \downarrow{\scriptstyle \widetilde{L}_nT(f)} \\
(L_nT)B & \xrightarrow{\tau_B} & (\widetilde{L}_nT)B.
\end{array}$$

To evaluate in the clockwise direction, consider

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & P_1 & \longrightarrow & P_0 & \longrightarrow & A & \longrightarrow & 0 \\
 & & & & & & \downarrow{\scriptstyle 1_A} & & \\
\cdots & \longrightarrow & \widetilde{P}_1 & \longrightarrow & \widetilde{P}_0 & \longrightarrow & A & \longrightarrow & 0 \\
 & & & & & & \downarrow{\scriptstyle f} & & \\
\cdots & \longrightarrow & \widetilde{Q}_1 & \longrightarrow & \widetilde{Q}_0 & \longrightarrow & B & \longrightarrow & 0,
\end{array}$$

where the bottom row is some projective resolution of $B$. The Comparison Theorem gives a chain map $\mathbf{P}_A \to \widetilde{\mathbf{Q}}_B$ over $f1_A = f$. Going counterclockwise, the picture will now have the chosen projective resolution of $B$ as its middle row, and we get a chain map $\mathbf{P}_A \to \widetilde{\mathbf{Q}}_B$ over $1_B f = f$. The uniqueness statement in the Comparison Theorem tells us that these two chain maps are homotopic, so that they give the same homomorphism in homology. Thus, the appropriate diagram commutes, showing that $\tau\colon L_nT \to \widetilde{L}_nT$ is a natural equivalence. •

**Corollary C-3.51.** *The module* $\operatorname{Tor}_n^R(A,B)$ *is independent of the choices of projective resolutions of* $A$ *and of* $B$.

**Proof.** The proposition applies at once to the left derived functors of $- \otimes_R B$, namely, $\operatorname{Tor}_n^R(\ ,B)$, and to the left derived functors of $A \otimes_R -$, namely $\operatorname{tor}_n^R(A,\quad)$. But we have already cited the fact that $\operatorname{Tor}_n^R(A,B) \cong \operatorname{tor}_n^R(A,B)$. •

**Corollary C-3.52.** *Let* $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ *be an additive covariant functor. If* $P$ *is a projective module, then* $L_nT(P) = \{0\}$ *for all* $n \geq 1$.

*In particular, if* $A$ *and* $P$ *are right* $R$*-modules, with* $P$ *projective, and if* $B$ *and* $Q$ *are left* $R$*-modules, with* $Q$ *projective, then*

$$\operatorname{Tor}_n^R(P,B) = \{0\} \quad \textit{and} \quad \operatorname{Tor}_n^R(A,Q) = \{0\}$$

*for all* $n \geq 1$.

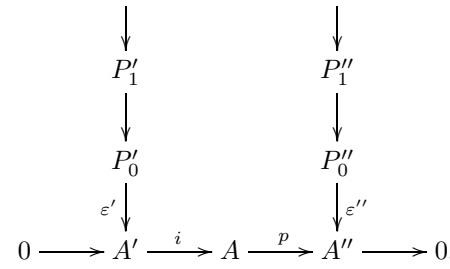
**Proof.** Since $P$ is projective, a projective resolution of it is

$$\mathbf{C}_\bullet = \cdots \to 0 \to 0 \to P \xrightarrow{1_P} P \to 0,$$

and so the corresponding deleted projective resolution $\mathbf{C}_P$ has only one nonzero term, namely, $C_0 = P$. It follows that $T\mathbf{C}_P$ is a complex having $n$th term $\{0\}$ for all $n \geq 1$, and so $L_nTP = H_n(T\mathbf{C}_P) = \{0\}$ for all $n \geq 1$, by Exercise C-3.23 on page 269. •

We are now going to show that there is a long exact sequence of left derived functors. We begin with a useful lemma; it says that if we are given an exact sequence of modules as well as a projective resolution of its first and third terms, then we can "fill in the horseshoe"; that is, there is a projective resolution of the middle term that fits in between them.

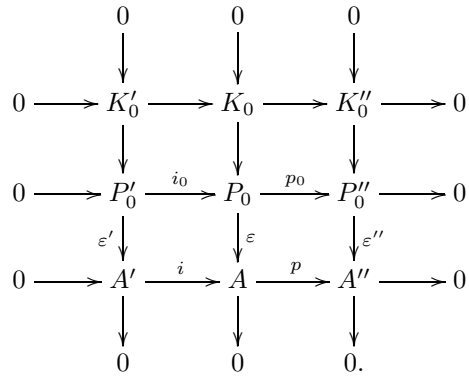
**Lemma C-3.53 (Horseshoe Lemma).** *Given a diagram*



*where the columns are projective resolutions and the row is exact, then there exists a projective resolution of $A$ and chain maps so that the three columns form an exact sequence of complexes.*

**Remark.** The dual theorem, in which projective resolutions are replaced by injective resolutions, is also true. ◀

**Proof.** We show first that there is a projective $P_0$ and a commutative $3\times 3$ diagram with exact columns and rows:



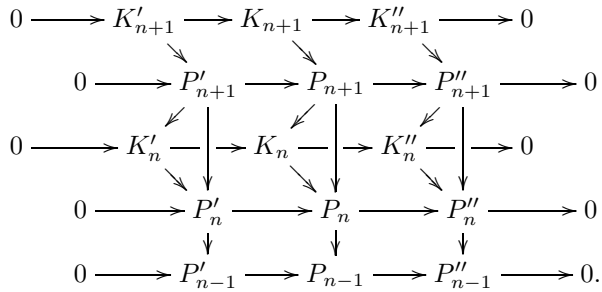
Define $P_0 = P_0' \oplus P_0''$; it is projective because both $P_0'$ and $P_0''$ are projective. Define $i_0 \colon P_0' \to P_0' \oplus P_0''$ by $x' \mapsto (x', 0)$, and define $p_0 \colon P_0' \oplus P_0'' \to P_0''$ by $(x', x'') \mapsto x''$. It is clear that

$$0 \to P_0' \xrightarrow{i_0} P_0 \xrightarrow{p_0} P_0'' \to 0$$

is exact. Since $P_0''$ is projective, there exists a map $\sigma \colon P_0'' \to A$ with $p\sigma = \varepsilon''$. Now define $\varepsilon \colon P_0 \to A$ by $\varepsilon \colon (x', x'') \mapsto i\varepsilon' x' + \sigma x''$. It is left as a routine exercise that if $K_0 = \ker \varepsilon$, then there are maps $K_0' \to K_0$ and $K_0 \to K_0''$ (where $K_0' = \ker \varepsilon'$ and $K_0'' = \ker \varepsilon''$), so that the resulting $3 \times 3$ diagram commutes. Exactness of the top row is Exercise B-1.58 on page 310 in Part 1, and surjectivity of $\varepsilon$ follows from the Five Lemma (Proposition B-1.48 in Part 1).

We now prove, by induction on $n \geq 0$, that the bottom $n$ rows of the desired diagram can be constructed. For the inductive step, assume that the first $n$ steps have been filled in, and let $K_n = \ker(P_n \to P_{n-1})$, etc. Now construct the $3 \times 3$ diagram whose bottom row is $0 \to K_n' \to K_n \to K_n'' \to 0$, and splice it to the

$n$th diagram, as illustrated next (note that the map $P_{n+1} \to P_n$ is defined as the composite $P_{n+1} \to K_n \to P_n$):



The columns of the new diagram are exact because, for example, $\operatorname{im}(P_{n+1} \to P_n) = K_n = \ker(P_n \to P_{n-1})$. •

**Theorem C-3.54.** *If* $0 \to A' \xrightarrow{i} A \xrightarrow{p} A'' \to 0$ *is an exact sequence of modules and if* $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ *is a covariant additive functor, then there is a long exact sequence*

$$\cdots \to L_nTA' \xrightarrow{L_nTi} L_nTA \xrightarrow{L_nTp} L_nTA'' \xrightarrow{\partial_n} L_{n-1}TA' \xrightarrow{L_{n-1}Ti} L_{n-1}TA \xrightarrow{L_{n-1}Tp} L_{n-1}TA'' \xrightarrow{\partial_{n-1}} \cdots$$

*that ends with*

$$\cdots \to L_0TA' \to L_0TA \to L_0TA'' \to 0.$$

**Proof.** Let $\mathbf{P}'_{A'}$ and $\mathbf{P}''_{A''}$ be the chosen deleted projective resolutions of $A'$ and of $A''$, respectively. By Lemma C-3.53, there is a deleted projective resolution $\widetilde{\mathbf{P}}_A$ of $A$ with

$$\mathbf{0}_\bullet \to \mathbf{P}'_{A'} \xrightarrow{j} \widetilde{\mathbf{P}}_A \xrightarrow{q} \mathbf{P}''_{A''} \to \mathbf{0}_\bullet$$

(in the notation of the Comparison Theorem, $j = \breve{i}$ is a chain map over $i$, and $q = \breve{p}$ is a chain map over $p$). Applying $T$ gives the sequence of complexes

$$\mathbf{0}_\bullet \to T\mathbf{P}'_{A'} \xrightarrow{Tj} T\widetilde{\mathbf{P}}_A \xrightarrow{Tq} T\mathbf{P}''_{A''} \to \mathbf{0}_\bullet.$$

To see that this sequence is exact,[13] note that each row $0 \to P'_n \xrightarrow{j_n} \widetilde{P}_n \xrightarrow{q_n} P''_n \to 0$ is a split exact sequence (because $P''_n$ is projective) and additive functors preserve split short exact sequences. There is thus a long exact sequence

$$\cdots \to H_n(T\mathbf{P}'_{A'}) \xrightarrow{(Tj)^*} H_n(T\widetilde{\mathbf{P}}_A) \xrightarrow{(Tq)^*} H_n(T\mathbf{P}''_{A''}) \xrightarrow{\partial_n} H_{n-1}(T\mathbf{P}'_{A'}) \to \cdots;$$

that is, there is an exact sequence

$$\cdots \to L_nTA' \xrightarrow{(Tj)^*} \widetilde{L}_nTA \xrightarrow{(Tq)^*} L_nTA'' \xrightarrow{\partial_n} L_{n-1}TA' \to \cdots.$$

[13] The exact sequence of complexes is *not* split because the sequence of splitting maps need not constitute a chain map $\mathbf{P}''_{A''} \to \widetilde{\mathbf{P}}_A$.

We do not know that the projective resolution of $A$ given by the Horseshoe Lemma is the resolution originally chosen, and this is why we have $\widetilde{L}_nTA$ instead of $L_nTA$. But there is a natural equivalence $\tau\colon L_nT \to \widetilde{L}_nT$, and so there is an exact sequence

$$\cdots \to L_nTA' \xrightarrow{\tau_A^{-1}(Tj)_*} L_nTA \xrightarrow{(Tq)_*\tau_A} L_nTA'' \xrightarrow{\partial_n} L_{n-1}TA' \to \cdots .$$

The sequence does terminate with $\{0\}$, for $L_nT = \{0\}$ for all negative $n$, by Proposition C-3.49.

It remains to show that $\tau_A^{-1}(Tj)_* = L_nT(i) = T(j)_*$ (remember that $j = \check{i}$, a chain map over $i$) and $(Tq)_*\tau_A = L_nT(p)$. Now $\tau_A^{-1} = (T\kappa)_*$, where $\kappa\colon \widetilde{\mathbf{P}}_A \to \mathbf{P}_A$ is a chain map over $1_A$, and so

$$\tau_A^{-1}(Tj)_* = (T\kappa)_*(Tj)_* = (T\kappa Tj)_* = \big(T(\kappa j)\big)_*.$$

Both $\kappa j$ and $j$ are chain maps $\widetilde{\mathbf{P}}_A \to \mathbf{P}_A$ over $1_A$, so they are homotopic, by the Comparison Theorem. Hence, $T(\kappa j)$ and $Tj$ are homotopic, and so they induce the same map in homology: $(T(\kappa j))_* = (Tj)_* = L_nT(i)$. Therefore, $\tau_A^{-1}(Tj)_* = L_nT(i)$. That $(Tq)_*\tau_A = L_nT(p)$ is proved in the same way. •

**Corollary C-3.55.** *If $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is a covariant additive functor, then the functor $L_0T$ is right exact.*

**Proof.** If $A \to B \to C \to 0$ is exact, then $L_0A \to L_0B \to L_0C \to 0$ is exact. •

**Theorem C-3.56.**

(i) *If an additive covariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is right exact, then $T$ is naturally equivalent to $L_0T$.*

(ii) *The functor $-\otimes_R B$ is naturally equivalent to $\operatorname{Tor}_0^R(\ , B)$. Hence, for all right $R$-modules $A$, there is an isomorphism*

$$A \otimes_R B \cong \operatorname{Tor}_0^R(A, B).$$

**Proof.**

(i) Let $\mathbf{P}_A$ be the chosen deleted projective resolution of $A$, and let

$$\cdots \to P_1 \xrightarrow{d_1} P_0 \xrightarrow{\varepsilon} A \to 0$$

be the chosen projective resolution. By definition,

$$L_0TA = \operatorname{coker} T(d_1).$$

But right exactness of $T$ gives an exact sequence

$$TP_1 \xrightarrow{Td_1} TP_0 \xrightarrow{T\varepsilon} TA \to 0.$$

Now $T\varepsilon$ induces an isomorphism $\sigma_A\colon \operatorname{coker} T(d_1) \to TA$, by the First Isomorphism Theorem; that is, $\sigma_A\colon L_0TA \to TA$. It is left as a routine exercise that $\sigma\colon L_0T \to T$ is a natural equivalence.

(ii) Immediate from (i), for $-\otimes_R B$ satisfies the hypotheses. •

We have shown that Tor repairs the loss of exactness that may occur after tensoring a short exact sequence.

**Corollary C-3.57.** *If* $0 \to A' \to A \to A'' \to 0$ *is a short exact sequence of modules, then there is a long exact sequence*

$$\begin{aligned}\cdots \to \operatorname{Tor}_2^R(A', B) &\to \operatorname{Tor}_2^R(A, B) \to \operatorname{Tor}_2^R(A'', B)\\ &\to \operatorname{Tor}_1^R(A', B) \to \operatorname{Tor}_1^R(A, B) \to \operatorname{Tor}_1^R(A'', B)\\ &\to A' \otimes_R B \to A \otimes_R B \to A'' \otimes_R B \to 0.\end{aligned}$$

The next proposition shows that the functors $\operatorname{Tor}_n(\ , B)$ satisfy the covariant version of Theorem C-3.45.

**Proposition C-3.58.** *Given a commutative diagram of modules having exact rows,*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A' & \xrightarrow{i} & A & \xrightarrow{p} & A'' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & C' & \xrightarrow{j} & C & \xrightarrow{q} & C'' & \longrightarrow & 0, \end{array}$$

*there is, for all* $n$*, a commutative diagram with exact rows,*

$$\begin{array}{ccccccc} \operatorname{Tor}_n^R(A', B) & \xrightarrow{i_*} & \operatorname{Tor}_n^R(A, B) & \xrightarrow{p_*} & \operatorname{Tor}_n^R(A'', B) & \xrightarrow{\partial_n} & \operatorname{Tor}_{n-1}^R(A', B) \\ \downarrow{\scriptstyle f_*} & & \downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle h_*} & & \downarrow{\scriptstyle f_*} \\ \operatorname{Tor}_n^R(C', B) & \xrightarrow{j_*} & \operatorname{Tor}_n^R(C, B) & \xrightarrow{q_*} & \operatorname{Tor}_n^R(C'', B) & \xrightarrow{\partial'_n} & \operatorname{Tor}_{n-1}^R(C', B). \end{array}$$

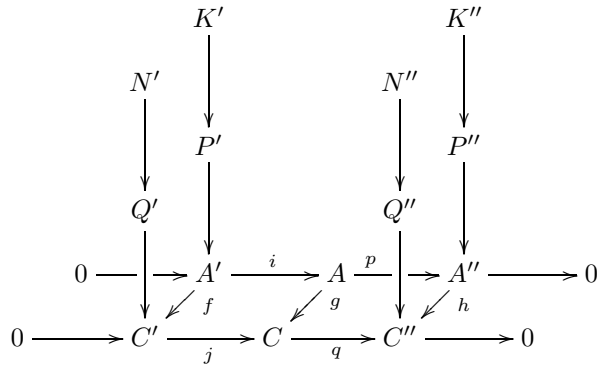
*There is a similar diagram if the first variable is fixed.*

**Proof.** Given the diagram in the statement, erect the chosen deleted projective resolutions on the corners $\mathbf{P}'_{A'}$, $\mathbf{P}''_{A''}$, $\mathbf{Q}'_{C'}$, and $\mathbf{Q}''_{C''}$. We claim that there are deleted projective resolutions $\widetilde{\mathbf{P}}_A$ and $\widetilde{\mathbf{Q}}_C$, together with chain maps, giving a commutative diagram of complexes having exact rows:

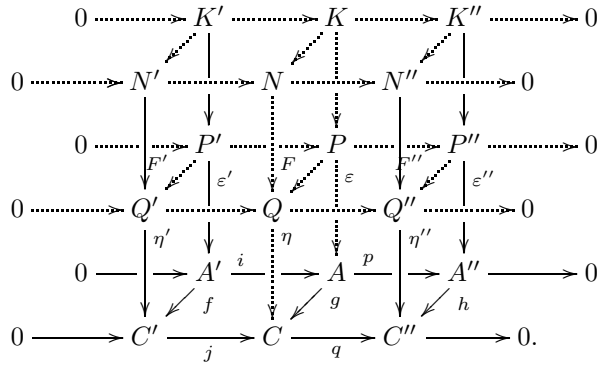
$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathbf{P}'_{A'} & \xrightarrow{\check{i}} & \widetilde{\mathbf{P}}_A & \xrightarrow{\check{p}} & \mathbf{P}''_{A''} & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \check{f}} & & \downarrow{\scriptstyle \check{g}} & & \downarrow{\scriptstyle \check{h}} & & \\ 0 & \longrightarrow & \mathbf{Q}'_{C'} & \xrightarrow{\check{j}} & \widetilde{\mathbf{Q}}_C & \xrightarrow{\check{q}} & \mathbf{Q}''_{C''} & \longrightarrow & 0. \end{array}$$

Once this is done, the result will follow from the naturality of the connecting homomorphism. As in the inductive proof of Lemma C-3.53, it suffices to prove a three-dimensional version of the Horseshoe Lemma. We complete the following commutative diagram, whose columns are short exact sequences and in which $P'$,

$P''$, $Q'$, and $Q''$ are projectives and $N'$, $N''$, $K'$, and $K''$ are kernels,




to the following commutative diagram, whose rows and columns are short exact sequences and in which $P$ and $Q$ are projective:




Step 1. By the Comparison Theorem, there are chain maps $\check{f}\colon \mathbf{P}'_{A'} \to \mathbf{Q}'_{C'}$ over $f$ and $\check{h}\colon \mathbf{P}''_{A''} \to \mathbf{Q}''_{C''}$ over $h$. To simplify notation, we will write $F' = \check{f}_0$ and $F'' = \check{h}_0$.

Step 2. Define $P = P' \oplus P''$, and insert the usual injection and projection maps $P' \to P$ and $P \to P''$, namely, $x' \mapsto (x', 0)$ and $(x', x'') \mapsto x''$. Similarly, define $Q = Q' \oplus Q''$, and insert the injection and projection maps $Q' \to Q$ and $Q \to Q''$. Of course, the sequences $0 \to P' \to P \to P'' \to 0$ and $0 \to Q' \to Q \to Q'' \to 0$ are exact.

Step 3. As in the proof of the Horseshoe Lemma, define $\varepsilon\colon P \to A$ by $\varepsilon\colon (x', x'') \mapsto i\varepsilon' x' + \sigma x''$, where $\sigma\colon P'' \to A$ satisfies $p\sigma = \varepsilon''$ (such a map $\sigma$ was shown to exist in the proof of the Horseshoe Lemma); indeed, the Horseshoe Lemma shows that the rear face of the diagram commutes. Similarly, define $\eta\colon Q \to C$ by $\eta\colon (y', y'') \mapsto j\eta' y' + \tau y''$, where $\tau\colon Q'' \to C$ satisfies $q\tau = \eta''$; the front face commutes as well.

Step 4. Define $F\colon P \to Q$ by

$$F\colon (x', x'') \mapsto (F'x' + \gamma x'', F''x''),$$

where $\gamma\colon P'' \to Q'$ is to be constructed. It is easy to see that the plane containing the $P$'s and $Q$'s commutes, no matter how $\gamma$ is defined.

Step 5. It remains to choose $\gamma$ so that the square with vertices $P, Q, C$, and $A$ commutes; that is, we want $g\varepsilon = \eta F$. Evaluating each side leads to the equation

$$gi\varepsilon' x' + g\sigma x'' = j\eta' F' x' + j\eta'\gamma x'' + \tau F'' x''.$$

Now $gi\varepsilon' = jf\varepsilon' = j\eta' F'$ (because $F'$ is the 0th term in the chain map $\check{f}$ over $f$), and so it suffices to find $\gamma$ so that

$$j\eta'\gamma = g\sigma - \tau F''.$$

Consider the diagram with exact row:

$$\begin{array}{ccccc} & & P'' & & \\ & & \downarrow{\scriptstyle g\sigma - \tau F''} & & \\ Q' & \xrightarrow{j\eta'} & C & \xrightarrow{q} & C''. \end{array}$$

Now $\operatorname{im}(g\sigma - \tau F'') \subseteq \operatorname{im} j\eta' = \ker q$, for

$$qg\sigma - q\tau F'' = hp\sigma - \eta'' F'' = h\varepsilon'' - \eta'' F'' = 0.$$

Since $P''$ is projective, there exists a map $\gamma\colon P'' \to Q'$ making the diagram commute.

Step 6. By the $3 \times 3$ Lemma (Exercise B-1.58 on page 310 in Part 1), the rows $0 \to K' \to K \to K'' \to 0$ and $0 \to N' \to N \to N'' \to 0$ are exact, and we let the reader show that there are maps on the top face making every square commute. •

In the next section, we will show how Tor can be computed and used. But, before leaving this section, let us give the same treatment to Hom that we have just given to tensor product.

*Left* derived functors of a functor $T$ are defined so that $T\mathbf{P}_A$ is a complex with all its nonzero terms on the *left* side; that is, all terms of negative degree are $\{0\}$. One consequence of this is Corollary C-3.55: if $T$ is right exact, then $L_0 T$ is naturally equivalent to $T$. As the Hom functors are left exact, we are now going to define right derived functors $R^n T$, in terms of deleted resolutions $\mathbf{C}_\bullet$ for which $T\mathbf{C}_\bullet$ is on the right. We shall see that $R^0 T$ is naturally equivalent to $T$ when $T$ is left exact.

## C-3.7. Right Derived Functors

Given an additive covariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, where $R$ and $S$ are rings, we are now going to construct, for all $n \in \mathbb{Z}$, its ***right derived functors*** $R^n T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$.

Choose, once for all, a deleted injective resolution $\mathbf{E}^A$ of every module $A$, form the complex $T\mathbf{E}^A$, and take homology:

$$R^nT(A) = H^n(T\mathbf{E}^A) = \frac{\ker Td^n}{\operatorname{im} Td^{n-1}}.$$

The reader should reread Example C-3.34(x) to recall the index raising convention; if the indices are lowered, then the definition would be

$$R^nT(A) = H_{-n}(T\mathbf{E}^A) = \frac{\ker Td_{-n}}{\operatorname{im} Td_{-n+1}}.$$

Notice that we have raised the index on homology modules as well; we write $H^n$ instead of $H_{-n}$.

The definition of $R^nT(f)$, where $f\colon A \to A'$ is a homomorphism, is similar to that for left derived functors. By the dual of the Comparison Theorem, there is a chain map $\check{f}\colon \mathbf{E}^A \to \mathbf{E}'^{A'}$ over $f$, unique to homotopy, and so a unique map $R^nT(f)\colon H^n(T\mathbf{E}^A) \to H^n(T\mathbf{E}^{A'})$, namely, $(T\check{f}_n)_*$, is induced in homology.

In pictures, look at the chosen injective resolutions:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A' & \longrightarrow & E'^0 & \longrightarrow & E'^1 & \longrightarrow & \cdots \\
 & & \uparrow{\scriptstyle f} & & & & & & \\
0 & \longrightarrow & A & \longrightarrow & E^0 & \longrightarrow & E^1 & \longrightarrow & \cdots .
\end{array}$$

Fill in a chain map $\check{f}$ over $f$, then apply $T$ to this diagram, and then take the map induced by $T\check{f}$ in homology.

**Proposition C-3.59.** *Given a pair of rings $R$ and $S$ and an additive covariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, then*

$$R^nT\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$$

*is an additive covariant functor for every $n$.*

The proof of this proposition, as well as the proofs of other propositions about right derived functors soon to be stated, are essentially duals of the proofs we have already given, and so they will be omitted.

**Example C-3.60.** If $T$ is a covariant additive functor that preserves multiplications and if $\mu_r\colon A \to A$ is multiplication by $r$, where $r \in Z(R)$ is a central element, then $R^nT$ also preserves multiplications (see Example C-3.47). ◀

**Proposition C-3.61.** *If $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is a covariant additive functor, then $R^nTA = \{0\}$ for all negative $n$ and for all $A$.*

**Definition.** If $T = \mathrm{Hom}_R(B, \ )$, define $\mathrm{Ext}^n_R(B, \ ) = R^nT$. Thus, if

$$\mathbf{E}^A = 0 \to E^0 \xrightarrow{d^0} E^1 \xrightarrow{d^1} E^2 \to \cdots$$

is the chosen deleted injective resolution of a module $A$, then

$$\mathrm{Ext}^n_R(B, A) = H^n(\mathrm{Hom}_R(B, \mathbf{E}^A)) = \frac{\ker(d^n)_*}{\mathrm{im}(d^{n-1})_*},$$

where $(d^n)_*\colon \mathrm{Hom}_R(B, E^n) \to \mathrm{Hom}_R(B, E^{n+1})$ is defined, as usual, by

$$(d^n)_*\colon f \mapsto d^n f.$$

The domain of $R^nT$, in particular, the domain of $\mathrm{Ext}^n_R(B, \ )$, is ${}_R\mathbf{Mod}$, the category of all left $R$-modules, and its target is **Ab**, the category of abelian groups. The target may be larger; for example, it is ${}_R\mathbf{Mod}$ if $R$ is commutative.

Assume that new choices $\widetilde{\mathbf{E}}^A$ of deleted injective resolutions have been made, and let us denote the right derived functors arising from these new choices by $\widetilde{R}^nT$.

**Proposition C-3.62.** *Given a pair of rings $R$ and $S$ and an additive covariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, then, for each $n$, the functors $R^nT$ and $\widetilde{R}^nT$ are naturally equivalent. In particular, for all $A$,*

$$(R^nT)A \cong (\widetilde{R}^nT)A,$$

*and so these modules are independent of the choice of (deleted) injective resolution of $A$.*

**Corollary C-3.63.** *The module $\mathrm{Ext}^n_R(B, A)$ is independent of the choice of injective resolution of $A$.*

**Corollary C-3.64.** *Let $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ be an additive covariant functor. If $E$ is an injective module, then $R^nT(E) = \{0\}$ for all $n \geq 1$.*

*In particular, if $E$ is an injective $R$-module, then $\mathrm{Ext}^n_R(B, E) = \{0\}$ for all $n \geq 1$ and all modules $B$.*

**Theorem C-3.65.** *If $0 \to A' \xrightarrow{i} A \xrightarrow{p} A'' \to 0$ is an exact sequence of modules and if $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is a covariant additive functor, then there is a long exact sequence*

$$\cdots \to R^nTA' \xrightarrow{R^nTi} R^nTA \xrightarrow{R^nTp} R^nTA'' \xrightarrow{\partial^n}$$
$$R^{n+1}TA' \xrightarrow{R^{n+1}Ti} R^{n+1}TA \xrightarrow{R^{n+1}Tp} R^{n+1}TA'' \xrightarrow{\partial^{n+1}} \cdots$$

*that begins with*

$$0 \to R^0TA' \to R^0TA \to R^0TA'' \to \cdots.$$

**Corollary C-3.66.** *If $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is a covariant additive functor, then the functor $R^0T$ is left exact.*

**Theorem C-3.67.**

(i) *If an additive covariant functor* $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ *is left exact, then* $T$ *is naturally equivalent to* $R^0T$.

(ii) *If* $B$ *is a left* $R$*-module, the functor* $\operatorname{Hom}_R(B,\ )$ *is naturally equivalent to* $\operatorname{Ext}^0_R(B,\ )$. *Hence, for all left* $R$*-modules* $A$, *there is an isomorphism*

$$\operatorname{Hom}_R(B,A) \cong \operatorname{Ext}^0_R(B,A).$$

We have shown that Ext repairs the loss of exactness that may occur after applying Hom to a short exact sequence.

**Corollary C-3.68.** *If* $0 \to A' \to A \to A'' \to 0$ *is a short exact sequence of modules, then there is a long exact sequence*

$$\begin{aligned} 0 \to \operatorname{Hom}_R(B,A') &\to \operatorname{Hom}_R(B,A) \to \operatorname{Hom}_R(B,A'') \\ &\to \operatorname{Ext}^1_R(B,A') \to \operatorname{Ext}^1_R(B,A) \to \operatorname{Ext}^1_R(B,A'') \\ &\to \operatorname{Ext}^2_R(B,A') \to \operatorname{Ext}^2_R(B,A) \to \operatorname{Ext}^2_R(B,A'') \to \cdots . \end{aligned}$$

**Proposition C-3.69.** *Given a commutative diagram of modules having exact rows,*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A' & \xrightarrow{i} & A & \xrightarrow{p} & A'' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & C' & \xrightarrow{j} & C & \xrightarrow{q} & C'' & \longrightarrow & 0, \end{array}$$

*there is, for all* $n$, *a commutative diagram with exact rows,*

$$\begin{array}{ccccccccc} \operatorname{Ext}^n_R(B,A') & \xrightarrow{i_*} & \operatorname{Ext}^n_R(B,A) & \xrightarrow{p_*} & \operatorname{Ext}^n_R(B,A'') & \xrightarrow{\partial^n} & \operatorname{Ext}^{n+1}_R(B,A') \\ \downarrow{\scriptstyle f_*} & & \downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle h_*} & & \downarrow{\scriptstyle f_*} \\ \operatorname{Ext}^n_R(B,C') & \xrightarrow{j_*} & \operatorname{Ext}^n_R(B,C) & \xrightarrow{q_*} & \operatorname{Ext}^n_R(B,C'') & \xrightarrow{\partial'^n} & \operatorname{Ext}^{n+1}_R(B,C'). \end{array}$$

Finally, we discuss derived functors of contravariant functors $T$. If we define right derived functors $R^nT$, in terms of deleted resolutions $\mathbf{C}_\bullet$ for which $T\mathbf{C}_\bullet$ is on the right, then we start with a deleted projective resolution $\mathbf{P}_A$, for then the contravariance of $T$ puts $T\mathbf{P}_A$ on the right.[14]

Given an additive contravariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, where $R$ and $S$ are rings, we are now going to construct, for all $n \in \mathbb{Z}$, its ***right derived functors*** $R^nT\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$.

Choose, once for all, a deleted projective resolution $\mathbf{P}_A$ of every module $A$, form the complex $T\mathbf{P}_A$, and take homology:

$$R^nT(A) = H^n(T\mathbf{P}_A) = \frac{\ker Td_{n+1}}{\operatorname{im} Td_n}.$$

[14] If we were interested in left derived functors of a contravariant $T$, but we are not, then we would use injective resolutions.

If $f\colon A \to A'$, define $R^nT(f)\colon R^nT(A') \to R^nT(A)$ as we did for left derived functors. By the Comparison Theorem, there is a chain map $\check{f}\colon \mathbf{P}_A \to \mathbf{P}'_{A'}$ over $f$, unique to homotopy, which induces a map $R^nT(f)\colon H^n(T\mathbf{P}'_{A'}) \to H^n(T\mathbf{P}_A)$, namely, $(T\check{f}_n)_*$, in homology.

**Example C-3.70.** If $T$ is an additive contravariant functor that preserves multiplications and if $\mu_r\colon A \to A$ is multiplication by $r$, where $r \in Z(R)$ is a central element, then $R^nT$ also preserves multiplications (see Example C-3.47). ◀

**Proposition C-3.71.** *Given a pair of rings $R$ and $S$ and an additive contravariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, then*

$$R^nT\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$$

*is an additive contravariant functor for every $n$.*

**Proposition C-3.72.** *If $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is a contravariant additive functor, then $R^nTA = \{0\}$ for all negative $n$ and for all $A$.*

**Definition.** If $T = \mathrm{Hom}_R(\ ,C)$, define $\mathrm{ext}^n_R(\ ,C) = R^nT$. Thus, if

$$\cdots \to P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \to 0$$

is the chosen deleted projective resolution of a module $A$, then

$$\mathrm{ext}^n_R(A,C) = H^n(\mathrm{Hom}_R(\mathbf{P}_A, C)) = \frac{\ker(d_{n+1})^*}{\mathrm{im}(d_n)^*},$$

where $(d^n)^*\colon \mathrm{Hom}_R(P_{n-1}, C) \to \mathrm{Hom}_R(P_n, C)$ is defined, as usual, by

$$(d_n)^*\colon f \mapsto fd_n.$$

The same phenomenon that holds for Tor holds for Ext: for all $A$ and $C$ (and for all $R$ and $n$),

$$\mathrm{Ext}^n_R(A,C) \cong \mathrm{ext}^n_R(A,C).$$

The same proof that shows that Tor is independent of the variable resolved also works for Ext (see Rotman [**187**], p. 197). In light of this theorem, we will dispense with the two notations for Ext.

Assume that new choices $\widetilde{\mathbf{P}}_A$ of deleted projective resolutions have been made, and let us denote the right derived functors arising from these new choices by $\widetilde{R}^nT$.

**Proposition C-3.73.** *Given a pair of rings $R$ and $S$ and an additive contravariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$, then, for each $n$, the functors $R^nT$ and $\widetilde{R}^nT$ are naturally equivalent. In particular, for all $A$,*

$$(R^nT)A \cong (\widetilde{R}^nT)A,$$

*and so these modules are independent of the choice of (deleted) projective resolution of $A$.*

**Corollary C-3.74.** *The module $\mathrm{Ext}^n_R(A,C)$ is independent of the choice of projective resolution of $A$.*

**Corollary C-3.75.** *Let $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ be an additive contravariant functor. If $P$ is a projective module, then $R^nT(P) = \{0\}$ for all $n \geq 1$.*

*In particular, if $P$ is a projective $R$-module, then $\operatorname{Ext}^n_R(P, B) = \{0\}$ for all $n \geq 1$ and all modules $B$.*

**Theorem C-3.76.** *If $0 \to A' \xrightarrow{i} A \xrightarrow{p} A'' \to 0$ is an exact sequence of modules and if $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is a contravariant additive functor, then there is a long exact sequence*

$$\cdots \to R^nTA'' \xrightarrow{R^nTp} R^nTA \xrightarrow{R^nTi} R^nTA' \xrightarrow{\partial^n} R^{n+1}TA'' \xrightarrow{R^{n+1}Tp} R^{n+1}TA \xrightarrow{R^{n+1}Ti} R^{n+1}TA' \xrightarrow{\partial^{n+1}} \cdots$$

*that begins with*

$$0 \to R^0TA'' \to R^0TA \to R^0TA' \to \cdots.$$

**Corollary C-3.77.** *If $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is a contravariant additive functor, then the functor $R^0T$ is left exact.*

**Theorem C-3.78.**

(i) *If an additive contravariant functor $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ is left exact, then $T$ is naturally equivalent to $R^0T$.*

(ii) *If $C$ is a left $R$-module, the functor $\operatorname{Hom}_R(\ , C)$ is naturally equivalent to $\operatorname{Ext}^0_R(\ , C)$. Hence, for all left $R$-modules $A$, there is an isomorphism*

$$\operatorname{Hom}_R(A, C) \cong \operatorname{Ext}^0_R(A, C).$$

We have shown that Ext repairs the loss of exactness that may occur after applying Hom to a short exact sequence.

**Corollary C-3.79.** *If $0 \to A' \to A \to A'' \to 0$ is a short exact sequence of modules, then there is a long exact sequence*

$$\begin{aligned} 0 \to \operatorname{Hom}_R(A'', C) &\to \operatorname{Hom}_R(A, C) \to \operatorname{Hom}_R(A', C) \\ &\to \operatorname{Ext}^1_R(A'', C) \to \operatorname{Ext}^1_R(A, C) \to \operatorname{Ext}^1_R(A', C) \\ &\to \operatorname{Ext}^2_R(A'', C) \to \operatorname{Ext}^2_R(A, C) \to \operatorname{Ext}^2_R(A', C) \to \cdots. \end{aligned}$$

**Proposition C-3.80.** *Given a commutative diagram of modules having exact rows,*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A' & \xrightarrow{i} & A & \xrightarrow{p} & A'' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & C' & \xrightarrow{j} & C & \xrightarrow{q} & C'' & \longrightarrow & 0, \end{array}$$

*there is, for all $n$, a commutative diagram with exact rows,*

$$\begin{array}{ccccccc} \operatorname{Ext}^n_R(A'', B) & \xrightarrow{p^*} & \operatorname{Ext}^n_R(A, B) & \xrightarrow{i^*} & \operatorname{Ext}^n_R(A', B) & \xrightarrow{\partial^n} & \operatorname{Ext}^{n+1}_R(A'', B) \\ \uparrow{\scriptstyle h^*} & & \uparrow{\scriptstyle g^*} & & \uparrow{\scriptstyle f^*} & & \uparrow{\scriptstyle h^*} \\ \operatorname{Ext}^n_R(C'', B) & \xrightarrow{q^*} & \operatorname{Ext}^n_R(C, B) & \xrightarrow{j^*} & \operatorname{Ext}^n_R(C', B) & \xrightarrow{\partial^{n\prime}} & \operatorname{Ext}^{n+1}_R(C'', B). \end{array}$$

**Remark.** When $T$ is a covariant functor, then we call the ingredients of $L_nT$ chains, cycles, boundaries, and homology. When $T$ is contravariant, we often add the prefix "co", and the ingredients of $R^nT$ are usually called *cochains*, *cocycles*, *coboundaries*, and *cohomology*. Unfortunately, this clear distinction is blurred because the Hom functor is contravariant in one variable but covariant in the other. In spite of this, we usually use the "co" prefix for the derived functors $\mathrm{Ext}^n$ of Hom. ◀

Derived functors are one way to construct functors like Ext and Tor. In the next section, along with more properties of Ext and Tor, we shall describe another construction of Ext, due to Yoneda, and another construction of Tor, due to Mac Lane. Indeed, derived functors will rarely be mentioned in the rest of the book.

## Exercises

**C-3.34.** If $\tau\colon F \to G$ is a natural transformation between additive functors, prove that $\tau$ gives chain maps $\tau_{\mathbf{C}}\colon F\mathbf{C} \to G\mathbf{C}$ for every complex $\mathbf{C}$. If $\tau$ is a natural equivalence, prove that $F\mathbf{C} \cong G\mathbf{C}$.

**C-3.35.** (i) Let $T\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ be an exact additive functor, where $R$ and $S$ are rings, and suppose that $P$ projective implies $TP$ projective. If $B$ is a left $R$-module and $\mathbf{P}_B$ is a deleted projective resolution of $B$, prove that $T\mathbf{P}_{TB}$ is a deleted projective resolution of $TB$.

(ii) Let $A$ be an $R$-algebra, where $R$ is a commutative ring, which is flat as an $R$-module. Prove that if $B$ is an $A$-module (and hence an $R$-module), then

$$A \otimes_R \mathrm{Tor}_n^R(B, C) \cong \mathrm{Tor}_n^A(B, A \otimes_R C)$$

for all $R$-modules $C$ and all $n \geq 0$.

**C-3.36.** Let $R$ be a semisimple ring.

(i) Prove, for all $n \geq 1$, that $\mathrm{Tor}_n^R(A, B) = \{0\}$ for all right $R$-modules $A$ and all left $R$-modules $B$.
**Hint.** If $R$ is semisimple, then every (left or right) $R$-module is projective.

(ii) Prove, for all $n \geq 1$, that $\mathrm{Ext}_R^n(A, B) = \{0\}$ for all left $R$-modules $A$ and $B$.

∗ **C-3.37.** If $R$ is a PID, prove, for all $n \geq 2$, that $\mathrm{Tor}_n^R(A, B) = \{0\} = \mathrm{Ext}_R^n(A, B)$ for all $R$-modules $A$ and $B$.

**Hint.** Use Theorem B-2.28 in Part 1.

**C-3.38.** Let $R$ be a domain and let $A$ be an $R$-module.

(i) Prove that if the multiplication $\mu_r\colon A \to A$ is an injection for all $r \neq 0$, then $A$ is torsion-free.

(ii) Prove that if the multiplication $\mu_r\colon A \to A$ is a surjection for all $r \neq 0$, then $A$ is divisible.

(iii) Prove that if the multiplication $\mu_r\colon A \to A$ is an isomorphism for all $r \neq 0$, then $A$ is a vector space over $Q$, where $Q = \mathrm{Frac}(R)$.
**Hint.** A module $A$ is a vector space over $Q$ if and only if it is torsion-free and divisible.

(iv) If either $C$ or $A$ is a vector space over $Q$, prove that $\operatorname{Tor}_n^R(C,A)$ and $\operatorname{Ext}_R^n(C,A)$ are also vector spaces over $Q$.

∗ **C-3.39.** Let $R$ be a domain and let $Q = \operatorname{Frac}(R)$.

(i) If $r \in R$ is nonzero and $A$ is an $R$-module for which $rA = \{0\}$, that is, $ra = 0$ for all $a \in A$, prove that $\operatorname{Ext}_R^n(Q,A) = \{0\} = \operatorname{Tor}_n^R(Q,A)$ for all $n \geq 0$.
**Hint.** If $V$ is a vector space over $Q$ for which $rV = \{0\}$, then $V = \{0\}$.

(ii) Prove that $\operatorname{Ext}_R^n(V,A) = \{0\} = \operatorname{Tor}_n^R(V,A)$ for all $n \geq 0$ whenever $V$ is a vector space over $Q$ and $A$ is an $R$-module for which $rA = \{0\}$ for some nonzero $r \in R$.

**C-3.40.** Let $A$ and $B$ be $R$-modules. For $f\colon A' \to B$, where $A'$ is a submodule of $A$, define its ***obstruction*** to be $\partial(f)$, where $\partial\colon \operatorname{Hom}_R(A',B) \to \operatorname{Ext}_R^1(A/A',B)$ is the connecting homomorphism. Prove that $f$ can be extended to a homomorphism $\widetilde{f}\colon A \to B$ if and only if its obstruction is 0.

**C-3.41.** If $T\colon \mathbf{Ab} \to \mathbf{Ab}$ is a left exact functor, prove that $L_0T$ is an exact functor. Conclude, for any abelian group $B$, that $L_0 \operatorname{Hom}(B, \ )$ is not naturally equivalent to $\operatorname{Hom}(B, \ )$.

∗ **C-3.42.** Let

$$\begin{array}{ccc} D & \xrightarrow{\alpha} & C \\ {\scriptstyle\beta}\downarrow & & \downarrow{\scriptstyle g} \\ B & \xrightarrow[f]{} & A \end{array}$$

be a pullback diagram in **Ab**. If there are $c \in C$ and $b \in B$ with $gc = fb$, prove that there exists $d \in D$ with $c\alpha(d)$ and $b = \beta(d)$,

**Hint.** Define $p\colon \mathbb{Z} \to C$ by $p(n) = nc$, and define $q\colon \mathbb{Z} \to B$ by $q(n) = nb$. There is a map $\theta\colon \mathbb{Z} \to D$ making the diagram commute; define $d = \theta(1)$.

## C-3.8. Ext and Tor

We now examine Ext and Tor more closely. As we said in the last section, all properties of these functors should follow from versions of Theorem C-3.45, the axioms characterizing them (see Exercises C-3.43 and C-3.44 on page 308); in particular, their construction as derived functors need not be used.

We begin by showing that Ext behaves like Hom with respect to sums and products.

**Proposition C-3.81.** *If $\{A_k : k \in K\}$ is a family of modules, then there are natural isomorphisms, for all $n$,*

$$\operatorname{Ext}_R^n\Big(\sum_{k\in K} A_k, B\Big) \cong \prod_{k\in K} \operatorname{Ext}_R^n(A_k,B).$$

**Proof.** The proof is by dimension shifting, that is, by induction on $n \geq 0$. The base step is Theorem C-3.67, for $\operatorname{Ext}^0(\ ,B)$ is naturally equivalent to the contravariant functor $\operatorname{Hom}(\ ,B)$.

For the inductive step, choose, for each $k \in K$, a short exact sequence

$$0 \to L_k \to P_k \to A_k \to 0,$$

where $P_k$ is projective. There is an exact sequence

$$0 \to \sum_k L_k \to \sum_k P_k \to \sum_k A_k \to 0,$$

and $\sum_k P_k$ is projective, for every sum of projectives is projective. There is a commutative diagram with exact rows:

$$\begin{array}{ccccccc}
\operatorname{Hom}(\sum P_k, B) & \to & \operatorname{Hom}(\sum L_k, B) & \xrightarrow{\partial} & \operatorname{Ext}^1(\sum A_k, B) & \to & \operatorname{Ext}^1(\sum P_k, B) \\
\downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle \sigma} & & \vdots & & \\
\prod \operatorname{Hom}(P_k, B) & \to & \prod \operatorname{Hom}(L_k, B) & \xrightarrow{d} & \prod \operatorname{Ext}^1(A_k, B) & \to & \prod \operatorname{Ext}^1(P_k, B),
\end{array}$$

where the maps in the bottom row are just the usual induced maps in each coordinate and the maps $\tau$ and $\sigma$ are the isomorphisms given by Proposition C-3.69. Now $\operatorname{Ext}^1(\sum P_k, B) = \{0\} = \prod \operatorname{Ext}^1(P_k, B)$, because $\sum P_k$ and each $P_k$ are projective, so that the maps $\partial$ and $d$ are surjective. This is precisely the sort of diagram in Proposition B-1.46 in Part 1, and so there exists an isomorphism $\operatorname{Ext}^1(\sum A_k, B) \to \prod \operatorname{Ext}^1(A_k, B)$ making the augmented diagram commute.

We may now assume that $n \geq 1$, and we look further out in the long exact sequence. There is a commutative diagram

$$\begin{array}{ccccccc}
\operatorname{Ext}^n\big(\sum P_k, B\big) & \to & \operatorname{Ext}^n\big(\sum L_k, B\big) & \xrightarrow{\partial} & \operatorname{Ext}^{n+1}\big(\sum A_k, B\big) & \to & \operatorname{Ext}^{n+1}\big(\sum P_k, B\big) \\
 & & \downarrow{\scriptstyle \sigma} & & \vdots & & \\
\prod \operatorname{Ext}^n(P_k, B) & \to & \prod \operatorname{Ext}^n(L_k, B) & \xrightarrow{d} & \prod \operatorname{Ext}^{n+1}(A_k, B) & \to & \prod \operatorname{Ext}^{n+1}(P_k, B),
\end{array}$$

where $\sigma\colon \operatorname{Ext}^n(\sum L_k, B) \to \prod \operatorname{Ext}^n(L_k, B)$ is an isomorphism that exists by the inductive hypothesis. Since $n \geq 1$, all four Ext's whose first variable is projective are $\{0\}$; it follows from exactness of the rows that both $\partial$ and $d$ are isomorphisms. Finally, the composite $d\sigma\partial^{-1}\colon \operatorname{Ext}^{n+1}(\sum A_k, B) \to \prod \operatorname{Ext}^{n+1}(A_k, B)$ is an isomorphism, as desired. •

There is a dual result in the second variable.

**Proposition C-3.82.** *If $\{B_k : k \in K\}$ is a family of modules, then there are natural isomorphisms, for all $n$,*

$$\operatorname{Ext}^n_R\Big(A, \prod_{k\in K} B_k\Big) \cong \prod_{k\in K} \operatorname{Ext}^n_R(A, B_k).$$

**Proof.** The proof is by dimension shifting. The base step is Theorem B-4.8 in Part 1, for $\operatorname{Ext}^0(A, \ )$ is naturally equivalent to the covariant functor $\operatorname{Hom}(A, \ )$.

For the inductive step, choose, for each $k \in K$, a short exact sequence

$$0 \to B_k \to E_k \to N_k \to 0,$$

where $E_k$ is injective. There is an exact sequence $0 \to \prod_k B_k \to \prod_k E_k \to \prod_k N_k \to 0$, and $\prod_k E_k$ is injective, for every product of injectives is injective, by Proposition B-4.55 in Part 1. There is a commutative diagram with exact rows:

$$\begin{array}{ccccccc}
\operatorname{Hom}(A, \prod E_k) & \to & \operatorname{Hom}(A, \prod N_k) & \xrightarrow{\partial} & \operatorname{Ext}^1(A, \prod B_k) & \to & \operatorname{Ext}^1(A, \prod E_k) \\
\downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle \sigma} & & \| & & \\
\prod \operatorname{Hom}(A, E_k) & \to & \prod \operatorname{Hom}(A, N_k) & \xrightarrow{d} & \prod \operatorname{Ext}^1(A, B_k) & \to & \prod \operatorname{Ext}^1(A, E_k),
\end{array}$$

where the maps in the bottom row are just the usual induced maps in each coordinate and the maps $\tau$ and $\sigma$ are the isomorphisms given by Theorem B-4.8 in Part 1. The proof now finishes as that of Proposition C-3.81. •

It follows that $\operatorname{Ext}^n$ commutes with finite direct sums in either variable (actually, Proposition B-4.18 in Part 1 says that every additive functor commutes with finite direct sums).

**Remark.** These last two propositions cannot be generalized by replacing sums by direct limits or products by inverse limits; the reason is, in general, that direct limits of projectives need not be projective and inverse limits of injectives need not be injective. ◀

When the ring $R$ is noncommutative, $\operatorname{Hom}_R(A, B)$ is an abelian group, but it need not be an $R$-module.

**Proposition C-3.83.**

(i) $\operatorname{Ext}^n_R(A, B)$ *is a* $Z(R)$*-module. In particular, if* $R$ *is a commutative ring, then* $\operatorname{Ext}^n_R(A, B)$ *is an* $R$*-module.*

(ii) *If* $A$ *and* $B$ *are left* $R$*-modules and* $r \in Z(R)$ *is a central element, then the induced map* $\mu_r^*\colon \operatorname{Ext}^n_R(A, B) \to \operatorname{Ext}^n_R(A, B)$*, where* $\mu_r\colon B \to B$ *is multiplication by* $r$*, is also multiplication by* $r$*. A similar statement is true in the other variable.*

**Proof.**

(i) By Example C-3.47, $\mu_r$ is an $R$-map, and so it induces a homomorphism on $\operatorname{Ext}^n_R(A, B)$. It is straightforward to check that $x \mapsto \mu_r^*(x)$ defines a scalar multiplication $Z(R) \times \operatorname{Ext}^n_R(A, B) \to \operatorname{Ext}^n_R(A, B)$.

(ii) This follows from (i) if we define scalar multiplication by $r$ to be $\mu_r^*$. •

**Example C-3.84.**

(i) We show, for every abelian group $B$, that

$$\operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_n, B) \cong B/nB.$$

There is an exact sequence

$$0 \to \mathbb{Z} \xrightarrow{\mu_n} \mathbb{Z} \to \mathbb{Z}_n \to 0,$$

where $\mu_n$ is multiplication by $n$. Applying $\mathrm{Hom}(\ ,B)$ gives exactness of $\mathrm{Hom}(\mathbb{Z},B) \xrightarrow{\mu_n^*} \mathrm{Hom}(\mathbb{Z},B) \to \mathrm{Ext}^1(\mathbb{Z}_n,B) \to \mathrm{Ext}^1(\mathbb{Z},B)$. Now $\mathrm{Ext}^1(\mathbb{Z},B) = \{0\}$ because $\mathbb{Z}$ is projective. Moreover, $\mu_n^*$ is also multiplication by $n$, while $\mathrm{Hom}(\mathbb{Z},B) = B$. More precisely, $\mathrm{Hom}(\mathbb{Z},\ )$ is naturally equivalent to the identity functor on **Ab**, and so there is a commutative diagram with exact rows

$$\begin{array}{ccccccc} B & \xrightarrow{\mu_n} & B & \longrightarrow & B/nB & \longrightarrow & 0 \\ \downarrow{\scriptstyle \tau_B} & & \downarrow{\scriptstyle \tau_B} & & \vdots & & \\ \mathrm{Hom}(\mathbb{Z},B) & \xrightarrow{\mu_n^*} & \mathrm{Hom}(\mathbb{Z},B) & \longrightarrow & \mathrm{Ext}^1(\mathbb{Z}_n,B) & \longrightarrow & 0. \end{array}$$

By Proposition B-1.46 in Part 1, there is an isomorphism $B/nB \cong \mathrm{Ext}^1(\mathbb{Z}_n,B)$.

(ii) We can now compute $\mathrm{Ext}^1_{\mathbb{Z}}(A,B)$ whenever $A$ and $B$ are finitely generated abelian groups. By the Fundamental Theorem of Finite Abelian Groups, both $A$ and $B$ are direct sums of cyclic groups. Since Ext commutes with finite direct sums, $\mathrm{Ext}^1_{\mathbb{Z}}(A,B)$ is the direct sum of groups $\mathrm{Ext}^1_{\mathbb{Z}}(C,D)$, where $C$ and $D$ are cyclic. We may assume that $C$ is finite, otherwise, it is projective, and $\mathrm{Ext}^1(C,D) = \{0\}$. This calculation can be completed using part (i) and Exercise B-3.15 on page 377 in Part 1, which says that if $D$ is a cyclic group of finite order $m$, then $D/nD$ is a cyclic group of order $d$, where $d = \gcd(m,n)$. ◀

We now define extensions of modules in the obvious way.

**Definition.** Given $R$-modules $C$ and $A$, an ***extension*** of $A$ by $C$ is a short exact sequence

$$0 \to A \xrightarrow{i} B \xrightarrow{p} C \to 0.$$

An extension is ***split*** if there exists an $R$-map $s\colon C \to B$ with $ps = 1_C$.

Of course, if $0 \to A \to B \to C \to 0$ is a split extension, then $B \cong A \oplus C$.

Whenever meeting a homology group, we must ask what it means for it to be zero, for its elements can then be construed as being obstructions. For example, factor sets explain why a group extension may not be split. In this section, we will show that $\mathrm{Ext}^1_R(C,A) = \{0\}$ if and only if every extension of $A$ by $C$ splits. Thus, nonzero elements of any $\mathrm{Ext}^1_R(C',A')$ describe nonsplit extensions (indeed, this result is why Ext is so called).

We begin with a definition motivated by Proposition C-3.17.

**Definition.** Given modules $C$ and $A$, two extensions $\xi : 0 \to A \to B \to C \to 0$ and $\xi' : 0 \to A \to B' \to C \to 0$ of $A$ by $C$ are ***equivalent*** if there exists a map $\varphi\colon B \to B'$ making the following diagram commute:

$$\begin{array}{ccccccccc} \xi : 0 & \longrightarrow & A & \xrightarrow{i} & B & \xrightarrow{p} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle 1_A} & & \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle 1_C} & & \\ \xi' : 0 & \longrightarrow & A & \xrightarrow{i'} & B' & \xrightarrow{p'} & C & \longrightarrow & 0. \end{array}$$

We denote the equivalence class of an extension $\xi$ by $[\xi]$, and we define

$$e(C, A) = \big\{[\xi] : \xi \text{ is an extension of } A \text{ by } C\big\}.$$

If two extensions are equivalent, then the Five Lemma shows that the map $\varphi$ must be an isomorphism; it follows that equivalence is, indeed, an equivalence relation (for we can now prove symmetry). However, the converse is false: there can be inequivalent extensions having isomorphic middle terms, as we saw in Example C-3.18 (all groups in this example are abelian, and so we may view it as an example of $\mathbb{Z}$-modules).

**Proposition C-3.85.** *If* $\operatorname{Ext}^1_R(C, A) = \{0\}$*, then every extension*

$$0 \to A \xrightarrow{i} B \xrightarrow{p} C \to 0$$

*is split.*

**Proof.** Apply the functor $\operatorname{Hom}(C, \ )$ to the extension to obtain an exact sequence

$$\operatorname{Hom}(C, B) \xrightarrow{p_*} \operatorname{Hom}(C, C) \xrightarrow{\partial} \operatorname{Ext}^1(C, A).$$

By hypothesis, $\operatorname{Ext}^1(C, A) = \{0\}$, so that $p_*$ is surjective. Hence, there exists $s \in \operatorname{Hom}(C, B)$ with $1_C = p_*(s)$; that is, $1_C = ps$, and this says that the extension splits. •

**Corollary C-3.86.** *An $R$-module $P$ is projective if and only if* $\operatorname{Ext}^1_R(P, B) = \{0\}$ *for every $R$-module $B$.*

**Proof.** If $P$ is projective, then $\operatorname{Ext}^1_R(P, B) = \{0\}$ for all $B$, by Corollary C-3.75. Conversely, if $\operatorname{Ext}^1_R(P, B) = \{0\}$ for all $B$, then every exact sequence $0 \to B \to X \to P \to 0$ splits, by Proposition C-3.85, and so $P$ is projective, by Proposition B-4.41 in Part 1. •

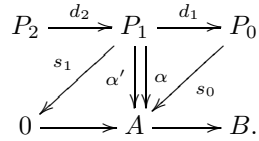
We are going to prove the converse of Proposition C-3.85 by showing that there is a bijection $\psi\colon e(C, A) \to \operatorname{Ext}^1(C, A)$. Let us construct the function $\psi$.

Given an extension $\xi : 0 \to A \to B \to C \to 0$ and a projective resolution of $C$, form the diagram

$$\begin{array}{ccccccccc}
P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \longrightarrow & C & \longrightarrow & 0 \\
\downarrow & & \downarrow{\scriptstyle \alpha} & & \downarrow & & \downarrow{\scriptstyle 1_C} & & \\
0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & 0.
\end{array}$$

By the Comparison Theorem (Theorem C-3.46), we may fill in dashed arrows to obtain a commutative diagram. In particular, there is a map $\alpha\colon P_1 \to A$ with $\alpha d_2 = 0$; that is, $d_2^*(\alpha) = 0$, so that $\alpha \in \ker d_2^*$ is a cocycle. The Comparison Theorem also says that any two fillings in of the diagram are homotopic; thus, if

$\alpha' \colon P_1 \to A$ is part of a second filling in, there are maps $s_0$ and $s_1$ with $\alpha' - \alpha = 0 \cdot s_1 + s_0 d_1 = s_0 d_1$:




Thus, $\alpha' - \alpha \in \operatorname{im} d_1^*$, and so the homology class $\alpha + \operatorname{im} d_1^* \in \operatorname{Ext}^1(C, A)$ is well-defined. We leave as an exercise for the reader that equivalent extensions $\xi$ and $\xi'$ determine the same element of Ext. Thus,

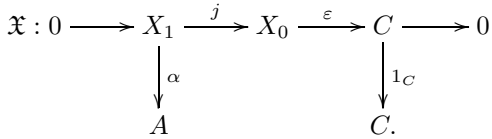
$$\psi \colon e(C, A) \to \operatorname{Ext}^1(C, A),$$

given by

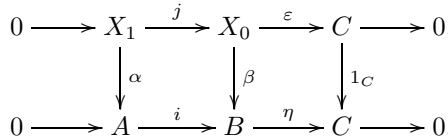
$$\psi([\xi]) = \alpha + \operatorname{im} d_1^*,$$

is a well-defined function. Note that if $\xi$ is a split extension, then $\psi([\xi]) = 0$. In order to prove that $\psi$ is a bijection, we first analyze the diagram containing the map $\alpha$.

**Lemma C-3.87.** *Let* $\mathfrak{X} : \ 0 \to X_1 \xrightarrow{j} X_0 \xrightarrow{\varepsilon} C \to 0$ *be an extension of a module* $X_1$ *by a module* $C$*. Given a module* $A$*, consider the diagram*




(i) *There exists a commutative diagram with exact rows completing the given diagram:*




(ii) *Any two bottom rows of completed diagrams are equivalent extensions.*

**Proof.**

(i) We define $B$ as the pushout of $j$ and $\alpha$. Thus, if

$$S = \big\{(\alpha x_1, -j x_1) \in A \oplus X_0 : x_1 \in X_1\big\},$$

then define $B = (A \oplus X_0)/S$,

$$i \colon a \mapsto (a, 0) + S, \ \beta \colon x_0 \mapsto (0, x_0) + S, \text{ and } \eta \colon (a, x_0) + S \mapsto \varepsilon x_0.$$

That $\eta$ is well-defined, that the diagram commutes, and that the bottom row is exact are left for the reader to check.

(ii) Let

$$\begin{array}{ccccccccc} 0 & \longrightarrow & X_1 & \xrightarrow{j} & X_0 & \xrightarrow{\varepsilon} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta'} & & \downarrow{\scriptstyle 1_C} & & \\ 0 & \longrightarrow & A & \xrightarrow{i'} & B' & \xrightarrow{\eta'} & C & \longrightarrow & 0 \end{array}$$

be a second completion of the diagram. Define $f\colon A \oplus X_0 \to B'$ by

$$f\colon (a, x_0) \mapsto i'a + \beta' x_0.$$

We claim that $f$ is surjective. If $b' \in B'$, then $\eta' b' \in C$, and so there is $x_0 \in X_0$ with $\varepsilon x_0 = \eta' b'$. Commutativity gives $\eta'\beta' x_0 = \varepsilon x_0 = \eta' b'$. Hence, $b' - \beta' x_0 \in \ker \eta' = \operatorname{im} i'$, and so there is $a \in A$ with $i'a = b' - \beta' x_0$. Therefore, $b' = i'a + \beta' x_0 \in \operatorname{im} f$, as desired.

We now show that $\ker f = S$. If $(\alpha x_1, -j x_1) \in S$, then $f(\alpha x_1, -j x_1) = i'\alpha x_1 - \beta' j x_1 = 0$, by commutativity of the first square of the diagram, and so $S \subseteq \ker f$. For the reverse inclusion, let $(a, x_0) \in \ker f$, so that $i'a + \beta' x_0 = 0$. Commutativity of the second square gives $\varepsilon x_0 = \eta'\beta' x_0 = -\eta' i' a = 0$. Hence, $x_0 \in \ker \varepsilon = \operatorname{im} j$, so there is $x_1 \in X_1$ with $j x_1 = x_0$. Thus, $i'a = -\beta' x_0 = -\beta' j x_1 = -i'\alpha x_1$. Since $i'$ is injective, we have $a = -\alpha x_1$. Replacing $x_1$ by $y_1 = -x_1$, we have $(a, x_0) = (\alpha y_1, -j y_1) \in S$, as desired.

Finally, define $\varphi\colon B \to B'$ by

$$\varphi\colon (a, x_0) + S \mapsto f(a, x_0) = i'a + \beta' x_0$$

($\varphi$ is well-defined because $B = (A \oplus X_0)/S$ and $S = \ker f$). To show commutativity of the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{i} & B & \xrightarrow{\eta} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle 1_A} & & \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle 1_C} & & \\ 0 & \longrightarrow & A & \xrightarrow{i'} & B' & \xrightarrow{\eta'} & C & \longrightarrow & 0 \end{array}$$

we use the definitions of the maps $i$ and $\eta$ in part (i). For the first square, if $a \in A$, then $\varphi i a = \varphi((a, 0) + S) = i'a$. For the second square,

$$\begin{aligned} \eta'\varphi\colon (a, x_0) + S &\mapsto \eta'(i'a + \beta' x_0) \\ &= \eta'\beta' x_0 \\ &= \varepsilon x_0 \\ &= \eta((a, x_0) + S). \end{aligned}$$

Therefore, the two bottom rows are equivalent extensions. •

**Notation.** Denote the extension of $A$ by $C$ just constructed by

$$\alpha\mathfrak{X}.$$

The dual result is true; it is related to the construction of Ext using injective resolutions of the second variable $A$.

**Lemma C-3.88.** *Let $A$ and $Y_0$ be modules, and let $\mathfrak{X}' : 0 \to A \to Y_0 \to Y_1 \to 0$ be an extension of $A$ by $Y_1$. Given a module $C$, consider the diagram*

$$\begin{array}{ccccccccc} & & & & A & & C & & \\ & & & & \downarrow{\scriptstyle 1_A} & & \downarrow{\scriptstyle \gamma} & & \\ \mathfrak{X}' : & 0 & \longrightarrow & A & \longrightarrow & Y_0 & \xrightarrow{p} & Y_1 & \longrightarrow 0. \end{array}$$

(i) *There exists a commutative diagram with exact rows completing the given diagram:*

$$\begin{array}{cccccccccc} \mathfrak{X}'\gamma : & 0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & 0 \\ & & & \downarrow{\scriptstyle 1_A} & & \downarrow & & \downarrow{\scriptstyle \gamma} & & \\ \mathfrak{X}' : & 0 & \longrightarrow & A & \longrightarrow & Y_0 & \xrightarrow{p} & Y_1 & \longrightarrow & 0. \end{array}$$

(ii) *Any two top rows of completed diagrams are equivalent extensions.*

**Proof.** Dual to that of Lemma C-3.87; in particular, construct the top row using the pullback of $\gamma$ and $p$. •

**Notation.** Denote the extension of $A$ by $C$ just constructed by

$$\mathfrak{X}'\gamma.$$

**Theorem C-3.89.** *The function $\psi\colon e(C, A) \to \operatorname{Ext}^1(C, A)$ is a bijection.*

**Proof.** We construct an inverse $\theta\colon \operatorname{Ext}^1(C, A) \to e(C, A)$ for $\psi$. Choose a projective resolution of $C$, so there is an exact sequence

$$\cdots \to P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \to C \to 0,$$

and choose a 1-cocycle $\alpha\colon P_1 \to A$. Since $\alpha$ is a cocycle, we have $0 = d_2^*(\alpha) = \alpha d_2$, so that $\alpha$ induces a homomorphism $\alpha'\colon P_1/\operatorname{im} d_2 \to A$ (if $x_1 \in P_1$, then $\alpha' : x_1 + \operatorname{im} d_2 \mapsto \alpha(x_1)$). Let $\mathfrak{X}$ denote the extension

$$\mathfrak{X} : 0 \to P_1/\operatorname{im} d_2 \to P_0 \to C \to 0.$$

As in Lemma C-3.87(i), there is a commutative diagram with exact rows:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & P_1/\operatorname{im} d_2 & \longrightarrow & P_0 & \longrightarrow & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha'} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle 1_C} & & \\ 0 & \longrightarrow & A & \xrightarrow{i} & B & \longrightarrow & C & \longrightarrow & 0. \end{array}$$

Define $\theta\colon \operatorname{Ext}^1(C, A) \to e(C, A)$ using the construction in the lemma:

$$\theta(\alpha + \operatorname{im} d_1^*) = [\alpha'\mathfrak{X}].$$

We begin by showing that $\theta$ is independent of the choice of cocycle $\alpha$. If $\zeta$ is another representative of the coset $\alpha + \operatorname{im} d_1^*$, then there is a map $s\colon P_0 \to A$ with

$\zeta = \alpha + sd_1$. But it is easy to see that the following diagram commutes:

$$\begin{array}{ccccccccc} P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \longrightarrow & C & \longrightarrow & 0 \\ \downarrow & & \downarrow{\scriptstyle \alpha+sd_1} & & \downarrow{\scriptstyle \beta+is} & & \downarrow{\scriptstyle 1_C} & & \\ 0 & \longrightarrow & A & \xrightarrow{i} & B & \longrightarrow & C & \longrightarrow & 0. \end{array}$$

As the bottom row has not changed, we have $[\alpha'\mathfrak{X}] = [\zeta'\mathfrak{X}]$.

It remains to show that the composites $\psi\theta$ and $\theta\psi$ are identities. If $\alpha + \operatorname{im} d_1^* \in \operatorname{Ext}^1(C, A)$, then $\theta(\alpha + \operatorname{im} d_1^*)$ is the bottom row of the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & P_1/\operatorname{im} d_2 & \longrightarrow & P_0 & \longrightarrow & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha'} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle 1_C} & & \\ 0 & \longrightarrow & A & \xrightarrow{i} & B & \longrightarrow & C & \longrightarrow & 0 \end{array}$$

and $\psi\theta(\alpha + \operatorname{im} d_1^*)$ is the homology class of a cocycle fitting this diagram. Clearly, $\alpha$ is such a cocycle; and so $\psi\theta$ is the identity. For the other composite, start with an extension $\xi$, and then imbed it as the bottom row of a diagram

$$\begin{array}{ccccccccc} P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \longrightarrow & C & \longrightarrow & 0 \\ \vdots & & \downarrow{\scriptstyle \alpha} & & \downarrow & & \downarrow{\scriptstyle 1_C} & & \\ 0 & \longrightarrow & A & \xrightarrow{i} & B & \longrightarrow & C & \longrightarrow & 0. \end{array}$$

Both $\xi$ and $\alpha'\mathfrak{X}$ are bottom rows of such a diagram, and so Lemma C-3.87(ii) shows that $[\xi] = [\alpha'\mathfrak{X}]$. •

We can now prove the converse of Proposition C-3.85.

**Corollary C-3.90.** *For any modules $C$ and $A$, every extension of $A$ by $C$ is split if and only if $\operatorname{Ext}^1_R(C, A) = \{0\}$.*

**Proof.** If every extension is split, then $|e(C, A)| = 1$, so that $|\operatorname{Ext}^1_R(C, A)| = 1$, by Theorem C-3.89; hence, $\operatorname{Ext}^1_R(C, A) = \{0\}$. Conversely, if $\operatorname{Ext}^1_R(C, A) = \{0\}$, then Proposition C-3.85 says that every extension is split. •

**Example C-3.91.** If $p$ is a prime, then $\operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_p, \mathbb{Z}_p) \cong \mathbb{Z}_p$, as we saw in Example C-3.84. On the other hand, it follows from Theorem C-3.89 that there are $p$ equivalence classes of extensions $0 \to \mathbb{Z}_p \to B \to \mathbb{Z}_p \to 0$. But $|B| = p^2$, so there are only two choices for middle groups $B$ up to isomorphism: $B \cong \mathbb{Z}_{p^2}$ or $B \cong \mathbb{Z}_p \oplus \mathbb{Z}_p$. Of course, this is consistent with Example C-3.18. ◀

The torsion subgroup of a group may not be a direct summand; the following proof by homology is quite different from that of Exercise B-4.61 on page 507 in Part 1.

**Proposition C-3.92.** *There exists an abelian group $G$ whose torsion subgroup is not a direct summand of $G$; in fact, we may choose $tG = \sum_p \mathbb{Z}_p$, where the sum is over all primes $p$.*

**Proof.** It suffices to prove that $\mathrm{Ext}^1\big(\mathbb{Q}, \sum_p \mathbb{Z}_p\big) \neq 0$, for this will give a nonsplit extension $0 \to \sum_p \mathbb{Z}_p \to G \to \mathbb{Q} \to 0$; moreover, since $\mathbb{Q}$ is torsion-free, it follows that $\sum_p \mathbb{Z}_p = tG$.

Consider the exact sequence $0 \to \sum_p \mathbb{Z}_p \to \prod_p \mathbb{Z}_p \to D \to 0$. By Exercise B-4.64 on page 507 in Part 1, we know that $D$ is divisible.[15] There is an exact sequence

$$\mathrm{Hom}\Big(\mathbb{Q}, \prod_p \mathbb{Z}_p\Big) \to \mathrm{Hom}(\mathbb{Q}, D) \xrightarrow{\partial} \mathrm{Ext}^1\Big(\mathbb{Q}, \sum_p \mathbb{Z}_p\Big) \to \mathrm{Ext}^1(\mathbb{Q}, \prod \mathbb{Z}_p).$$

But $\partial$ is an isomorphism: $\mathrm{Ext}^1\big(\mathbb{Q}, \prod_p \mathbb{Z}_p\big) \cong \prod \mathrm{Ext}^1(\mathbb{Q}, \mathbb{Z}_p) = \{0\}$, by Propositions C-3.81 and C-3.93, while $\mathrm{Hom}(\mathbb{Q}, \prod_p \mathbb{Z}_p) \cong \prod \mathrm{Hom}(\mathbb{Q}, \mathbb{Z}_p) = \{0\}$, by Theorem B-4.8 in Part 1. Since $\mathrm{Hom}(\mathbb{Q}, D) \neq \{0\}$, we have $\mathrm{Ext}^1(\mathbb{Q}, \sum_p \mathbb{Z}_p) \neq \{0\}$. •

**Remark.** We can prove that a torsion abelian group $T$ has the property that it is a direct summand of any group containing it as its torsion subgroup if and only if $T \cong B \oplus D$, where $B$ has bounded order and $D$ is divisible. ◀

Here is another application of Ext.

**Proposition C-3.93.**

(i) *If $F$ is a torsion-free abelian group and $T$ is an abelian group of* ***bounded order*** *(that is, $nT = \{0\}$ for some positive integer $n$), then* $\mathrm{Ext}^1(F, T) = \{0\}$.

(ii) *Let $G$ be an abelian group. If the torsion subgroup $tG$ of $G$ is of bounded order, then $tG$ is a direct summand of $G$.*

**Proof.**

(i) Since $F$ is torsion-free, it is a flat $\mathbb{Z}$-module, by Corollary B-4.105 in Part 1, so that exactness of $0 \to \mathbb{Z} \to \mathbb{Q}$ gives exactness of $0 \to \mathbb{Z} \otimes F \to \mathbb{Q} \otimes F$. Thus, $F \cong \mathbb{Z} \otimes F$ can be imbedded in a vector space $V$ over $\mathbb{Q}$; namely, $V = \mathbb{Q} \otimes F$. Applying the contravariant functor $\mathrm{Hom}(\ , T)$ to $0 \to F \to V \to V/F \to 0$ gives an exact sequence

$$\mathrm{Ext}^1(V, T) \to \mathrm{Ext}^1(F, T) \to \mathrm{Ext}^2(V/F, T).$$

The last term is $\{0\}$, by Exercise C-3.37 on page 291, and $\mathrm{Ext}^1(V, T)$ is (torsion-free) divisible, by Example C-3.70, so that $\mathrm{Ext}^1(F, T)$ is divisible. Since $T$ has bounded order, Exercise C-3.39 on page 292 gives $\mathrm{Ext}^1(F, T) = \{0\}$.

(ii) To prove that the extension $0 \to tG \to G \to G/tG \to 0$ splits, it suffices to prove that $\mathrm{Ext}^1(G/tG, tG) = \{0\}$. Since $G/tG$ is torsion-free, this follows from part (i) and Corollary C-3.90. •

[15] In truth, $D \cong \mathbb{R}$: it is a torsion-free divisible group, hence it is a vector space over $\mathbb{Q}$, by Proposition B-4.69 in Part 1, and we can check that $\dim(D) =$ continuum, which is the dimension of $\mathbb{R}$ as a vector space over $\mathbb{Q}$.

If $\mathcal{E}$ is a set and $\psi\colon \mathcal{E} \to G$ is a bijection to a group $G$, then there is a unique group structure on $\mathcal{E}$ that makes it a group and $\psi$ an isomorphism (if $e, e' \in \mathcal{E}$, then $e = \psi^{-1}(g)$ and $e' = \psi^{-1}(g')$; define $ee' = \psi^{-1}(gg')$). In particular, Theorem C-3.89 implies that there is a group structure on $e(C, A)$; here are the necessary definitions.

Define the ***diagonal map*** $\Delta_C\colon C \to C \oplus C$ by $\Delta_C\colon c \mapsto (c, c)$, and define the ***codiagonal map*** $\nabla_A\colon A \oplus A \to A$ by $\nabla_A\colon (a_1, a_2) \mapsto a_1 + a_2$. Note that if $f, f'\colon C \to A$ is a homomorphism, then the composite $\nabla_A(f \oplus f')\Delta$ maps $C \to C \oplus C \to A \oplus A \to A$. It is easy to check that $\nabla_A(f \oplus f')\Delta = f + f'$, so that this formula describes addition in $\operatorname{Hom}(C, A)$,

$$\begin{aligned}\nabla_A(f \oplus f')\Delta(c) &= \nabla_A(f \oplus f')(c, c)\\ &= \nabla_A(fc, f'c) = fc + f'c = (f + f')(c).\end{aligned}$$

Now Ext is a generalized Hom, and so we mimic this definition to define addition in $e(C, A)$.

If $\xi : 0 \to A \to B \to C \to 0$ and $\xi' : 0 \to A' \to B' \to C' \to 0$ are extensions, then their ***direct sum*** is the extension

$$\xi \oplus \xi' : 0 \to A \oplus A' \to B \oplus B' \to C \oplus C' \to 0.$$

The ***Baer sum*** $[\xi] + [\xi']$ is defined to be the equivalence class $[\nabla_A(\xi \oplus \xi')\Delta_C]$ (we have already defined $\alpha\mathfrak{X}$ and $\mathfrak{X}'\gamma$). To show that Baer sum is well-defined, we first show that $\alpha(\mathfrak{X}'\gamma)$ is equivalent to $(\alpha\mathfrak{X}')\gamma$. One then shows that $e(C, A)$ is a group under this operation by showing that $\psi([\xi] + [\xi']) = \psi([\nabla_A(\xi \oplus \xi')\Delta_C])$. The identity element is the class of the split extension, and the inverse of $[\xi]$ is $[(-1_A)\xi]$.

This description of $\operatorname{Ext}^1$ has been generalized by Yoneda to a description of $\operatorname{Ext}^n$ for all $n$. Elements of Yoneda's $\operatorname{Ext}^n(C, A)$ are certain equivalence classes of exact sequences

$$0 \to A \to B_1 \to \cdots \to B_n \to C \to 0,$$

and we add them by a generalized Baer sum (see Mac Lane [**145**], pp. 82–87). Thus, there is a construction of Ext that does not use derived functors. Indeed, we can construct $\operatorname{Ext}^n$ without using projectives or injectives.

In their investigation of finite-dimensional algebras, M. Auslander and Reiten introduced the following notion.

**Definition.** An exact sequence of left $R$-modules, over any ring $R$,

$$\mathfrak{X}: \quad 0 \to N \to X \to M \to 0$$

is ***almost split*** if it is not split, if both $N$ and $M$ are indecomposable modules, and if for all $R$-modules $C$ and every $R$-map $\varphi\colon C \to M$ that is not an isomorphism, the exact sequence $\mathfrak{X}\varphi$ is split.

Another way to say this is that $[\mathfrak{X}]$ is a nonzero element of $\operatorname{Ext}^1_R(N, M)$, where $N$ and $M$ are indecomposable, and $[\mathfrak{X}] \in \ker \varphi^*$ for every $\varphi\colon C \to M$ that is not an isomorphism. Auslander and Reiten proved that for every indecomposable module $M$ that is not projective, there exists an almost split exact sequence ending with $M$. Dually, they proved that for every indecomposable module $N$ that is not injective, there exists an almost split exact sequence beginning with $N$.

It is now Tor's turn. We begin with a result that has no analog for Ext.

**Theorem C-3.94.** *If $R$ is a ring, $A$ is a right $R$-module, and $B$ is a left $R$-module, then*

$$\operatorname{Tor}_n^R(A,B) \cong \operatorname{Tor}_n^{R^{\mathrm{op}}}(B,A)$$

*for all $n \geq 0$, where $R^{\mathrm{op}}$ is the opposite ring of $R$.*

**Proof.** Recall Proposition B-1.23 in Part 1: every left $R$-module is a right $R^{\mathrm{op}}$-module, and every right $R$-module is a left $R^{\mathrm{op}}$-module. Choose a deleted projective resolution $\mathbf{P}_A$ of $A$. It is easy to see that $t\colon \mathbf{P}_A \otimes_R B \to B \otimes_{R^{\mathrm{op}}} \mathbf{P}_A$ is a chain map of $\mathbb{Z}$-complexes, where

$$t_n\colon P_n \otimes_R B \to B \otimes_{R^{\mathrm{op}}} P_n$$

is given by

$$t_n\colon x_n \otimes b \mapsto b \otimes x_n.$$

Since each $t_n$ is an isomorphism of abelian groups (its inverse is $b \otimes x_n \mapsto x_n \otimes b$), the chain map $t$ is an isomorphism of complexes. By Exercise C-3.24 on page 269,

$$\operatorname{Tor}_n^R(A,B) = H_n(\mathbf{P}_A \otimes_R B) \cong H_n(B \otimes_{R^{\mathrm{op}}} \mathbf{P}_A)$$

for all $n$. But $\mathbf{P}_A$, viewed as a complex of left $R^{\mathrm{op}}$-modules, is a deleted projective resolution of $A$ as a left $R^{\mathrm{op}}$-module, and so $H_n(B \otimes_{R^{\mathrm{op}}} \mathbf{P}_A) \cong \operatorname{Tor}_n^{R^{\mathrm{op}}}(B,A)$. •

In light of this result, theorems about $\operatorname{Tor}(A,\ )$ will yield results about $\operatorname{Tor}(\ ,B)$; we will not have to say "similarly in the other variable".

**Corollary C-3.95.** *If $R$ is a commutative ring and $A$ and $B$ are $R$-modules, then for all $n \geq 0$,*

$$\operatorname{Tor}_n^R(A,B) \cong \operatorname{Tor}_n^R(B,A).$$

We know that $\operatorname{Tor}_n$ vanishes on projectives; we now show that it vanishes on flat modules.

**Proposition C-3.96.** *A right $R$-module $F$ is flat if and only if $\operatorname{Tor}_n^R(F,M) = \{0\}$ for all $n \geq 1$ and every left $R$-module $M$.*

**Proof.** Let $0 \to N \xrightarrow{i} P \to M \to 0$ be exact, where $P$ is projective. There is an exact sequence

$$\operatorname{Tor}_1(F,P) \to \operatorname{Tor}_1(F,M) \to F \otimes N \xrightarrow{1\otimes i} F \otimes P.$$

Now $\operatorname{Tor}_1(F,P) = \{0\}$, because $P$ is projective, so that $\operatorname{Tor}_1(F,M) = \ker(1 \otimes i)$. Since $F$ is flat, however, $\ker(1 \otimes i) = \{0\}$, and so $\operatorname{Tor}_1(F,M) = \{0\}$. The result for all $n \geq 1$ follows by dimension shifting.

For the converse, $0 \to A \xrightarrow{i} B$ exact implies exactness of

$$0 = \operatorname{Tor}_1(F, B/A) \to F \otimes A \xrightarrow{1\otimes i} F \otimes B.$$

Hence, $1 \otimes i$ is an injection, and so $F$ is flat. (Notice that we have only assumed the vanishing of $\operatorname{Tor}_1$ in proving the converse.) •

**Proposition C-3.97.** *If $\{B_k : k \in K\}$ is a family of left $R$-modules, then there are natural isomorphisms, for all $n$,*

$$\operatorname{Tor}_n^R\Big(A, \sum_{k\in K} B_k\Big) \cong \sum_{k\in K} \operatorname{Tor}_n^R(A, B_k).$$

*There is also an isomorphism if the direct sum is in the first variable.*

**Proof.** The proof is by dimension shifting. The base step is Theorem B-4.86 in Part 1, for $\operatorname{Tor}_0(A, \ )$ is naturally equivalent to $A \otimes -$.

For the inductive step, choose, for each $k \in K$, a short exact sequence

$$0 \to N_k \to P_k \to B_k \to 0,$$

where $P_k$ is projective. There is an exact sequence

$$0 \to \sum_k N_k \to \sum_k P_k \to \sum_k B_k \to 0,$$

and $\sum_k P_k$ is projective, for every sum of projectives is projective. There is a commutative diagram with exact rows:

$$\begin{array}{ccccccc}
\operatorname{Tor}_1(A, \sum P_k) & \longrightarrow & \operatorname{Tor}_1(A, \sum B_k) & \xrightarrow{\partial} & A \otimes \sum N_k & \longrightarrow & A \otimes \sum P_k \\
 & & \downarrow & & \downarrow \tau & & \downarrow \sigma \\
\sum \operatorname{Tor}_1(A, P_k) & \longrightarrow & \sum \operatorname{Tor}_1(A, B_k) & \xrightarrow{\partial'} & \sum A \otimes N_k & \longrightarrow & \sum A \otimes P_k,
\end{array}$$

where the maps in the bottom row are just the usual induced maps in each coordinate and the maps $\tau$ and $\sigma$ are the isomorphisms given by Theorem B-4.86 in Part 1. The proof is completed by dimension shifting. •

**Example C-3.98.**

(i) We show, for every abelian group $B$, that

$$\operatorname{Tor}_1^{\mathbb{Z}}(\mathbb{Z}_n, B) \cong B[n] = \{b \in B : nb = 0\}.$$

There is an exact sequence

$$0 \to \mathbb{Z} \xrightarrow{\mu_n} \mathbb{Z} \to \mathbb{Z}_n \to 0,$$

where $\mu_n$ is multiplication by $n$. Applying $- \otimes B$ gives exactness of

$$\operatorname{Tor}_1(\mathbb{Z}, B) \to \operatorname{Tor}_1(\mathbb{Z}_n, B) \to \mathbb{Z} \otimes B \xrightarrow{1\otimes\mu_n} \mathbb{Z} \otimes B.$$

Now $\operatorname{Tor}_1(\mathbb{Z}, B) = \{0\}$, because $\mathbb{Z}$ is projective. Moreover, $1 \otimes \mu_n$ is also multiplication by $n$, while $\mathbb{Z} \otimes B = B$. More precisely, $\mathbb{Z} \otimes -$ is naturally equivalent to the identity functor on **Ab**, and so there is a commutative diagram with exact rows

$$\begin{array}{ccccccc}
0 & \longrightarrow & B[n] & \longrightarrow & B & \xrightarrow{\mu_n} & B \\
 & & \downarrow & & \downarrow \tau_B & & \downarrow \tau_B \\
0 & \longrightarrow & \operatorname{Tor}_1(\mathbb{Z}_n, B) & \longrightarrow & \mathbb{Z} \otimes B & \xrightarrow{1\otimes\mu_n} & \mathbb{Z} \otimes B.
\end{array}$$

By Proposition B-1.47 in Part 1, $B[n] \cong \operatorname{Tor}_1(\mathbb{Z}_n, B)$.

(ii) We can now compute $\operatorname{Tor}_1^{\mathbb{Z}}(A, B)$ whenever $A$ and $B$ are finitely generated abelian groups. By the Fundamental Theorem of Finite Abelian Groups, both $A$ and $B$ are direct sums of cyclic groups. Since Tor commutes with direct sums, $\operatorname{Tor}_1^{\mathbb{Z}}(A, B)$ is the direct sum of groups $\operatorname{Tor}_1^{\mathbb{Z}}(C, D)$, where $C$ and $D$ are cyclic. We may assume that $C$ and $D$ are finite; otherwise, they are projective and $\operatorname{Tor}_1 = \{0\}$. This calculation can be completed using part (i) and Exercise B-3.16 on page 377 in Part 1, which says that if $D$ is a cyclic group of finite order $m$, then $D[n]$ is a cyclic group of order $d$, where $d = \gcd(m, n)$. ◀

In contrast to Ext, Proposition C-3.97 can be generalized by replacing sums by direct limits.

**Proposition C-3.99.** *If $\{B_i, \varphi_j^i\}$ is a direct system of left $R$-modules over a directed index set $I$, then there is an isomorphism, for all right $R$-modules $A$ and for all $n \geq 0$,*

$$\operatorname{Tor}_n^R\big(A, \varinjlim B_i\big) \cong \varinjlim \operatorname{Tor}_n^R(A, B_i).$$

**Proof.** The proof is by dimension shifting. The base step is Theorem B-7.15 in Part 1, for $\operatorname{Tor}_0(A, \ )$ is naturally equivalent to $A \otimes -$.

For the inductive step, choose, for each $i \in I$, a short exact sequence

$$0 \to N_i \to P_i \to B_i \to 0,$$

where $P_i$ is projective. Since the index set is directed, Proposition B-7.14 in Part 1 says that there is an exact sequence

$$0 \to \varinjlim N_i \to \varinjlim P_i \to \varinjlim B_i \to 0.$$

Now $\varinjlim P_i$ is flat, for every projective module is flat, and a direct limit of flat modules is flat, by Corollary B-7.17 in Part 1. There is a commutative diagram with exact rows:

$$\begin{array}{ccccccc}
\operatorname{Tor}_1(A, \varinjlim P_i) & \longrightarrow & \operatorname{Tor}_1(A, \varinjlim B_i) & \xrightarrow{\partial} & A \otimes \varinjlim N_i & \longrightarrow & A \otimes \varinjlim P_i \\
 & & \downarrow & & \downarrow \tau & & \downarrow \sigma \\
\varinjlim \operatorname{Tor}_1(A, P_i) & \longrightarrow & \varinjlim \operatorname{Tor}_1(A, B_i) & \xrightarrow{\vec{\partial}} & \varinjlim A \otimes N_i & \longrightarrow & \varinjlim A \otimes P_i,
\end{array}$$

where the maps in the bottom row are just the usual induced maps between direct limits and the maps $\tau$ and $\sigma$ are the isomorphisms given by Theorem B-7.15 in Part 1. The step $n \geq 2$ is routine. •

This last proposition generalizes Lemma B-4.103 in Part 1, which says that if every finitely generated submodule of a module $M$ is flat, then $M$ itself is flat. After all, by Example B-7.13 in Part 1, $M$ is a direct limit, over a directed index set, of its finitely generated submodules.

When the ring $R$ is noncommutative, $A \otimes_R B$ is an abelian group, but it need not be an $R$-module.

**Proposition C-3.100.**

(i) *Let $r \in Z(R)$ be a central element, let $A$ be a right $R$-module, and let $B$ be a left $R$-module. If $\mu_r\colon B \to B$ is multiplication by $r$, then the induced map*

$$\mu_{r*}\colon \operatorname{Tor}_n^R(A,B) \to \operatorname{Tor}_n^R(A,B)$$

*is also multiplication by $r$.*

(ii) *If $R$ is a commutative ring, then $\operatorname{Tor}_n^R(A,B)$ is an $R$-module.*

**Proof.**

(i) This follows at once from Example C-3.47.

(ii) This follows from (i) if we define scalar multiplication by $r$ to be $\mu_{r*}$. •

We are now going to assume that $R$ is a domain, so that the notion of torsion submodule is defined, and we shall see why Tor is so called.

**Lemma C-3.101.** *Let $R$ be a domain, let $Q = \operatorname{Frac}(R)$, and let $K = Q/R$.*

(i) *If $A$ is a torsion $R$-module, then $\operatorname{Tor}_1^R(K,A) \cong A$.*

(ii) *For every $R$-module $A$, we have $\operatorname{Tor}_n(K,A) = \{0\}$ for all $n \geq 2$.*

(iii) *If $A$ is a torsion-free $R$-module, then $\operatorname{Tor}_1(K,A) = \{0\}$.*

**Proof.**

(i) Exactness of $0 \to R \to Q \to K \to 0$ gives exactness of

$$\operatorname{Tor}_1(Q,A) \to \operatorname{Tor}_1(K,A) \to R \otimes A \to Q \otimes A.$$

Now $Q$ is flat, by Corollary B-4.106 in Part 1, and so $\operatorname{Tor}_1(Q,A) = \{0\}$, by Proposition C-3.96. The last term $Q \otimes A = \{0\}$ because $Q$ is divisible and $A$ is torsion, by Exercise B-4.95 on page 542 in Part 1, and so the middle map $\operatorname{Tor}_1(K,A) \to R \otimes A$ is an isomorphism.

(ii) There is an exact sequence

$$\operatorname{Tor}_n(Q,A) \to \operatorname{Tor}_n(K,A) \to \operatorname{Tor}_{n-1}(R,A).$$

Since $n \geq 2$, we have $n-1 \geq 1$, and so both the first and third Tor's are $\{0\}$, because $Q$ and $R$ are flat. Exactness gives $\operatorname{Tor}_n(K,A) = \{0\}$.

(iii) By Theorem B-4.64 in Part 1, there is an injective $R$-module $E$ containing $A$ as a submodule. Since $A$ is torsion-free, however, $A \cap tE = \{0\}$, and so $A$ is imbedded in $E/tE$. By Lemma B-4.60 in Part 1, injective modules are divisible, and so $E$ is divisible, as is its quotient $E/tE$. Now $E/tE$ is a vector space over $Q$, for it is a torsion-free divisible $R$-module (Exercise B-4.58 on page 507 in Part 1). Let us denote $E/tE$ by $V$. Since every vector space has a basis, $V$ is a direct sum of copies of $Q$. Corollary B-4.106 in Part 1 says that $Q$ is flat, and Lemma B-4.101 in Part 1 says that a direct sum of flat modules is flat. We conclude that $V$ is flat.[16]

[16] Torsion-free $\mathbb{Z}$-modules are flat, but there exist domains $R$ having torsion-free modules that are not flat. In fact, domains for which every torsion-free module is flat, called *Prüfer rings*, are characterized as those domains in which every finitely generated ideal is a projective module.

Exactness of $0 \to A \to V \to V/A \to 0$ gives exactness of

$$\operatorname{Tor}_2(K, V/A) \to \operatorname{Tor}_1(K, A) \to \operatorname{Tor}_1(K, V).$$

Now $\operatorname{Tor}_2(K, V/A) = \{0\}$, by (ii), and $\operatorname{Tor}_1(K, V) = \{0\}$, because $V$ is flat. We conclude from exactness that $\operatorname{Tor}_1(K, A) = \{0\}$. •

The next result shows why Tor is so called.

**Theorem C-3.102.**

(i) *If $R$ is a domain, $Q = \operatorname{Frac}(R)$, and $K = Q/R$, then the functor $\operatorname{Tor}_1^R(K, \ )$ is naturally equivalent to the torsion functor.*

(ii) $\operatorname{Tor}_1^R(K, A) \cong tA$ *for all $R$-modules $A$.*

**Proof.** Use the exactness of the sequence

$$\operatorname{Tor}_2(K, A/tA) \to \operatorname{Tor}_1(K, tA) \xrightarrow{\iota_A} \operatorname{Tor}_1(K, A) \to \operatorname{Tor}_1(K, A/tA).$$

The first and last terms are $\{0\}$, by Lemma C-3.101(ii) and Lemma C-3.101(iii). Therefore, the map $\iota_A \colon \operatorname{Tor}_1(K, tA) \to \operatorname{Tor}_1(K, A)$ is an isomorphism.

Let $f \colon A \to B$ and let $f' \colon tA \to tB$ be its restriction. The following diagram commutes, because $\operatorname{Tor}_1(K, \ )$ is a functor, and this says that the isomorphisms $\iota_A$ constitute a natural transformation:

$$\begin{array}{ccc} \operatorname{Tor}_1(K, tA) & \xrightarrow{\iota_A} & \operatorname{Tor}_1(K, A) \\ {\scriptstyle f'_*}\downarrow & & \downarrow{\scriptstyle f_*} \\ \operatorname{Tor}_1(K, tB) & \xrightarrow{\iota_B} & \operatorname{Tor}_1(K, B). \end{array} \quad \bullet$$

There is a construction of $\operatorname{Tor}_1^{\mathbb{Z}}(A, B)$ by generators and relations. Consider all triples $(a, n, b)$, where $a \in A$, $b \in B$, $na = 0$, and $nb = 0$. Then $\operatorname{Tor}_1^{\mathbb{Z}}(A, B)$ is generated by all such triples subject to the relations (whenever both sides are defined)

$$\begin{aligned} (a + a', n, b) &= (a, n, b) + (a', n, b), \\ (a, n, b + b') &= (a, n, b) + (a, n, b'), \\ (ma, n, b) &= (a, mn, b) = (a, m, nb). \end{aligned}$$

For a proof of this result and its generalization to $\operatorname{Tor}_n^R(A, B)$ for arbitrary rings $R$, see Mac Lane [**145**], pp. 150–159.

The Tor functors are very useful in algebraic topology. The *Universal Coefficients Theorem* gives a formula for the homology groups $H_n(X; G)$ with coefficients in an abelian group $G$.

**Theorem C-3.103 (Universal Coefficients).** *For every topological space $X$ and every abelian group $G$, there are isomorphisms for all $n \geq 0$,*

$$H_n(X; G) \cong H_n(X) \otimes_{\mathbb{Z}} G \oplus \operatorname{Tor}_1^{\mathbb{Z}}(H_{n-1}(X), G).$$

**Proof.** See Rotman [**191**], p. 261. •

If we know the homology groups of spaces $X$ and $Y$, then the *Künneth formula* gives a formula for the homology groups of $X \times Y$, and this, too, involves Tor in an essential way.

**Theorem C-3.104 (Künneth Formula).** *For every pair of topological spaces $X$ and $Y$, there are isomorphisms for every $n \geq 0$,*

$$H_n(X \times Y) \cong \sum_i H_i(X) \otimes_{\mathbb{Z}} H_{n-i}(Y) \oplus \sum_p \operatorname{Tor}_1^{\mathbb{Z}}(H_p(X), H_{n-1-p}(Y)).$$

**Proof.** See Rotman [**191**], p. 269. •

## Exercises

∗ **C-3.43.** Prove the following analog of Theorem C-3.45. Let $\mathcal{E}^n\colon {}_R\mathbf{Mod} \to \mathbf{Ab}$ be a sequence of covariant functors, for $n \geq 0$, such that

(i) for every short exact sequence $0 \to A \to B \to C \to 0$, there is a long exact sequence and natural connecting homomorphisms

$$\cdots \to \mathcal{E}^n(A) \to \mathcal{E}^n(B) \to \mathcal{E}^n(C) \xrightarrow{\Delta_n} \mathcal{E}^{n+1}(A) \to \cdots;$$

(ii) there is a left $R$-module $M$ such that $\mathcal{E}^0$ and $\operatorname{Hom}_R(M, \ )$ are naturally equivalent;

(iii) $\mathcal{E}^n(E) = \{0\}$ for all injective modules $E$ and all $n \geq 1$.

Prove that $\mathcal{E}^n$ is naturally equivalent to $\operatorname{Ext}^n(M, \ )$ for all $n \geq 0$.

∗ **C-3.44.** Let $\operatorname{TOR}^n\colon {}_R\mathbf{Mod} \to \mathbf{Ab}$ be a sequence of covariant functors, for $n \geq 0$, such that

(i) for every short exact sequence $0 \to A \to B \to C \to 0$, there is a long exact sequence and natural connecting homomorphisms

$$\cdots \to \operatorname{TOR}_n(A) \to \operatorname{TOR}_n(B) \to \operatorname{TOR}_n(C) \xrightarrow{\Delta_n} \operatorname{TOR}_{n-1}(A) \to \cdots;$$

(ii) there is a left $R$-module $M$ such that $\operatorname{TOR}_0$ and $- \otimes_R M$ are naturally equivalent;

(iii) $\operatorname{TOR}_n(P) = \{0\}$ for all projective modules $P$ and all $n \geq 1$.

Prove that $\operatorname{TOR}_n$ is naturally equivalent to $\operatorname{Tor}_n(\ , M)$ for all $n \geq 0$. (There is a similar result if the first variable is fixed.)

**C-3.45.** Prove that any two split extensions of modules $A$ by $C$ are equivalent.

**C-3.46.** Prove that if $A$ is an abelian group with $nA = A$ for some positive integer $n$, then every extension $0 \to A \to E \to \mathbb{Z}_n \to 0$ splits.

**C-3.47.** If $A$ is a torsion abelian group, prove that $\operatorname{Ext}^1(A, \mathbb{Z}) \cong \operatorname{Hom}(A, S^1)$, where $S^1$ is the circle group.

∗ **C-3.48.** Prove that a left $R$-module $E$ is injective if and only if $\operatorname{Ext}_R^1(A, E) = \{0\}$ for every left $R$-module $A$.

**C-3.49.** For a ring $R$, prove that a left $R$-module $B$ is injective if and only if $\operatorname{Ext}^1(R/I, B) = \{0\}$ for every left ideal $I$.

**Hint.** Use the Baer criterion.

**C-3.50.** Prove that an abelian group $G$ is injective if and only if $\operatorname{Ext}^1(\mathbb{Q}/\mathbb{Z}, G) = \{0\}$.

**C-3.51.** Prove that an abelian group $G$ is free abelian if and only if $\mathrm{Ext}^1(G, F) = \{0\}$ for every free abelian group $F$.[17]

**C-3.52.** If $0 \to A \to B \to C \to 0$ is an exact sequence of right $R$-modules with both $A$ and $C$ flat, prove that $B$ is flat.

**C-3.53.** If $A$ and $B$ are finite abelian groups, prove that $\mathrm{Tor}_1^{\mathbb{Z}}(A, B) \cong A \otimes_{\mathbb{Z}} B$.

**C-3.54.** Let $R$ be a domain, $Q = \mathrm{Frac}(R)$, and $K = Q/R$.

(i) Prove, for every $R$-module $A$, that there is an exact sequence

$$0 \to tA \to A \to Q \otimes A \to K \otimes A \to 0.$$

(ii) Prove that a module $A$ is torsion if and only if $Q \otimes A = \{0\}$.

**C-3.55.** Let $R$ be a domain.

(i) If $B$ is a torsion $R$-module, prove that $\mathrm{Tor}_n(A, B)$ is a torsion $R$-module for all $R$-modules $A$ and for all $n \geq 0$.

(ii) For all $R$-modules $A$ and $B$, prove that $\mathrm{Tor}_n(A, B)$ is a torsion $R$-module for all $n \geq 1$.

**C-3.56.** Let $k$ be a field, let $R = k[x, y]$, and let $I$ be the ideal $(x, y)$.

(i) Prove that $x \otimes y - y \otimes x \in I \otimes_R I$ is nonzero.
**Hint.** Consider $(I/I^2) \otimes (I/I^2)$.

(ii) Prove that $x(x \otimes y - y \otimes x) = 0$, and conclude that $I \otimes_R I$ is not torsion-free.

## C-3.9. Cohomology of Groups

We will see, in this section, that our earlier discussion of the extension problem for groups fits nicely into homological algebra. Recall that Proposition C-3.31 says, for any group $G$, that there is an exact sequence

$$B_3 \xrightarrow{d_3} B_2 \xrightarrow{d_2} B_1 \xrightarrow{d_1} B_0 \xrightarrow{\varepsilon} \mathbb{Z} \to 0,$$

where $B_0$, $B_1$, $B_2$, and $B_3$ are free $G$-modules and $\mathbb{Z}$ is viewed as a trivial $G$-module. In light of the calculations in Section C-3.3, the following definition should now seem reasonable.

**Definition.** Let $G$ be a group, let $A$ be a $G$-module (i.e., a left $\mathbb{Z}G$-module), and let $\mathbb{Z}$ be the integers viewed as a trivial $G$-module (i.e., $gm = m$ for all $g \in G$ and $m \in \mathbb{Z}$). The ***cohomology groups*** of $G$ are

$$H^n(G, A) = \mathrm{Ext}^n_{\mathbb{Z}G}(\mathbb{Z}, A);$$

the ***homology groups*** of $G$ are

$$H_n(G, A) = \mathrm{Tor}_n^{\mathbb{Z}G}(\mathbb{Z}, A).$$

[17] The question whether $\mathrm{Ext}^1(G, \mathbb{Z}) = \{0\}$ implies $G$ is free abelian is known as *Whitehead's problem*. It turns out that if $G$ is countable, then it must be free abelian, but Shelah proved that it is undecideable whether uncountable such $G$ must be free abelian (see Eklof [**61**]).

The history of cohomology of groups is quite interesting. The subject began with the discovery, by the topologist Hurewicz in the 1930s, that if $X$ is a connected *aspherical space* (that is, if the higher homotopy groups of $X$ are all trivial), then all the homology and cohomology groups of $X$ are determined by the fundamental group $\pi = \pi_1(X)$. This led to the question of whether $H_n(X)$ could be described algebraically in terms of $\pi$. For example, Hurewicz proved that $H_1(X) \cong \pi/\pi'$, where $\pi'$ is the commutator subgroup. In 1942, H. Hopf proved that if $\pi$ has a presentation $F/R$, where $F$ is free, then $H_2(X) \cong (R \cap F')/[F, R]$, where $[F, R]$ is the subgroup generated by all commutators of the form $frf^{-1}r^{-1}$ for $f \in F$ and $r \in R$. These results led Eilenberg, Mac Lane, Hopf, Freudenthal, and Eckmann to create cohomology of groups. In addition to its links with group theory and algebraic topology, cohomology of groups is used extensively in algebraic number theory (see [**38**] and [**166**]).

In what follows, we will write $\mathrm{Hom}_G$ instead of $\mathrm{Hom}_{\mathbb{Z}G}$ and $-\otimes_G$ instead of $-\otimes_{\mathbb{Z}G}$. Because of the special role of the trivial $G$-module $\mathbb{Z}$, the augmentation

$$\varepsilon\colon \mathbb{Z}G \to \mathbb{Z},$$

defined by

$$\varepsilon\colon \sum_{x\in G} m_x x \mapsto \sum_{x\in G} m_x,$$

is important. We have seen, in Exercise C-2.12 on page 145, that $\varepsilon$ is a surjective ring homomorphism, and so its kernel, $\mathcal{G}$, is a two-sided ideal in $\mathbb{Z}G$, called the ***augmentation ideal***. Thus, there is an exact sequence of rings

$$0 \to \mathcal{G} \to \mathbb{Z}G \xrightarrow{\varepsilon} \mathbb{Z} \to 0.$$

**Proposition C-3.105.** *Let $G$ be a group with augmentation ideal $\mathcal{G}$. As an abelian group, $\mathcal{G}$ is free abelian with basis $G - 1 = \{x - 1 : x \in G, x \neq 1\}$.*

**Proof.** An element $u = \sum_x m_x x \in \mathbb{Z}G$ lies in $\ker \varepsilon = \mathcal{G}$ if and only if $\sum_x m_x = 0$. Therefore, if $u \in \mathcal{G}$, then

$$u = u - \Big(\sum_x m_x\Big)1 = \sum_x m_x(x - 1).$$

Thus, $\mathcal{G}$ is generated by the nonzero $x - 1$ for $x \in G$.

Suppose that $\sum_{x\neq 1} n_x(x - 1) = 0$. Then $\sum_{x\neq 1} n_x x - \big(\sum_{x\neq 1} n_x\big)1 = 0$ in $\mathbb{Z}G$, which, as an abelian group, is free abelian with basis the elements of $G$. Hence, $n_x = 0$ for all $x \neq 1$. Therefore, the nonzero $x - 1$ comprise a basis of $\mathcal{G}$. •

We begin by examining homology groups.

**Proposition C-3.106.** *If $A$ is a $G$-module, then*

$$H_0(G, A) = \mathbb{Z} \otimes_G A \cong A/\mathcal{G}A.$$

**Proof.** By definition, $H_0(G, A) = \mathrm{Tor}_0^{\mathbb{Z}G}(\mathbb{Z}, A) = \mathbb{Z}\otimes_G A$. Applying the right exact functor $- \otimes_G A$ to the exact sequence

$$0 \to \mathcal{G} \to \mathbb{Z}G \to \mathbb{Z} \to 0$$

gives exactness of the top row of the following commutative diagram:

$$\begin{array}{ccccccc}
\mathcal{G}\otimes_G A & \longrightarrow & \mathbb{Z}G\otimes_G A & \longrightarrow & \mathbb{Z}\otimes_G A & \longrightarrow & 0 \\
\downarrow & & \downarrow & & \vdots & & \\
\mathcal{G}A & \longrightarrow & A & \longrightarrow & A/\mathcal{G}A & \longrightarrow & 0.
\end{array}$$

The two solid vertical arrows are given by $u\otimes a \mapsto ua$. By Proposition B-1.46 in Part 1, there is an isomorphism $\mathbb{Z}\otimes_G A \cong A/\mathcal{G}A$. •

It is easy to see that $A/\mathcal{G}A$ is $G$-trivial; indeed, it is the largest $G$-trivial quotient of $A$.

**Example C-3.107.** Suppose that $E$ is a semidirect product of an abelian group $A$ by a group $G$. Recall that $[G, A]$ is the subgroup generated by all commutators of the form $[x, a] = xax^{-1}a^{-1}$, where $x \in G$ and $a \in A$. If we write commutators additively, as we did at the beginning of this chapter, then

$$[x, a] = x + a - x - a = xa - a = (x - 1)a$$

(recall that $G$ acts on $A$ by conjugation). Therefore, $A/\mathcal{G}A = A/[G, A]$ here. ◀

We are now going to use the independence of the choice of projective resolution to compute the homology groups of a finite cyclic group $G$.

**Lemma C-3.108.** *Let $G = \langle x\rangle$ be a cyclic group of finite order $k$. Define elements $D$ and $N$ in $\mathbb{Z}G$ by*

$$D = x - 1 \quad \textit{and} \quad N = 1 + x + x^2 + \cdots + x^{k-1}.$$

*Then the following sequence is a $G$-free resolution of $Z$:*

$$\cdots \to \mathbb{Z}G \xrightarrow{N} \mathbb{Z}G \xrightarrow{D} \mathbb{Z}G \xrightarrow{N} \mathbb{Z}G \xrightarrow{D} \mathbb{Z}G \xrightarrow{\varepsilon} \mathbb{Z} \to 0,$$

*where $\varepsilon$ is the augmentation and the other maps are multiplication by $N$ and $D$, respectively.*

**Proof.** Obviously, every term $\mathbb{Z}G$ is free; moreover, since $\mathbb{Z}G$ is commutative, the maps are $G$-maps. Now $DN = ND = x^k - 1 = 0$, while if $u \in \mathbb{Z}G$, then

$$\varepsilon D(u) = \varepsilon\big((x - 1)u\big) = \varepsilon(x - 1)\varepsilon(u) = 0,$$

because $\varepsilon$ is a ring map. Thus, we have a complex, and it only remains to prove exactness.

We have already noted that $\varepsilon$ is surjective. Now $\ker\varepsilon = \mathcal{G} = \operatorname{im} D$, by Proposition C-3.105, and so we have exactness at the 0th step.

Suppose $u = \sum_{i=0}^{k-1} m_i x^i \in \ker D$; that is, $(x - 1)u = 0$. Expanding, and using the fact that $\mathbb{Z}G$ has basis $\{1, x, x^2, \ldots, x^{k-1}\}$, we have

$$m_0 = m_1 = \cdots = m_{k-1},$$

so that $u = m_0 N \in \operatorname{im} N$, as desired.

Finally, if $u = \sum_{i=0}^{k-1} m_i x^i \in \ker N$, then $0 = \varepsilon(Nu) = \varepsilon(N)\varepsilon(u) = k\varepsilon(u)$, so that $\varepsilon(u) = \sum_{i=0}^{k-1} m_i = 0$. Therefore,

$$u = -D\big(m_0 1 + (m_0 + m_1)x + \cdots + (m_0 + \cdots + m_{k-1})x^{k-1}\big) \in \operatorname{im} D. \quad \bullet$$

**Definition.** If $A$ is a $G$-module, define submodules

$$A[N] = \big\{a \in A : Na = 0\big\}$$

and

$$A^G = \big\{a \in A : ga = a \text{ for all } g \in G\big\}.$$

**Theorem C-3.109.** *If $G$ is a cyclic group of finite order $k$ and $A$ is a $G$-module, then*

$$\begin{aligned} H_0(G, A) &= A/\mathcal{G}A, \\ H_{2n-1}(G, A) &= A^G/NA \quad \text{for all } n \geq 1, \\ H_{2n}(G, A) &= A[N]/\mathcal{G}A \quad \text{for all } n \geq 1. \end{aligned}$$

**Proof.** Apply $-\otimes_G A$ to the resolution of $\mathbb{Z}$ in Lemma C-3.108. As $\mathbb{Z}G \otimes_{\mathbb{Z}G} A \cong A$, the calculation of ker/im is now simple; use $\operatorname{im} D = \mathcal{G}A$ (which follows from Proposition C-3.105) and the fact that $(x-1) \mid (x^i - 1)$. $\bullet$

**Corollary C-3.110.** *If $G$ is a finite cyclic group of order $k$ and $A$ is a trivial $G$-module, then*

$$\begin{aligned} H_0(G, A) &= A, \\ H_{2n-1}(G, A) &= A/kA \quad \text{for all } n \geq 1, \\ H_{2n}(G, A) &= A[k] \quad \text{for all } n \geq 1. \end{aligned}$$

*In particular,*

$$\begin{aligned} H_0(G, \mathbb{Z}) &= \mathbb{Z}, \\ H_{2n-1}(G, \mathbb{Z}) &= \mathbb{Z}/k\mathbb{Z} \quad \text{for all } n \geq 1, \\ H_{2n}(G, \mathbb{Z}) &= \{0\} \quad \text{for all } n \geq 1. \end{aligned}$$

**Proof.** Since $A$ is $G$-trivial, we have $A^G = A$ and $\mathcal{G}A = \{0\}$ (for $Da = (x-1)a = 0$ because $xa = a$). $\bullet$

We continue computing low-dimensional homology groups of not necessarily cyclic groups; we have already computed $H_0(G, A) \cong A/\mathcal{G}A$ in Proposition C-3.106.

**Lemma C-3.111.** *For any group $G$, we have*

$$H_1(G, \mathbb{Z}) \cong \mathcal{G}/\mathcal{G}^2.$$

**Proof.** The long exact sequence arising from

$$0 \to \mathcal{G} \to \mathbb{Z}G \xrightarrow{\varepsilon} \mathbb{Z} \to 0$$

ends with

$$H_1(G, \mathbb{Z}G) \to H_1(G, \mathbb{Z}) \xrightarrow{\partial} H_0(G, \mathcal{G}) \to H_0(G, \mathbb{Z}G) \xrightarrow{\varepsilon_*} H_0(G, \mathbb{Z}) \to 0.$$

Now $H_1(G, \mathbb{Z}G) = \{0\}$, because $\mathbb{Z}G$ is projective, so that $\partial$ is an injection. Also,

$$H_0(G, \mathbb{Z}G) \cong \mathbb{Z},$$

by Proposition C-3.106. Since $\varepsilon_*$ is surjective, it must be injective as well (if $\ker \varepsilon_* \neq \{0\}$, then $\mathbb{Z}/\ker \varepsilon_*$ is finite; on the other hand, $\mathbb{Z}/\ker \varepsilon_* \cong \operatorname{im} \varepsilon_* = \mathbb{Z}$, which is torsion-free). Exactness of the sequence of homology groups ($\partial$ is surjective if and only if $\varepsilon_*$ is injective) now gives $\partial$ a surjection. We conclude that

$$\partial\colon H_1(G, \mathbb{Z}) \cong H_0(G, \mathcal{G}) \cong \mathcal{G}/\mathcal{G}^2,$$

by Proposition C-3.106. •

**Proposition C-3.112.** *For any group $G$, we have*

$$H_1(G, \mathbb{Z}) \cong G/G',$$

*where $G'$ is the commutator subgroup of $G$.*

**Proof.** It suffices to prove that $G/G' \cong \mathcal{G}/\mathcal{G}^2$. Define $\theta\colon G \to \mathcal{G}/\mathcal{G}^2$ by

$$\theta\colon x \mapsto (x-1) + \mathcal{G}^2.$$

To see that $\theta$ is a homomorphism, note that

$$xy - 1 - (x-1) - (y-1) = (x-1)(y-1) \in \mathcal{G}^2,$$

so that

$$\begin{aligned}\theta(xy) &= xy - 1 + \mathcal{G}^2 \\ &= (x-1) + (y-1) + \mathcal{G}^2 \\ &= x - 1 + \mathcal{G}^2 + y - 1 + \mathcal{G}^2 \\ &= \theta(x) + \theta(y).\end{aligned}$$

Since $\mathcal{G}/\mathcal{G}^2$ is abelian, $G' \subseteq \ker \theta$, and so $\theta$ induces a homomorphism $\theta'\colon G/G' \to \mathcal{G}/\mathcal{G}^2$; namely, $xG' \mapsto x - 1 + \mathcal{G}^2$.

We now construct the inverse of $\theta'$. By Proposition C-3.105, $\mathcal{G}$ is a free abelian group with basis all $x - 1$, where $x \in G$ and $x \neq 1$. It follows that there is a (well-defined) homomorphism $\varphi\colon \mathcal{G} \to G/G'$, given by

$$\varphi\colon x - 1 \mapsto xG'.$$

If $\mathcal{G}^2 \subseteq \ker \varphi$, then $\varphi$ induces a homomorphism $\mathcal{G}/\mathcal{G}^2 \to G/G'$ that, obviously, is the inverse of $\theta'$, and this will complete the proof.

If $u \in \mathcal{G}^2$, then

$$\begin{aligned}u &= \Big(\sum_{x\neq 1} m_x(x-1)\Big)\Big(\sum_{y\neq 1} n_y(y-1)\Big) \\ &= \sum_{x,y} m_x n_y (x-1)(y-1) \\ &= \sum_{x,y} m_x n_y \big((xy-1) - (x-1) - (y-1)\big).\end{aligned}$$

Therefore, $\varphi(u) = \prod_{x,y}(xyx^{-1}y^{-1})^{m_x n_y} G' = G'$, and so $u \in \ker \varphi$, as desired. •

Since $H_1(\ ,\mathbb{Z})$ is a contravariant functor, every group homomorphism $f\colon S \to G$ induces a map $f_*\colon H_1(G,\mathbb{Z}) \to H_1(S,\mathbb{Z})$; that is, $f_*\colon G/G' \to S/S'$. If $S \subseteq G$ is a subgroup of finite index with inclusion $i\colon S \to G$, there is a well-known homomorphism $V_{G\to S}\colon G/G' \to S/S'$, called the ***transfer***, and Eckmann proved that $i_* = V_{G\to S}$ (see Rotman [**187**], p. 578).

The group $H_2(G,\mathbb{Z})$ is useful; it is called the ***Schur multiplier*** of $G$. For example, suppose that $G = F/R$, where $F$ is a free group; that is, we have a presentation of a group $G$. Then ***Hopf's formula*** is

$$H_2(G,\mathbb{Z}) \cong (R \cap F)/[F,R]$$

(see Rotman [**187**], p. 274). It follows that the group $(R \cap F)/[F,R]$ depends only on $G$ and not upon the choice of presentation of $G$.

**Definition.** An exact sequence $0 \to A \to E \to G \to 1$ is a ***central extension*** of a group $G$ if $A \subseteq Z(E)$. A ***universal central extension*** of $G$ is a central extension $0 \to M \to U \to G \to 1$ for which there always exists a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & G & \longrightarrow & 1 \\ & & \vdots & & \vdots & & \downarrow{\scriptstyle 1_G} & & \\ 0 & \longrightarrow & M & \longrightarrow & U & \longrightarrow & G & \longrightarrow & 1. \end{array}$$

**Theorem C-3.113.** *If $G$ is a finite group, then $G$ has a universal central extension if and only if $G = G'$, in which case $M \cong H_2(G,\mathbb{Z})$. In particular, every finite simple group has a universal central extension.*

**Proof.** See Milnor [**156**], pp. 43–46. •

This theorem is used to construct "covers" of simple groups.

If $G$ is a finitely presented group, say, with presentation

$$G = (x_1, \ldots, x_n \mid y_1, \ldots, y_r),$$

then $n - r \leq \operatorname{rank}(G/G') + d(H_2(G,\mathbb{Z}))$, where $d(H_2(G,\mathbb{Z}))$ is the minimal number of generators of $H_2(G,\mathbb{Z})$ (Rotman [**187**], p. 551). This result implies that $n \leq r$; that is, a finitely presented group usually has more relations than generators.

There is an explicit upper bound on the order of the Schur multiplier of a finite $p$-group.

**Theorem C-3.114 (Green).** *If $G$ is a finite $p$-group of order $p^m$, then*

$$|H_2(G,\mathbb{Z})| \leq p^{m(m-1)/2}.$$

**Proof.** Rotman [**187**], p. 663. •

We now consider cohomology groups.

**Proposition C-3.115.** *Let $G$ be a group, let $A$ be a $G$-module, and let $\mathbb{Z}$ be viewed as a trivial $G$-module. Then*

$$H^0(G,A) = \operatorname{Hom}_G(\mathbb{Z},A) \cong A^G.$$

**Proof.** Recall that $A^G = \{a \in A : ga = a \text{ for all } g \in G\}$. By definition,

$$H^0(G, A) = \operatorname{Ext}^0_{\mathbb{Z}G}(\mathbb{Z}, A) = \operatorname{Hom}_G(\mathbb{Z}, A).$$

Define $\tau_A \colon \operatorname{Hom}_G(\mathbb{Z}, A) \to A^G$ by $f \mapsto f(1)$. Note that $f(1) \in A^G$: if $g \in G$, then $gf(1) = f(g \cdot 1)$ (because $f$ is a $G$-map), and $g \cdot 1 = 1$ (because $\mathbb{Z}$ is $G$-trivial); therefore, $gf(1) = f(1)$, and $f(1) \in A^G$. That $\tau_A$ is an isomorphism is a routine calculation. •

It follows that $H^0(G, A)$ is the largest $G$-trivial submodule of $A$.

**Theorem C-3.116.** *Let $G = \langle \sigma \rangle$ be a cyclic group of finite order $k$, and let $A$ be a $G$-module. If $N = \sum_{i=0}^{k-1} \sigma^i$ and $D = \sigma - 1$, then*

$$H^0(G, A) = A^G,$$
$$H^{2n-1}(G, A) = \ker N/(\sigma - 1)A \quad \textit{for all } n \geq 1,$$
$$H^{2n}(G, A) = A^G/NA \quad \textit{for all } n \geq 1.$$

**Proof.** Apply the contravariant $\operatorname{Hom}_G(\ \ , A)$ to the resolution of $\mathbb{Z}$ in Lemma C-3.108, noting that $\operatorname{Hom}_G(\mathbb{Z}G, A) \cong A$. The calculation of ker/im is now as given in the statement. •

Note that Proposition C-3.105 gives $\operatorname{im} D = \mathcal{G}A$.

**Corollary C-3.117.** *If $G$ is a cyclic group of finite order $k$ and $A$ is a trivial $G$-module, then*

$$H^0(G, A) = A,$$
$$H^{2n-1}(G, A) = A[k] \quad \textit{for all } n \geq 1,$$
$$H^{2n}(G, A) = A/kA \quad \textit{for all } n \geq 1.$$

*In particular,*

$$H^0(G, \mathbb{Z}) = \mathbb{Z},$$
$$H^{2n-1}(G, \mathbb{Z}) = \{0\} \quad \textit{for all } n \geq 1,$$
$$H^{2n}(G, \mathbb{Z}) = \mathbb{Z}/k\mathbb{Z} \quad \textit{for all } n \geq 1.$$

**Remark.** A finite group $G$ for which there exists a nonzero integer $d$ such that

$$H^n(G, A) \cong H^{n+d}(G, A),$$

for all $n \geq 1$ and all $G$-modules $A$, is said to have ***periodic cohomology***. Corollary C-3.117 shows that all finite cyclic groups have periodic cohomology, and it can be proved that a group $G$ has periodic cohomology if and only if its Sylow $p$-subgroups are cyclic, for all odd primes $p$, while its Sylow 2-subgroups are either cyclic or generalized quaternion (see Adem–Milgram [**1**], p. 148). For example, $G = \mathrm{SL}(2, 5)$ has periodic cohomology: it is a group of order $120 = 8 \cdot 3 \cdot 5$, so its Sylow 3-subgroups and its Sylow 5-subgroups are cyclic, having prime order, while we saw, in Exercise C-1.39 on page 32, that its Sylow 2-subgroups are isomorphic to the quaternions. ◄

We can interpret $H^1(G, A)$ and $H^2(G, A)$, where $G$ is any not necessarily cyclic group, in terms of derivations and extensions if we can show that the formulas in Section C-3.3 do, in fact, arise from a projective resolution of $\mathbb{Z}$. Here is a technical interlude.

**Definition.** If $G$ is a group, define $B_0(G)$ to be the free $G$-module on the single generator $[\ ]$ (hence, $B_0(G) \cong \mathbb{Z}G$) and, for $n \geq 1$, define $B_n(G)$ to be the free $G$-module with basis all symbols $[x_1 \mid x_2 \mid \cdots \mid x_n]$, where $x_i \in G$. Define $\varepsilon\colon B_0(G) \to \mathbb{Z}$ by $\varepsilon([\ ]) = 1$ and, for $n \geq 1$, define $d_n\colon B_n(G) \to B_{n-1}(G)$ by

$$\begin{aligned} d_n\colon [x_1 \mid \cdots \mid x_n] \mapsto\ & x_1[x_2 \mid \cdots \mid x_n] \\ & + \sum_{i=1}^{n-1} (-1)^i [x_1 \mid \cdots \mid x_i x_{i+1} \mid \cdots \mid x_n] \\ & + (-1)^n [x_1 \mid \cdots \mid x_{n-1}]. \end{aligned}$$

The ***bar resolution*** is the sequence

$$\mathbf{B}_\bullet(G) : \cdots \to B_2(G) \xrightarrow{d_2} B_1(G) \xrightarrow{d_1} B_0(G) \xrightarrow{\varepsilon} \mathbb{Z} \to 0.$$

Let us look at the low-dimensional part of the bar resolution,

$$\begin{aligned} &d_1\colon [x] \mapsto x[\ ] - [\ ], \\ &d_2\colon [x \mid y] \mapsto x[y] - [xy] + [x], \\ &d_3\colon [x \mid y \mid z] \mapsto x[y \mid z] - [xy \mid z] + [x \mid yz] - [x \mid y]. \end{aligned}$$

These are the formulas that arose in Section C-3.3 (see page 251), but without the added conditions $[x \mid 1] = 0 = [1 \mid y]$ and $[1] = 0$. In fact, there is another bar resolution, the *normalized bar resolution*, which we will soon define.

The bar resolution is a free resolution of $\mathbb{Z}$, although it is not a routine calculation to see this; calling it a resolution doesn't make it so. We prove $\mathbf{B}_\bullet(G)$ is a resolution by comparing it to a resolution familiar to algebraic topologists.

**Definition.** If $G$ is a group, let $P_n(G)$ be the free abelian group with basis all $(n+1)$-tuples of elements of $G$. Make $P_n(G)$ into a $G$-module by defining

$$x(x_0, x_1, \ldots, x_n) = (xx_0, xx_1, \ldots, xx_n).$$

Define $\partial_n\colon P_n(G) \to P_{n-1}(G)$, whenever $n \geq 1$, by

$$\partial_n\colon (x_0, x_1, \ldots, x_n) \mapsto \sum_{i=0}^{n} (-1)^i (x_0, \ldots, \widehat{x}_i, \ldots, x_n),$$

where $\widehat{x}_i$ means that $x_i$ has been deleted. $\mathbf{P}_\bullet(G)$ is called the ***homogenous resolution*** of $\mathbb{Z}$.

Note that $P_0(G)$ is the free abelian group with basis all $(y)$, for $y \in G$, made into a $G$-module by $x(y) = (xy)$. In other words, $P_0(G) = \mathbb{Z}G$.

The proof that $\mathbf{P}_\bullet(G)$ is a projective resolution of $\mathbb{Z}$ will be broken into two parts.

**Lemma C-3.118.** *The sequence*

$$\mathbf{P}_\bullet(G) : \cdots \to P_2(G) \xrightarrow{\partial_2} P_1(G) \xrightarrow{\partial_1} P_0(G) \xrightarrow{\varepsilon} \mathbb{Z} \to 0,$$

*where $\varepsilon$ is the augmentation, is a complex.*

**Proof.** It suffices to prove that $\partial_{n-1}\partial_n(x_0, x_1, \ldots, x_n) = 0$. Now

$$\begin{aligned}\partial_{n-1}\partial_n(x_0, x_1, \ldots, x_n) &= \sum_{i=0}^{n}(-1)^i\partial_{n-1}(x_0, \ldots, \widehat{x}_i, \ldots, x_n)\\ &= \sum_{i=0}^{n}(-1)^i\Big(\sum_{j<i}(-1)^j(x_0, \ldots, \widehat{x}_j, \ldots, \widehat{x}_i, \ldots, x_n)\Big)\\ &\quad + \sum_{i=0}^{n}(-1)^i\Big(\sum_{j>i}(-1)^{j-1}(x_0, \ldots, \widehat{x}_i, \ldots, \widehat{x}_j, \ldots, x_n)\Big).\end{aligned}$$

In the last equation, the first summation has inner sign $(-1)^j$, because $j < i$, and so $x_j$ is still in the $j$th position after the deletion of $x_i$ from the original $n$-tuple. In the second summation, however, the inner sign is $(-1)^{j-1}$, because $i < j$, and so $x_j$ is in position $j-1$ after deletion of the earlier $x_i$. Thus, $\partial_{n-1}\partial_n(x_0, x_1, \ldots, x_n)$ is a sum of $(n-2)$-tuples $(x_0, \ldots, \widehat{x}_i, \ldots, \widehat{x}_j, \ldots, x_n)$ with $i < j$, each of which occurs twice: once upon deleting $x_i$ by $\partial_n$ and then deleting $x_j$ by $\partial_{n-1}$; a second time upon deleting $x_j$ by $\partial_n$ and then deleting $x_i$ by $\partial_{n-1}$. In the first case, the sign of the $(n-2)$-tuple is $(-1)^{i+j-1}$; in the second case, its sign is $(-1)^{i+j}$. Therefore, the $(n-2)$-tuples cancel in pairs, and $\partial_{n-1}\partial_n = 0$. •

**Proposition C-3.119.** *The complex*

$$\mathbf{P}_\bullet(G) : \cdots \to P_2(G) \xrightarrow{\partial_2} P_1(G) \xrightarrow{\partial_1} P_0(G) \xrightarrow{\partial_0} \mathbb{Z} \to 0,$$

*where $\partial_0 = \varepsilon$ is the augmentation, is a $G$-free resolution of $\mathbb{Z}$.*

**Proof.** We let the reader prove that $P_n(G)$ is a free $G$-module with basis all symbols of the form $(1, x_1, \ldots, x_n)$.

To prove exactness of $\mathbf{P}_\bullet(G)$, it suffices, by Proposition C-3.40, to construct a contracting homotopy, that is, maps

$$\cdots \leftarrow P_2(G) \xleftarrow{s_1} P_1(G) \xleftarrow{s_0} P_0(G) \xleftarrow{s_{-1}} \mathbb{Z}$$

with $\varepsilon s_{-1} = 1_{\mathbb{Z}}$ and

$$\partial_{n+1}s_n + s_{n-1}\partial_n = 1_{P_n(G)}, \quad \text{for all } n \geq 0.$$

Define $s_{-1}\colon \mathbb{Z} \to P_0(G)$ by $m \mapsto m(1)$, where the 1 in the parentheses is the identity element of the group $G$, and define $s_n\colon P_n(G) \to P_{n+1}(G)$, for $n \geq 0$, by

$$s_n\colon (x_0, x_1, \ldots, x_n) \mapsto (1, x_0, x_1, \ldots, x_n).$$

These maps $s_n$ are only $\mathbb{Z}$-maps, but Exercise C-3.31 on page 270 says that this suffices to prove exactness. Here are the computations:

$$\varepsilon s_{-1}(1) = \varepsilon\big((1)\big) = 1.$$

If $n \geq 0$, then

$$\begin{aligned}\partial_{n+1}s_n(x_0,\dots,x_n) &= \partial_{n+1}(1,x_0,\dots,x_n)\\ &= (x_0,\dots,x_n) + \sum_{i=0}^{n}(-1)^{i+1}(1,x_0,\dots,\widehat{x}_i,\dots,x_n)\end{aligned}$$

(the range of summation has been rewritten because $x_i$ sits in the $(i+1)$st position in $(1,x_0,\dots,x_n)$). On the other hand,

$$\begin{aligned}s_{n-1}\partial_n(x_0,\dots,x_n) &= s_{n-1}\sum_{j=0}^{n}(-1)^j(x_0,\dots,\widehat{x}_j,\dots,x_n)\\ &= \sum_{j=0}^{n}(-1)^j(1,x_0,\dots,\widehat{x}_j,\dots,x_n).\end{aligned}$$

It follows that $\big(\partial_{n+1}s_n + s_{n-1}\partial_n\big)(x_0,\dots,x_n) = (x_0,\dots,x_n)$. •

**Proposition C-3.120.** *The bar resolution* $\mathbf{B}_\bullet(\mathbf{G})$ *is a G-free resolution of* $\mathbb{Z}$.

**Proof.** For each $n \geq 0$, define $\tau_n\colon P_n(G) \to B_n(G)$ by

$$\tau_n\colon (x_0,\dots,x_n) \mapsto x_0[x_0^{-1}x_1 \mid x_1^{-1}x_2 \mid \cdots \mid x_{n-1}^{-1}x_n],$$

and define $\sigma_n\colon B_n(G) \to P_n(G)$ by

$$\sigma_n\colon [x_1 \mid \cdots \mid x_n] \mapsto (1, x_1, x_1x_2, x_1x_2x_3, \dots, x_1x_2\cdots x_n).$$

It is routine to check that $\tau_n$ and $\sigma_n$ are inverse, and so each $\tau_n$ is an isomorphism.

The reader can also check that $\tau\colon \mathbf{P}_\bullet(\mathbf{G}) \to \mathbf{B}_\bullet(\mathbf{G})$ is a chain map; that is, the following diagram commutes:

$$\begin{array}{ccc} P_n(G) & \xrightarrow{\tau_n} & B_n(G) \\ \Big\downarrow{\scriptstyle \partial_n} & & \Big\downarrow{\scriptstyle d_n} \\ P_{n-1}(G) & \xrightarrow{\tau_{n-1}} & B_{n-1}(G). \end{array}$$

Finally, Exercise C-3.24 on page 269 shows that both complexes have the same homology groups. By Proposition C-3.119, the complex $\mathbf{P}_\bullet(G)$ is an exact sequence, so that all its homology groups are $\{0\}$. It follows that all the homology groups of $\mathbf{B}_\bullet(G)$ are $\{0\}$, and so it, too, is an exact sequence. •

**Definition.** Define

$$[x_1 \mid \cdots \mid x_n]^* = \begin{cases} [x_1 \mid \cdots \mid x_n] & \text{if all } x_i \neq 1,\\ 0 & \text{if some } x_i = 1.\end{cases}$$

The ***normalized bar resolution***, $\mathbf{B}_\bullet^*(\mathbf{G})$, is the sequence

$$\mathbf{B}_\bullet^*(\mathbf{G})\colon \cdots \to B_2^*(G) \xrightarrow{d_2} B_1^*(G) \xrightarrow{d_1} B_0^*(G) \xrightarrow{\varepsilon} \mathbb{Z} \to 0,$$

where $B_n^*(G)$ is the free $G$-module with basis all nonzero $[x_1 \mid \cdots \mid x_n]^*$, and the maps $d_n$ have the same formula as the maps in the bar resolution except that all symbols $[x_1|\cdots|x_n]$ now occur as $[x_1|\cdots|x_n]^*$; in particular, $[x_1|\cdots|x_n]^* = 0$ if some $x_i = 1$.

Since we are making some of the basis elements 0, it is not obvious that the normalized bar resolution $\mathbf{B}_\bullet^*(\mathbf{G})$ is a complex, let alone a resolution of $\mathbb{Z}$.

**Theorem C-3.121.** *The normalized bar resolution $\mathbf{B}_\bullet^*(\mathbf{G})$ is a $G$-free resolution of $\mathbb{Z}$.*

**Proof.** We begin by constructing a contracting homotopy

$$\cdots \leftarrow B_2^*(G) \xleftarrow{t_1} B_1^*(G) \xleftarrow{t_0} B_0^*(G) \xleftarrow{t_{-1}} \mathbb{Z},$$

where each $t_n$ is a $\mathbb{Z}$-map. Define $t_{-1}\colon \mathbb{Z} \to B_0^*(G)$ by $t_{-1}\colon m \mapsto m[\,]$. Note that $B_n^*(G)$ is a free $G$-module with basis all nonzero $[x_1 \mid \cdots \mid x_n]^*$; hence, it is a direct sum of copies of $\mathbb{Z}G$. Since $\mathbb{Z}G$ is a free abelian group, $B_n^*(G)$ is also a free abelian group; the reader may check that a basis of $B_n^*(G)$, as a free abelian group, consists of all nonzero $x[x_1 \mid \cdots \mid x_n]^*$. To define $t_n$ for $n \geq 0$, we take advantage of the fact that $t_n$ need only be a $\mathbb{Z}$-map, by giving its values on these $\mathbb{Z}$-basis elements (and freeness allows us to choose these values without restriction). Thus, for $n \geq 0$, define $t_n\colon B_n^*(G) \to B_{n+1}^*(G)$ by

$$t_n\colon x[x_1 \mid \cdots \mid x_n]^* \mapsto [x \mid x_1 \mid \cdots \mid x_n]^*.$$

That we have constructed a contracting homotopy is routine; the reader may check that $\varepsilon t_{-1} = 1_{\mathbb{Z}}$ and, for $n \geq 0$, that

$$d_{n+1}t_n + t_{n-1}d_n = 1_{B_n^*(G)}.$$

The proof will be complete once we show that $\mathbf{B}_\bullet^*(\mathbf{G})$ is a complex. Since $B_{n+1}^*(G)$ is generated, as a $G$-module, by $\operatorname{im} t_n$, it suffices to show that $d_n d_{n+1} = 0$ on this subgroup. We now prove, by induction on $n \geq -1$, that $d_n d_{n+1} t_n = 0$. The base step is true, for $\varepsilon = t_{-1}$ and $0 = \varepsilon d_1 = t_{-1}d_1$. For the inductive step, we use the identities in the definition of contracting homotopy and the inductive hypothesis $d_{n-1}d_n = 0$:

$$\begin{aligned} d_n d_{n+1} t_n &= d_n(1 - t_{n-1}d_n) \\ &= d_n - d_n t_{n-1} d_n \\ &= d_n - (1 - t_{n-2}d_{n-1})d_n \\ &= d_n - d_n - t_{n-2}d_{n-1}d_n \\ &= 0. \quad \bullet \end{aligned}$$

We can now interpret $H^1(G, A)$ and $H^2(G, A)$.

**Corollary C-3.122.** *For every group $G$ and every $G$-module $A$, the groups $H^1(G, A)$ and $H^2(G, A)$ constructed in Section* C-3.3 *coincide with the cohomology groups.*

**Proof.** We have proved that factor sets, coboundaries, derivations, and principal derivations do, in fact, arise from a projective resolution of $\mathbb{Z}$. $\bullet$

**Proposition C-3.123.** *If $G$ is a finite group of order $m$, then $mH^n(G, A) = \{0\}$ for all $n \geq 1$ and all $G$-modules $A$.*

**Proof.** For $f \in \mathrm{Hom}_G(B_n, A)$, define $\sigma_f(x_1|\dots|x_{n-1}) = \sum_{x\in G} f(x_1|\dots|x_{n-1}|x)$. As in the proof of Theorem C-3.21, sum the cocycle formula to obtain $df = d\sigma_f + m(-1)^n f$; thus, if $f$ is a cocycle, then $mf$ is a coboundary. •

**Corollary C-3.124.** *If $G$ is a finite group and $A$ is a finitely generated $G$-module, then $H^n(G, A)$ is finite for all $n \geq 0$.*

**Proof.** $H^n(G, A)$ is a finitely generated abelian group (because $A$ is finitely generated) of finite exponent. •

Both Proposition C-3.123 and its corollary are true for homology groups as well.

There are several aspects of the cohomology of groups that we have not mentioned. Aside from being a useful tool within group theory itself, these groups also form a link with algebraic topology. For every group $G$, there exists a topological space $K(G, 1)$, called an ***Eilenberg–Mac Lane space***, whose fundamental group is $G$ and whose cohomology groups coincide with the algebraically defined cohomology groups.[18] There is, in fact, a deep connection between group theory and algebraic topology, of which this is a first sign.

An important property of the cohomology of groups is the relation between the cohomology of a group and the cohomology of its subgroups and its quotient groups. If $\varphi\colon S \to G$ is a homomorphism, every $G$-module $A$ becomes an $S$-module if we define $sa = \varphi(s)a$ for all $s \in S$ and $a \in A$. What is the connection between $H^n(S, A)$ and $H^n(G, A)$? What is the connection between $H_n(S, A)$ and $H_n(G, A)$? (There is also a connection between homology groups and cohomology groups: $H^n(G, A)^* \cong H_n(G, A^*)$, where $A^* = \mathrm{Hom}_{\mathbb{Z}}(A, \mathbb{Q}/\mathbb{Z})$.)

There are three standard maps, which we will define in terms of the bar resolution. The first is ***restriction***. If $S$ is a subgroup of a group $G$, every function $f\colon B_n(G) \to A$ is defined on all $[x_1 \mid \cdots \mid x_n]$ with $x_i \in G$, and so it is obviously defined on all $n$-tuples of the form $[s_1 \mid \cdots \mid s_n]$ with $s_i \in S \subseteq G$; thus, its restriction, denoted by $f|S$, maps $B_n(S) \to A$. If $f$ is an $n$-cocycle, denote its cohomology class by

$$\mathrm{cls}\, f = f + \mathrm{im}\, d^*_{n+1}.$$

Define

$$\mathrm{Res}^n\colon H^n(G, A) \to H^n(S, A)$$

by $\mathrm{Res}^n(\mathrm{cls}\, f) = \mathrm{cls}(f|S)$. One result is that if $G$ is finite, $S_p$ is a Sylow $p$-subgroup, and $n \geq 1$, then $\mathrm{Res}^n\colon H^n(G, A) \to H^n(S_p, A)$ is injective on the $p$-primary component of $H^n(G, A)$ (Rotman [**187**], Corollary 9.90). Thus, the cohomology of $G$ is strongly influenced by the cohomology of its Sylow subgroups.

If $S \subseteq G$, there is a map

$$\mathrm{Cor}^n\colon H^n(S, A) \to H^n(G, A)$$

in the reverse direction, called ***corestriction***, which is defined when $S$ has finite index in $G$. First, define $\mathrm{Cor}^0\colon A^S \to A^G$ by $a \mapsto \sum_{t\in T} ta$, where $T$ is a left

[18] Because of this topological connection, many authors use the notation $H^n(\pi, \mathbb{Z})$ to denote cohomology groups, for $\pi_1$ is the standard notation for the fundamental group.

transversal of $S$ in $G$ (of course, we must check that $\mathrm{Cor}^0$ is a homomorphism that is independent of the choice of transversal). There is a standard way of extending a map in dimension 0 to maps in higher dimensions (essentially by dimension shifting), and if $[G : S] = m$, then

$$\mathrm{Cor}^n \circ \mathrm{Res}^n \colon H^n(G, A) \to H^n(G, A) = \mu_m;$$

that is, the composite is multiplication by $m$. Similarly, in homology, the map

$$\mathrm{Cor}_0 \colon A/\mathcal{G}A \to A/\mathcal{S}A,$$

defined by $a+\mathcal{G}A \mapsto \sum_{t\in T} t^{-1}a+\mathcal{S}A$, extends to maps in higher dimensions. When $n = 1$ and $A = \mathbb{Z}$, we have $\mathrm{Cor}_1 \colon H_1(G, \mathbb{Z}) \to H_1(S, \mathbb{Z})$; that is, $\mathrm{Cor}_1 \colon G/G' \to S/S'$, which turns out to be the transfer: $\mathrm{Cor}_1 = V_{G\to S}$ (see Rotman [**187**], p. 578).

The third standard map is called ***inflation***. Suppose that $N$ is a normal subgroup of a group $G$. If $A$ is a $G$-module, then $A^N$ is a $G/N$-module if we define $(gN)a = ga$ for $a \in A^N$ (if $gN = hN$, then $h = gx$ for some $x \in N$, and so $ha = (gx)a = g(xa) = ga$, because $xa = a$). Define

$$\mathrm{Inf}^n \colon H^n(G/N, A^N) \to H^n(G, A)$$

by $\mathrm{cls}\, f \mapsto \mathrm{cls}(f^{\#})$, where

$$f^{\#} \colon [g_1 \mid \cdots \mid g_m] \mapsto f[g_1N \mid \cdots \mid g_mN].$$

These fit together in the *Five Term Exact Sequence* as follows:

**Theorem C-3.125.** *Let $S$ be a normal subgroup of a group $G$, and let $A$ be a $G$-module. There is an exact sequence*

$$0 \to H^1(G/S, A^S) \xrightarrow{\mathrm{Inf}^1} H^1(G, A) \xrightarrow{\mathrm{Res}^1} H^1(S, A).$$

**Remark.** Using spectral sequences, the exact sequence can be lengthened to a five term sequence ending with

$$\to H^1(S, A)^{G/S} \xrightarrow{d} H^2(G/S, A^S) \xrightarrow{\mathrm{Inf}^2} H^2(G, A).$$

Using conjugation, the cohomology groups $H^n(G, A)$ can be made into $G/S$-modules (Rotman [**187**], p. 567), so that $H^1(S, A)^{G/S}$ makes sense and $\mathrm{im}\,\mathrm{Res}^1 \subseteq H^1(S, A)^{G/S}$. The map $d \colon H^1(S, A)^G \to H^2(G/S, A^S)$ is called the ***transgression*** (Mac Lane [**145**], pp. 332–335). ◀

**Sketch of Proof.** We use the interpretation of $H^1$ as Der/PDer (formally, we are using the normalized bar resolution).

(i) $\mathrm{Inf}^1$ *is an injection.* If $\zeta \colon G/S \to A^S$ is a derivation and $\dot{\zeta} \colon G \to A$ is the function obtained from $\zeta$ by making it constant on cosets of $S$, then $\mathrm{Inf}^1 \colon \mathrm{cls}(\zeta) \mapsto \mathrm{cls}(\dot{\zeta})$. If $\mathrm{cls}(\zeta) \in \ker \mathrm{Inf}^1$, then $\dot{\zeta}$ is a principal derivation; that is, there is $a \in A$ with $\dot{\zeta}(x) = xa - a$ for all $x \in G$. Since $\dot{\zeta}$ is constant on cosets of $S$, we have $xsa - a = xa - a$ for all $s \in S$ and $x \in G$; hence, $xsa = xa$. If $x = 1$, then $sa = a$; that is, $a \in A^S$. It follows that $\zeta$ is a principal derivation, and $\mathrm{cls}(\zeta) = 0$.

(ii) $\operatorname{im}\operatorname{Inf}^1 \subseteq \ker \operatorname{Res}^1$. If $Z\colon G \to A$ is a derivation, then $\operatorname{Res}^1 \operatorname{cls}(Z) = \operatorname{cls}(Z|S)$. In particular, if $\operatorname{cls}(Z) = \operatorname{cls}(\dot{\zeta}) \in \operatorname{im}\operatorname{Inf}^1$, then $Z$ is constant on the coset $S$. But $Z(1) = 0$ (because every derivation sends 1 to 0), and so $Z = 0$.

(iii) $\ker \operatorname{Res}^1 \subseteq \operatorname{im}\operatorname{Inf}^1$. If $\operatorname{cls}(Z) \in \ker \operatorname{Res}^1$, then $\operatorname{cls}(Z|S) = 0$; that is, $Z|S$ is a principal derivation. Thus, there is $a \in A$ with $Z(s) = sa - a$ for all $s \in S$. Replacing $Z$ by $Z - \delta$, where $\delta$ is the principal derivation $g \mapsto ga - a$, we may assume that $Z(s) = 0$ for all $s \in S$. But $Z$ is constant on cosets of $S$, for the definition of derivation gives $Z(gs) = gZ(s) + Z(g) = Z(g)$. Now $\zeta\colon G/S \to A$, defined by $\zeta(gS) = Z(g)$, is a well-defined derivation, and $Z = \dot{\zeta} \in \operatorname{im}\operatorname{Inf}^1$. •

There is a similar discussion for group homology.

We can force cohomology groups to be a graded ring by defining ***cup product*** on $H^\bullet(G, R) = \sum_{n\geq 0} H^n(G, R)$, where $R$ is any commutative ring (see Cassels–Fröhlich [**38**], pp. 105–108, or Evens [**62**]), and this added structure has important applications.

We now consider the cohomology of free groups.

**Lemma C-3.126.** *If $G$ is a free group with basis $X$, then its augmentation ideal $\mathcal{G}$ is a free $G$-module with basis*

$$X - 1 = \{x - 1 : x \in X\}.$$

**Proof.** We show first that $\mathcal{G}$ is generated by $X - 1$. The identities

$$xy - 1 = (x - 1) + x(y - 1)$$

and

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

show that if $w$ is any word in $X$, then $w-1$ can be written as a $G$-linear combination of $X - 1$.

To show that $\mathcal{G}$ is a free $G$-module with basis $X - 1$, it suffices, by Proposition B-4.36 in Part 1, to show that the following diagram can be completed:

$$\begin{array}{ccc} \mathcal{G} & & \\ {\scriptstyle i}\uparrow & \searrow^{\Phi} & \\ X-1 & \xrightarrow[\varphi]{} & A, \end{array}$$

where $A$ is any $G$-module, $i\colon X - 1 \to \mathcal{G}$ is the inclusion, and $\varphi$ is any function (uniqueness of such a map $\Phi$ follows from $X - 1$ generating $\mathcal{G}$). Thus, we are seeking $\Phi \in \operatorname{Hom}_G(\mathcal{G}, A)$. By Exercise C-3.58 on page 325, we have $\operatorname{Hom}_G(\mathcal{G}, A) \cong \operatorname{Der}(G, A)$ via $f\colon x \mapsto f(x - 1)$, where $f \in \mathcal{G} \to A$, and so we seek a derivation.

Consider the (necessarily split) extension $0 \to A \to E \to G \to 1$, so that $E$ consists of all ordered pairs $(a, g) \in A \times G$. The given function $\varphi\colon X - 1 \to A$ defines a lifting $\ell$ of the generating set $X$ of $G$; namely,

$$\ell(x) = (\varphi(x - 1), x).$$

Since $G$ is free with basis $X$, the function $\ell\colon X \to E$ extends to a homomorphism $L\colon G \to E$. We claim, for every $g \in G$, that $L(g) = (d(g), g)$, where $d\colon G \to A$.

Each $g \in G$ has a unique expression as a reduced word $g = x_1^{e_1} \cdots x_n^{e_n}$, where $x_i \in X$ and $e_i = \pm 1$. We prove the claim by induction on $n \geq 1$. The base step is clear, while

$$\begin{aligned} L(g) &= L(x_1^{e_1} \cdots x_n^{e_n}) \\ &= L(x_1^{e_1}) \cdots L(x_n^{e_n}) \\ &= (\varphi(x_1 - 1), x_1)^{e_1} \cdots (\varphi(x_n - 1), x_n)^{e_n} \\ &= (d(g), g), \end{aligned}$$

and so the first coordinate $d(g)$ lies in $A$. Finally, $d$ is a derivation because $L$ is a homomorphism.

Exercise C-3.58 on page 325 now says that there is a homomorphism $\Phi\colon \mathcal{G} \to A$ defined by $\Phi(g-1) = d(g)$ for all $g \in G$. In particular, $\Phi(x-1) = d(x) = \varphi(x-1)$, so that $\Phi$ does extend $\varphi$. •

**Theorem C-3.127.** *If $G$ is a free group, then $H^n(G, A) = \{0\}$ for all $n > 1$ and all $G$-modules $A$.*

**Proof.** The sequence $0 \to \mathcal{G} \to \mathbb{Z}G \to \mathbb{Z} \to 0$ is a free resolution of $\mathbb{Z}$ because $\mathcal{G}$ is now a free $G$-module. Thus, the only nonzero terms in the deleted resolution occur in positions 0 and 1, and so all cohomology groups vanish for $n > 1$. •

We are now going to state an interesting result (the Stallings–Swan Theorem), which was discovered using homological methods but which does not mention homology in its statement.

If $G$ is a group and $S \subseteq G$ is a subgroup, then every $G$-module $A$ can be viewed as an $S$-module, for $\mathbb{Z}S$ is a subring of $\mathbb{Z}G$.

**Definition.** A group $G$ has ***cohomological dimension*** $\leq n$, denoted by

$$\operatorname{cd}(G) \leq n,$$

if $H^{n+1}(S, A) = \{0\}$ for all $G$-modules $A$ and every subgroup $S$ of $G$. We write $\operatorname{cd}(G) = \infty$ if no such integer $n$ exists.

We say that $\operatorname{cd}(G) = n$ if $\operatorname{cd}(G) \leq n$ but it is not true that $\operatorname{cd}(G) \leq n - 1$.

**Example C-3.128.**

(i) If $G = \{1\}$, then $\operatorname{cd}(G) = 0$; this follows from Theorem C-3.116 because $G$ is a cyclic group of order 1.

(ii) If $G$ is a finite cyclic group of order $k > 1$, then $\operatorname{cd}(G) = \infty$, as we see from Corollary C-3.117 with $A = \mathbb{Z}$. Suppose that $G$ is an infinite cyclic group. Since every subgroup $S \subseteq G$ is cyclic, Theorem C-3.127 gives $\operatorname{cd}(G) \leq 1$. If $\operatorname{cd}(G) = 0$, then $H^1(S, A) = \{0\}$ for all subgroups $S$ and all modules $A$. In particular, if $S \cong \mathbb{Z}$ and $A \neq \{0\}$ is a trivial module, then $H^1(S, A) = \operatorname{Der}(S, A)/\operatorname{PDer}(S, A) \cong \operatorname{Hom}(\mathbb{Z}, A) \neq \{0\}$. Hence, $\operatorname{cd}(G) = 1$.

(iii) If $G \neq \{1\}$ is a free group, then Theorem C-3.127 shows that $\operatorname{cd}(G) = 1$, for every subgroup of a free group is free.

(iv) If $\mathrm{cd}(G) < \infty$, then $G$ must be torsion-free; otherwise, $G$ has a subgroup $S$ that is cyclic of finite order $k > 1$, and $H^n(S, \mathbb{Z}) \neq 0$ for all even $n$.

(v) It is known that if $G$ is a free abelian group of finite rank $n$, then $\mathrm{cd}(G) = n$. ◀

**Proposition C-3.129 (Shapiro's Lemma).** *Let $G$ be a group and let $S \subseteq G$ be a subgroup. If $A$ is a $\mathbb{Z}S$-module, then for all $n \geq 0$,*

$$H^n(S, A) \cong H^n(G, \mathrm{Hom}_{\mathbb{Z}S}(\mathbb{Z}G, A)).$$

**Proof.** Let $\mathbf{P}_\bullet = \cdots \to P_1 \to P_0 \to \mathbb{Z} \to 0$ be a $\mathbb{Z}G$-free resolution. If we denote $\mathrm{Hom}_{\mathbb{Z}S}(\mathbb{Z}G, A)$ by $A^*$, then

$$H^n(G, A^*) = H^n(\mathrm{Hom}_{\mathbb{Z}G}(\mathbf{P}_\bullet, A^*)).$$

By the adjoint isomorphism,

$$\begin{aligned} \mathrm{Hom}_{\mathbb{Z}G}(P_i, A^*) &= \mathrm{Hom}_{\mathbb{Z}G}(P_i, \mathrm{Hom}_{\mathbb{Z}S}(\mathbb{Z}G, A)) \\ &\cong \mathrm{Hom}_{\mathbb{Z}S}(P_i \otimes_{\mathbb{Z}G} \mathbb{Z}G, A) \\ &\cong \mathrm{Hom}_{\mathbb{Z}S}(P_i, A). \end{aligned}$$

But (the proof of) Lemma C-2.84(i) shows that $\mathbb{Z}G$ is a free $\mathbb{Z}S$-module, and so the free $\mathbb{Z}G$-modules $P_i$ are also free $\mathbb{Z}S$-modules. It follows that we may regard $\mathbf{P}_\bullet$ as a $\mathbb{Z}S$-free resolution of $\mathbb{Z}$, and there is an isomorphism of complexes:

$$\mathrm{Hom}_{\mathbb{Z}S}(\mathbf{P}_\bullet, A) \cong \mathrm{Hom}_{\mathbb{Z}G}(\mathbf{P}_\bullet, A^*).$$

Hence, their homology groups are isomorphic; that is, $H^n(S, A) \cong H^n(G, A^*)$. •

**Corollary C-3.130.** *If $G$ is a group and $S \subseteq G$ is a subgroup, then* $\mathrm{cd}(S) \leq \mathrm{cd}(G)$.

**Proof.** We may assume that $\mathrm{cd}(G) = n < \infty$. If $m > n$ and there is a $\mathbb{Z}S$-module $A$ with $H^m(S, A) \neq \{0\}$, then Shapiro's Lemma gives $H^m(G, \mathrm{Hom}_{\mathbb{Z}S}(\mathbb{Z}G, A)) \cong H^m(S, A) \neq \{0\}$, and this contradicts $\mathrm{cd}(G) = n$. •

**Corollary C-3.131.** *A group $G$ of finite cohomological dimension has no elements (other than 1) of finite order.*

**Proof.** Example C-3.128(ii) and the preceding corollary. •

**Corollary C-3.132.** *A group $G = \{1\}$ if and only if* $\mathrm{cd}(G) = 0$.

**Proof.** If $G = \{1\}$, then $\mathrm{cd}(G) = 0$, by Example C-3.128(i). Conversely, if $\mathrm{cd}(G) = 0$, then Corollary C-3.130 gives $\mathrm{cd}(S) = 0$ for every cyclic subgroup $S = \langle g \rangle \subseteq G$. By Theorem C-3.116, we have $\langle g \rangle = \{1\}$ for all $g \in G$, and so $G = \{1\}$. •

Are there groups $G$ with $\mathrm{cd}(G) = 1$ that are not free? In 1968, Stallings [**211**] proved the following nice theorem ($\mathbb{F}_2 G$ denotes the group algebra over $\mathbb{F}_2$).

**Theorem C-3.133.** *If $G$ is a finitely presented group for which $H^1(G, \mathbb{F}_2 G)$ has more than two elements, then $G$ is a nontrivial free product: $G = H * K$, where $H \neq \{1\}$ and $K \neq \{1\}$ (free product is the coproduct in* **Groups**).

As a consequence, he proves the following results. We refer the reader to Cohen [**41**] for proofs.

**Corollary C-3.134.** *If $G$ is a finitely generated group with* $\mathrm{cd}(G) = 1$*, then $G$ is free.*

**Corollary C-3.135.** *If $G$ is a torsion-free finitely generated group having a free subgroup of finite index, then $G$ is free.*

Swan showed that both corollaries remain true if we remove the hypothesis that $G$ be finitely generated.

**Theorem C-3.136 (Stallings–Swan).** *A torsion-free group having a free subgroup of finite index must be free.*

## Exercises

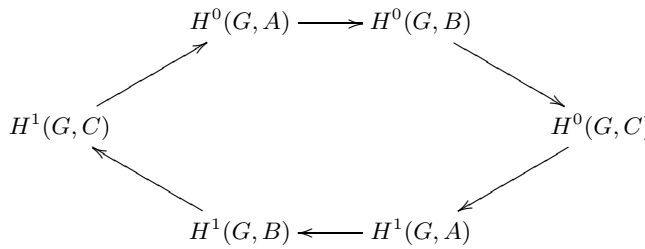
**C-3.57.** (i) Prove that the isomorphisms in Proposition C-3.106 constitute a natural equivalence $\mathbb{Z} \otimes_G -$ to $T$, where $T\colon A \mapsto A/\mathcal{G}A$.

(ii) Prove that the isomorphisms in Proposition C-3.115 constitute a natural equivalence $\mathrm{Hom}_G(\mathbb{Z}, \ )$ to $A \mapsto A^G$.

∗ **C-3.58.** For a fixed group $G$, prove that the functors $\mathrm{Hom}_G(\mathcal{G}, \ )$ and $\mathrm{Der}(G, \ )$ are naturally equivalent.

**Hint.** If $f\colon \mathcal{G} \to A$ is a homomorphism, then $d_f\colon x \mapsto f(x-1)$ is a derivation.

**C-3.59.** (i) If $G$ is a finite cyclic group and $0 \to A \to B \to C \to 0$ is an exact sequence of $G$-modules, prove that there is an ***exact hexagon***; that is, kernel = image at each vertex of the diagram




We remark that this exercise is a key lemma in class field theory.

(ii) If $G$ is a finite cyclic group and $A$ is a $G$-module, define the ***Herbrand quotient*** by

$$h(A) = |H^0(G,A)|/|H^1(G,A)|$$

($h(A)$ is defined only when both $H^0(G,A)$ and $H^1(G,A)$ are finite).

Let $0 \to A \to B \to C \to 0$ be an exact sequence of $G$-modules. Prove that if the Herbrand quotient is defined for two of the modules $A, B, C$, then it is defined for the third one, and

$$h(B) = h(A)h(C).$$

**C-3.60.** If $G$ is a group, prove that

$$P_n(G) \cong \bigotimes_1^{n+1} \mathbb{Z}G,$$

where $P_n(G)$ is the $n$th term in the homogeneous resolution $\mathbf{P}_\bullet(G)$ and

$$\bigotimes_1^{n+1} \mathbb{Z}G = \mathbb{Z}G \otimes_{\mathbb{Z}} \mathbb{Z}G \otimes_{\mathbb{Z}} \cdots \otimes_{\mathbb{Z}} \mathbb{Z}G,$$

the tensor product over $\mathbb{Z}$ of $\mathbb{Z}G$ with itself $n+1$ times.

**C-3.61.** If $G$ is a finite cyclic group, prove, for all $G$-modules $A$ and for all $n \geq 1$, that $H^n(G, A) \cong H_{n+1}(G, A)$.

**C-3.62.** Let $G$ be a group.

(i) Show, for any abelian group $A$, that $A^* = \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}G, A)$ is a left $\mathbb{Z}G$-module. We call $A^*$ a ***coinduced module***.
**Hint.** If $\varphi\colon \mathbb{Z}G \to A$ and $g \in G$, define $g\varphi$ by $x \mapsto g\varphi(g^{-1}x)$.

(ii) For any left $\mathbb{Z}G$-module $B$, prove that $\mathrm{Hom}_{\mathbb{Z}G}(B, A^*) \cong \mathrm{Hom}_{\mathbb{Z}}(B, A)$.
**Hint.** Use the adjoint isomorphism, Theorem B-4.98 in Part 1.

(iii) If $A^*$ is a coinduced module, prove that $H^n(G, A^*) = \{0\}$ for all $n \geq 1$.

**C-3.63.** If $G$ is a group and $A$ is an abelian group, call the $\mathbb{Z}G$-module $A_* = \mathbb{Z}G \otimes_{\mathbb{Z}} A$ an ***induced module***. Prove that $H_n(G, A_*) = \{0\}$ for all $n \geq 1$.

## C-3.10. Crossed Products

Cohomology groups can also be used to study division rings, and this section may be considered as a continuation of the Division Rings section in Chapter C-2, but now with homology available to help. We begin with a return to Galois theory.

**Theorem C-3.137.** *Let $E/k$ be a Galois extension with Galois group $G = \mathrm{Gal}(E/k)$. The multiplicative group $E^\times$ of the field $E$ is a $kG$-module, and*

$$H^1(G, E^\times) = \{0\}.$$

**Proof.** If $c\colon G \to E^\times$ is a 1-cocycle, denote $c(\sigma)$ by $c_\sigma$. In multiplicative notation, the cocycle condition is the identity $\sigma(c_\tau)c_{\sigma\tau}^{-1}c_\sigma = 1$ for all $\sigma, \tau \in G$; that is,

$$\sigma(c_\tau) = c_{\sigma\tau}c_\sigma^{-1}. \tag{1}$$

For $e \in E^\times$, consider

$$b = \sum_{\tau \in G} c_\tau \tau(e).$$

By Independence of Characters, Proposition A-5.38 in Part 1, there is some $e \in E^\times$ with $b \neq 0$. For such an element $e$, we have, using Eq. (1),

$$\begin{aligned}\sigma(b) &= \sum_{\tau\in G} \sigma(c_\tau)\sigma\tau(e)\\ &= \sum_{\tau\in G} c_{\sigma\tau}c_\sigma^{-1}\sigma\tau(e)\\ &= c_\sigma^{-1}\sum_{\tau\in G} c_{\sigma\tau}\sigma\tau(e)\\ &= c_\sigma^{-1}\sum_{\omega\in G} c_\omega\omega(e)\\ &= c_\sigma^{-1}b.\end{aligned}$$

Hence, $c_\sigma = b\sigma(b)^{-1}$, and $c$ is a coboundary. Therefore, $H^1(G, E^\times) = \{0\}$. •

Theorem C-3.137 implies Theorem A-5.59 in Part 1, which describes the elements of norm 1 in a cyclic extension. Recall that if $E/k$ is a finite Galois extension, then the *norm* is the function $N\colon E^\times \to E^\times$ with

$$N\colon e \mapsto \prod_{\sigma\in \mathrm{Gal}(E/k)} \sigma(e).$$

Here is another proof of Theorem A-5.59 in Part 1, using homology.

**Corollary C-3.138 (Hilbert's Theorem 90).** *Let $E/k$ be a Galois extension whose Galois group $G = \mathrm{Gal}(E/k)$ is cyclic, say, with generator $\sigma$. If $u \in E^\times$, then $Nu = 1$ if and only if there is $v \in E^\times$ with*

$$u = \sigma(v)v^{-1}.$$

**Proof.** By Theorem C-3.116, we have $H^1(G, E^\times) = \ker N/\operatorname{im} D$, where $N$ is the norm (remember that $E^\times$ is a multiplicative group) and $Dv = \sigma(v)v^{-1}$. Theorem C-3.137 gives $H^1(G, E^\times) = \{0\}$, so that $\ker N = \operatorname{im} D$. Hence, if $u \in E^\times$, then $Nu = 1$ if and only if there is $v \in E^\times$ with $u = \sigma(v)v^{-1}$. •

Theorem C-3.137 is one of the first results in what is called *Galois cohomology*. Another early result is that $H^n(G, E) = \{0\}$ for all $n \geq 1$, where $E$ (in contrast to $E^\times$) is the additive group of the Galois extension. This result follows easily from the Normal Basis Theorem (Theorem C-5.83), which says that if $E/k$ is a finite Galois extension, then $E \cong kG$ as $k$-algebras, where $G = \mathrm{Gal}(E/k)$.

We are now going to use $H^2(G, E^\times)$ to study division rings.

Only one example of a noncommutative division ring has been given in the text: the quaternions $\mathbb{H}$ (this is an $\mathbb{R}$-algebra) and its $k$-algebra analogs for every subfield $k \subseteq \mathbb{R}$ (actually, another example is given in Exercise C-2.64 on page 222). Hamilton discovered the quaternions in 1843, and Frobenius, in 1880, proved that the only $\mathbb{R}$-division algebras are $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ (see Theorem C-2.125). No other examples of noncommutative division rings were known until *cyclic algebras* were found in the early 1900s, by Wedderburn and by Dickson. In 1932, Albert found an example of a *crossed product algebra* that is not a cyclic algebra, and in 1972,

Amitsur found an example of a noncommutative division ring that is not a crossed product algebra.

Wedderburn proved that every finite division ring is a field (Theorem C-2.31). Are there any division rings of prime characteristic?

We begin with an elementary calculation. Suppose that $V$ is a vector space over a field $E$ having basis $\{u_\sigma : \sigma \in G\}$ for some set $G$, so that each $v \in V$ has a unique expression as a linear combination $v = \sum_\sigma a_\sigma u_\sigma$ for $a_\sigma \in E$. For a function $\mu\colon V \times V \to V$, with $\mu(u_\sigma, u_\tau)$ denoted by $u_\sigma u_\tau$, define ***structure constants*** $g_\alpha^{\sigma,\tau} \in E$ by

$$u_\sigma u_\tau = \sum_{\alpha \in G} g_\alpha^{\sigma,\tau} u_\alpha.$$

To have the associative law, we must have $u_\sigma(u_\tau u_\omega) = (u_\sigma u_\tau)u_\omega$; expanding this equation, the coefficient of each $u_\beta$ is

$$\sum_\alpha g_\alpha^{\sigma,\tau} g_\beta^{\alpha,\omega} = \sum_\gamma g_\gamma^{\tau,\omega} g_\beta^{\sigma,\gamma}.$$

Let us simplify these equations. Let $G$ be a group and suppose that $g_\alpha^{\sigma,\tau} = 0$ unless $\alpha = \sigma\tau$; that is, $u_\sigma u_\tau = f(\sigma,\tau)u_{\sigma\tau}$, where $f(\sigma,\tau) = g_{\sigma\tau}^{\sigma,\tau}$. The function $f\colon G \times G \to E^\times$, given by $f(\sigma,\tau) = g_{\sigma\tau}^{\sigma,\tau}$, satisfies the following equation for all $\sigma, \tau, \omega \in G$:

$$f(\sigma,\tau)f(\sigma\tau,\omega) = f(\tau,\omega)f(\sigma,\tau\omega),$$

an equation reminiscent of the cocycle identity but here written in multiplicative notation. This is why factor sets enter into the next definition.

Let $E/k$ be a Galois extension with $\mathrm{Gal}(E/k) = G$, and let $f\colon G \times G \to E^\times$ be a factor set: in multiplicative notation

$$f(\sigma,1) = 1 = f(1,\tau) \quad \text{for all } \sigma,\tau \in G$$

and, if we denote the action of $\sigma \in G$ on $a \in E^\times$ by $a^\sigma$, then

$$f(\sigma,\tau)f(\sigma\tau,\omega) = f(\tau,\omega)^\sigma f(\sigma,\tau\omega).$$

**Definition.** Given a Galois extension $E/k$ with Galois group $G = \mathrm{Gal}(E/k)$ and a factor set $f\colon G \times G \to E^\times$, define the ***crossed product algebra*** $(E, G, f)$ to be the vector space over $E$ having basis all symbols $\{u_\sigma : \sigma \in G\}$ and multiplication

$$(au_\sigma)(bu_\tau) = ab^\sigma f(\sigma,\tau)u_{\sigma\tau}$$

for all $a, b \in E$. If $G$ is a cyclic group, then the crossed product algebra $(E, G, f)$ is called a ***cyclic algebra***.

Since every element in $(E, G, f)$ has a unique expression of the form $\sum a_\sigma u_\sigma$, the definition of multiplication extends by linearity to all of $(E, G, f)$. We note two special cases:

$$u_\sigma b = b^\sigma u_\sigma,$$
$$u_\sigma u_\tau = f(\sigma,\tau)u_{\sigma\tau}.$$

Recall that if $k$ is a field, then a $k$-algebra $A$ is *central simple* if it is a simple algebra (i.e., no two-sided ideals other than $\{0\}$ and $A$ itself) with center $Z(A) = k$.

**Proposition C-3.139.** *If $E/k$ is a Galois extension with Galois group $G = \mathrm{Gal}(E/k)$ and if $f\colon G \times G \to E^\times$ is a factor set, then $(E, G, f)$ is a central simple $k$-algebra that is split by $E$.*

**Proof.** Denote $(E, G, f)$ by $A$. First, we show that $A$ is a $k$-algebra. To prove that $A$ is associative, it suffices to prove that

$$au_\sigma(bu_\tau cu_\omega) = (au_\sigma bu_\tau)cu_\omega,$$

where $a, b, c \in E$. Using the definition of multiplication,

$$\begin{aligned} au_\sigma(bu_\tau cu_\omega) &= au_\sigma(bc^\tau f(\tau,\omega)u_{\tau\omega}) \\ &= a\big(bc^\tau f(\tau,\omega)\big)^\sigma f(\sigma,\tau\omega)u_{\sigma\tau\omega} \\ &= ab^\sigma c^{\sigma\tau} f(\tau,\omega)^\sigma f(\sigma,\tau\omega)u_{\sigma\tau\omega}. \end{aligned}$$

We also have

$$\begin{aligned} (au_\sigma bu_\tau)cu_\omega &= ab^\sigma f(\sigma,\tau)u_{\sigma\tau}cu_\omega \\ &= ab^\sigma f(\sigma,\tau)c^{\sigma\tau} f(\sigma\tau,\omega)u_{\sigma\tau\omega} \\ &= ab^\sigma c^{\sigma\tau} f(\sigma,\tau)f(\sigma\tau,\omega)u_{\sigma\tau\omega}. \end{aligned}$$

The cocycle identity shows that multiplication in $A$ is associative.

That $u_1$ is the unit in $A$ follows from our assuming that factor sets are normalized:

$$u_1u_\tau = f(1,\tau)u_{1\tau} = u_\tau \quad \text{and} \quad u_\sigma u_1 = f(\sigma,1)u_{\sigma 1} = u_\sigma.$$

We have shown that $A$ is a ring. We claim that $ku_1 = \{au_1 : a \in k\}$ is the center $Z(A)$. If $a \in E$, then $u_\sigma a u_1 = a^\sigma u_\sigma$. If $a \in k = E^G$, then $a^\sigma = a$ for all $\sigma \in G$, and so $k \subseteq Z(A)$. For the reverse inclusion, suppose that $z = \sum_\sigma a_\sigma u_\sigma \in Z(A)$. For any $b \in E$, we have $zbu_1 = bu_1 z$. But

$$zbu_1 = \sum a_\sigma u_\sigma bu_1 = \sum a_\sigma b^\sigma u_\sigma.$$

On the other hand,

$$bu_1 z = \sum ba_\sigma u_\sigma.$$

For every $\sigma \in G$, we have $a_\sigma b^\sigma = ba_\sigma$, so that if $a_\sigma \neq 0$, then $b^\sigma = b$. If $\sigma \neq 1$ and $H = \langle\sigma\rangle$, then $E^H \neq E^{\{1\}} = E$, by Theorem A-5.41 in Part 1, and so there exists $b \in E$ with $b^\sigma \neq b$. We conclude that $z = a_1u_1$. For every $\sigma \in G$, the equation $(a_1u_1)u_\sigma = u_\sigma(a_1u_1)$ gives $a_1^\sigma = a_1$, and so $a_1 \in E^G = k$. Therefore, $Z(A) = ku_1$.

We now show that $A$ is simple. Observe first that each $u_\sigma$ is invertible, for its inverse is $f(\sigma^{-1},\sigma)^{-1}u_{\sigma^{-1}}$ (remember that $\operatorname{im} f \subseteq E^\times$, so that its values are nonzero). Let $I$ be a nonzero two-sided ideal in $A$, and choose a nonzero $y = \sum_\sigma c_\sigma u_\sigma \in I$ of shortest length; that is, $y$ has the smallest number of nonzero coefficients. Multiplying by $(c_\sigma u_\sigma)^{-1}$ if necessary, we may assume that $y = u_1 + c_\tau u_\tau + \cdots$. Suppose that $c_\tau \neq 0$. Since $\tau \neq 1_E$, there is $a \in E$ with $a^\tau \neq a$. Now $I$ contains $ay - ya = b_\tau u_\tau + \cdots$, where $b_\tau = c_\tau(a - a^\tau) \neq 0$. Hence, $I$ contains $y - c_\tau b_\tau^{-1}(ay - ya)$, which is shorter than $y$ (it involves $u_1$ but not $u_\tau$). We conclude that $y$ must have length 1; that is, $y = c_\sigma u_\sigma$. But $y$ is invertible, and so $I = A$ and $A$ is simple.

Finally, Theorem C-2.128 says that $A$ is split by $K$, where $K$ is any maximal subfield of $A$. The reader may show, using Lemma C-2.118, that $Eu_1 \cong E$ is a maximal subfield. •

In light of Proposition C-3.139, it is natural to expect a connection between relative Brauer groups and cohomology.

Recall that two central simple $k$-algebras $A$ and $B$ are ***similar***, denoted by

$$A \sim B,$$

if there are integers $n$ and $m$ with $A \otimes_k \mathrm{Mat}_n(k) \cong B \otimes_k \mathrm{Mat}_m(k)$. If $[A]$ denotes the equivalence class of a central simple $k$-algebra $A$ under similarity, then the ***Brauer group*** $\mathrm{Br}(k)$ is the set

$$\mathrm{Br}(k) = \big\{[A] : A \text{ is a central simple } k\text{-algebra}\big\}$$

with binary operation $[A][B] = [A \otimes_k B]$.

**Theorem C-3.140.** *Let $E/k$ be a Galois extension with $G = \mathrm{Gal}(E/k)$. There is an isomorphism $H^2(G, E^\times) \to \mathrm{Br}(E/k)$ with $\mathrm{cls}\, f \mapsto [(E, G, f)]$.*

**Sketch of Proof.** The usual proofs of this theorem are rather long. Each of the items: the isomorphism is a well-defined function; it is a homomorphism; it is injective; it is surjective, must be checked, and the proofs are computational. For example, the proof in Herstein [**98**], pp. 110–116 is fairly long; there is a less computational proof, but still long, in Serre [**199**], pp. 164–167, using the method of *descent*. •

What is the advantage of this isomorphism? In Corollary C-2.133, we saw that

$$\mathrm{Br}(k) = \bigcup_{E/k \text{ finite}} \mathrm{Br}(E/k).$$

**Corollary C-3.141.** *Let $k$ be a field.*

(i) *The Brauer group $\mathrm{Br}(k)$ is a torsion group.*

(ii) *If $A$ is a central simple $k$-algebra, then there is an integer $n$ so that the tensor product of $A$ with itself $r$ times (where $r$ is the order of $[A]$ in $\mathrm{Br}(k)$) is a matrix algebra:*

$$A \otimes_k A \otimes_k \cdots \otimes_k A \cong \mathrm{Mat}_n(k).$$

**Sketch of Proof.**

(i) By Corollary C-2.133, $\mathrm{Br}(k)$ is the union of the relative Brauer groups $\mathrm{Br}(E/k)$, where $E/k$ is a finite Galois extension. We may now invoke Proposition C-3.123, which says that $|G|\, H^2(G, E^\times) = \{0\}$.

(ii) Tensor product is the binary operation in the Brauer group. •

Recall Proposition C-2.130: there exists a noncommutative division $k$-algebra over a field $k$ if and only if $\mathrm{Br}(k) \neq \{0\}$.

**Corollary C-3.142.** *Let $k$ be a field. If there is a cyclic Galois extension $E/k$ such that the norm $N\colon E^\times \to k^\times$ is not surjective, then there exists a noncommutative $k$-division algebra.*

**Sketch of Proof.** If $G$ is a finite cyclic group, then Theorem C-3.116 gives

$$H^2(G, E^\times) = (E^\times)^G / \operatorname{im} N = k^\times / \operatorname{im} N.$$

Hence, $\operatorname{Br}(E/k) \neq \{0\}$ if $N$ is not surjective, and this implies that $\operatorname{Br}(k) \neq \{0\}$. •

If $k$ is a finite field and $E/k$ is a finite extension, then it follows from Wedderburn's Theorem on finite division rings (Theorem C-2.31) that the norm $N\colon E^\times \to k^\times$ is surjective.

**Corollary C-3.143.** *If $p$ is a prime, then there exists a noncommutative division algebra of characteristic $p$.*

**Proof.** If $k$ is a field of characteristic $p$, it suffices to find a cyclic extension $E/k$ for which the norm $N\colon E^\times \to k^\times$ is not surjective; that is, we must find some $z \in k^\times$ which is not a norm.

If $p$ is an odd prime, let $k = \mathbb{F}_p(x)$. Since $p$ is odd, $t^2 - x$ is a separable irreducible polynomial, and so $E = k(\sqrt{x})$ is a Galois extension of degree 2. If $u \in E$, then there are polynomials $a, b, c \in \mathbb{F}_p[x]$ with $u = (a + b\sqrt{x})/c$. Moreover,

$$N(u) = (a^2 - b^2 x)/c^2.$$

We claim that $x^2 + x$ is not a norm. Otherwise,

$$a^2 - b^2 x = c^2(x^2 + x).$$

Since $c \neq 0$, the polynomial $c^2(x^2 + x) \neq 0$, and it has even degree. On the other hand, if $b \neq 0$, then $a^2 - b^2 x$ has odd degree, and this is a contradiction. If $b = 0$, then $u = a/c$; since $a^2 = c^2(x^2 + x)$, we have $c^2 \mid a^2$, hence $c \mid a$, and so $u \in \mathbb{F}_p[x]$ is a polynomial. But it is easy to see that $x^2 + x$ is not the square of a polynomial. We conclude that $N\colon E^\times \to k^\times$ is not surjective.

Here is an example in characteristic 2. Let $k = \mathbb{F}_2(x)$, and let $E = k(\alpha)$, where $\alpha$ is a root of $f(t) = t^2 + t + x + 1$ ($f(t)$ is irreducible and separable; its other root is $\alpha + 1$). As before, each $u \in E$ can be written in the form $u = (a + b\alpha)/c$, where $a, b, c \in \mathbb{F}_2[x]$. Of course, we may assume that $x$ is not a divisor of all three polynomials $a$, $b$, and $c$. Moreover,

$$N(u) = \big((a + b\alpha)(a + b\alpha + b)\big)/c^2 = \big(a^2 + ab + b^2(x + 1)\big)/c^2.$$

We claim that $x$ is not a norm. Otherwise,

$$a^2 + ab + b^2(x + 1) = c^2 x. \tag{2}$$

Now $a(0)$, the constant term of $a$, is either 0 or 1. Consider the four cases arising from the constant terms of $a$ and $b$; that is, evaluate Eq. (2) at $x = 0$. We see that $a(0) = 0 = b(0)$; that is, $x \mid a$ and $x \mid b$. Hence, $x^2 \mid a^2$ and $x^2 \mid b^2$, so that Eq. (2) has the form $x^2 d = c^2 x$, where $d \in \mathbb{F}_2[x]$. Dividing by $x$ gives $xd = c^2$, which forces $c(0) = 0$; that is, $x \mid c$, and this is a contradiction. •

For further discussion of the Brauer group, see the article by Serre in Cassels–Fröhlich [**38**], Jacobson [**111**], pp. 471–481, Reiner [**180**], Chapters 5, 7, and 8, and the article by Platonov–Yanchevskii in Kostrikin–Shafarevich [**128**]. In particular, a ***global field*** is a field which is either an algebraic number field (i.e., a finite extension of $\mathbb{Q}$) or a *function field* (a finite extension of $k(x)$, where $k$ is a finite field). To each global field, we assign a family of *local fields*. These fields are best defined in terms of discrete valuations.

A ***discrete valuation*** on a field $L$ is a function $v\colon L^\times \to \mathbb{N}$ such that, for all $a, b \in L$,

$$\begin{aligned} v(a) &= 0 \quad \text{if and only if } a = 0, \\ v(ab) &= v(a)v(b), \\ v(a+b) &= \max\{v(a), v(b)\}. \end{aligned}$$

Now $R = \{a \in L : v(a) \le 1\}$ is a domain and $P = \{a \in L : v(a) < 1\}$ is a maximal ideal in $R$. We call $R/P$ the ***residue field*** of $L$ with respect to the discrete valuation $v$. A ***local field*** is a field which is complete with respect to the metric arising from a discrete valuation on it and whose residue field is finite. It turns out that every local field is either a finite extension of $\mathbb{Q}_p$, the $p$-adic numbers (which is the fraction field of the $p$-adic integers $\mathbb{Z}_p$), or it is isomorphic to $\mathbb{F}_q[[x]]$, the ring of formal power series in one variable over a finite field $\mathbb{F}_q$. If $k$ is a local field, then $\mathrm{Br}(k) \cong H^2(k_s, k^\times)$, where $k_s/k$ is the maximal separable extension of $k$ in the algebraic closure $\overline{k}$. If $A$ is a central simple $K$-algebra, where $K$ is a global field and if $K_v$ is a local field of $K$, then $K_v \otimes_K A$ is a central simple $K_v$-algebra. The ***Hasse–Brauer–Noether–Albert Theorem*** states that if $A$ is a central simple algebra over a global field $K$, then $A \sim K$ if and only if $K_v \otimes_K A \sim K_v$ for all associated local fields $K_v$. We merely mention that these results were used by Chevalley to develop *class field theory* (the branch of algebraic number theory involving Galois extensions (of possibly infinite degree) having abelian Galois groups). See Neukirch–Schmidt–Wingberg [**166**].

For generalizations of the Brauer group (e.g., $\mathrm{Br}(k)$, where $k$ is a commutative ring) and ties to Morita theory, see Orzech–Small [**173**] and Caenepeel [**32**].

## Exercises

**C-3.64.** Show that the structure constants in the crossed product $(E, G, f)$ are

$$g_\alpha^{\sigma,\tau} = \begin{cases} f(\sigma,\tau) & \text{if } \alpha = \sigma\tau, \\ 0 & \text{otherwise.} \end{cases}$$

**C-3.65.** Prove that $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathrm{Mat}_4(\mathbb{R})$.

## C-3.11. Introduction to Spectral Sequences

The last topic we discuss is spectral sequences, whose major uses are in computing homology groups and in comparing homology groups of composites of functors. This brief section merely describes the setting for spectral sequences, in the hope that it will ease the reader's first serious encounter with them. For a more complete account, we refer the reader to Mac Lane [**145**], Chapter XI, McCleary [**152**], or Rotman [**187**], Chapter 11.

Call a series of submodules of a module $K$,

$$K = K_0 \supseteq K_1 \supseteq K_2 \supseteq \cdots \supseteq K_\ell = \{0\},$$

a ***filtration*** (instead of a normal series), and call the quotients $K_i/K_{i+1}$ the ***factor modules*** of the filtration. We know that a module $K$ may not be determined by the factor modules of a filtration; on the other hand, knowledge of the factor modules does give some information about $K$. For example, if all the factor modules are zero, then $K = \{0\}$; if all the factor modules are finite, then $K$ is finite (and $|K|$ is the product of the orders of the factor modules); or, if all the factor modules are finitely generated, then $K$ is finitely generated.

**Definition.** If $K$ is a module, then a ***subquotient*** of $K$ is a module isomorphic to $S/T$, where $T \subseteq S \subseteq K$ are submodules.

Thus, a subquotient of $K$ is a quotient of a submodule. It is also easy to see that a subquotient of $K$ is also a submodule of a quotient ($S/T \subseteq K/T$).

**Example C-3.144.**

(i) The factor modules of a filtration of a module $K$ are subquotients of $K$.

(ii) The $n$th homology group $H_n(\mathbf{C}_\bullet, d_\bullet)$ of a complex $(\mathbf{C}_\bullet, d_\bullet)$ is a subquotient of $C_n$. ◀

A spectral sequence computes a homology group $H_n$ in the sense that it computes the factor modules of some filtration of $H_n$. In general, this gives only partial information about $H_n$, but, if the factor modules are heavily constrained, then they can give much more information and, indeed, might even determine $H_n$ completely. For example, suppose that only one of the factor modules of $K$ is nonzero, say, $K_i/K_{i+1} \cong A \neq \{0\}$; we claim that $K \cong A$. The beginning of the filtration is

$$K = K_0 \supseteq K_1 \supseteq \cdots \supseteq K_i.$$

Since $K_0/K_1 = \{0\}$, we have $K = K_0 = K_1$. Similarly, $K_1/K_2 = \{0\}$ gives $K_1 = K_2$; indeed, $K = K_0 = K_1 = \cdots = K_i$. Similar reasoning computes the end of the filtration. For example, since $K_{\ell-1}/K_\ell = \{0\}$, we have $K_{\ell-1} = K_\ell = \{0\}$. Thus, the filtration is

$$K = K_0 = \cdots = K_i \supsetneq K_{i+1} = \cdots = K_\ell = \{0\},$$

and so $K \cong K/\{0\} = K_i/K_{i+1} \cong A$.

In order to appreciate spectral sequences, we must recognize an obvious fact: very general statements can become useful if extra simplifying hypotheses can be imposed.

Spectral sequences usually arise in the following context. A ***bigraded module*** $M = M_{\bullet\bullet}$ is a doubly indexed family of modules $M_{p,q}$, where $p, q \in \mathbb{Z}$; we picture a bigraded module as a collection of modules, one sitting on each lattice point $(p, q)$ in the plane. Thus, there are ***first quadrant*** bigraded modules, for example, with $M_{p,q} = \{0\}$ if either $p$ or $q$ is negative; similarly, there are ***third quadrant*** bigraded modules. A ***bicomplex*** is a bigraded module that has vertical arrows $d''_{p,q}\colon M_{p,q} \to M_{p,q-1}$ making the columns complexes, horizontal arrows $d'_{p,q}\colon M_{p,q} \to M_{p-1,q}$ making the rows complexes, and whose squares anticommute:

$$\begin{array}{ccc} M_{p-1,q} & \xleftarrow{d'_{p,q}} & M_{p,q} \\ \downarrow{\scriptstyle d''_{p-1,q}} & & \downarrow{\scriptstyle d''_{p,q}} \\ M_{p-1,q-1} & \xleftarrow[d'_{p,q-1}]{} & M_{p,q-1}; \end{array}$$

that is, $d'd'' + d''d' = 0$. The reason for the anticommutativity is to allow us to define the ***total complex***, $\mathrm{Tot}(M)$, of a bicomplex $M$: its term in degree $n$ is

$$\mathrm{Tot}(M)_n = \bigoplus_{p+q=n} M_{p,q};$$

its differentiation $d_n\colon \mathrm{Tot}(M)_n \to \mathrm{Tot}(M)_{n-1}$ is given by

$$d_n = \sum_{p+q=n} d'_{p,q} + d''_{p,q}.$$

Anticommutativity forces $d_{n-1}d_n = 0$:

$$dd = (d' + d'')(d' + d'') = d'd' + (d'd'' + d''d') + d''d'' = 0;$$

thus, $\mathrm{Tot}(M)$ is a complex.

All bigraded modules form a category. Given an ordered pair of integers $(a, b)$, a family of maps $f_{p,q}\colon M_{p,q} \to L_{p+a,q+b}$ is called a ***map*** $f\colon M_{\bullet\bullet} \to L_{\bullet\bullet}$ of ***bidegree*** $(a, b)$. For example, the maps $d'$ and $d''$ above have respective bidegrees $(0, -1)$ and $(-1, 0)$. It is easy to check that all bigraded modules and all maps having some bidegree form a category. One nice feature of composition is that bidegrees add: if $f$ has bidegree $(a, b)$ and $f'$ has bidegree $(a', b')$, then their composite $f'f$ has bidegree $(a + a', b + b')$. Maps of bigraded modules are used in establishing certain exact sequences. For example, one proof of the Five Term Exact Sequence uses these maps.

A ***spectral sequence*** is a sequence of bicomplexes, $E^r_{p,q}$, for all $r \geq 2$, where each $E^{r+1}_{p,q}$ is a subquotient of $E^r_{p,q}$ (we must also specify that the homomorphisms of the bicomplex $E^{r+1}_{p,q}$ arise from those of $E^r_{p,q}$). Most spectral sequences arise from a filtration of $\mathrm{Tot}(M)$, where $M_{\bullet\bullet}$ is a bicomplex. In particular, there are two "usual" filtrations (if $M_{\bullet\bullet}$ is either first quadrant or third quadrant), and the spectral sequences they determine are denoted by ${}^{\mathrm{I}}E^r_{p,q}$ and ${}^{\mathrm{II}}E^r_{p,q}$. In more detail, the ***first filtration*** of $\mathrm{Tot}(M)$ is the subcomplex whose $n$th term is the direct sum of those $M_{i,n-i}$ lying to the left of the vertical line at $p$ (see Figure C-3.1). The ***second filtration*** of $\mathrm{Tot}(M)$ is the subcomplex whose $n$th term is the direct sum of those $M_{i,n-i}$ lying below the horizontal line at $p$ (see Figure C-3.2).

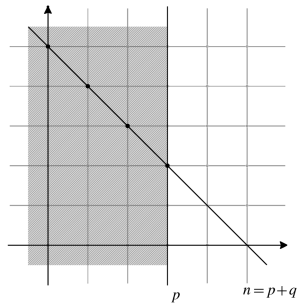



**Figure C-3.1.** First filtration.

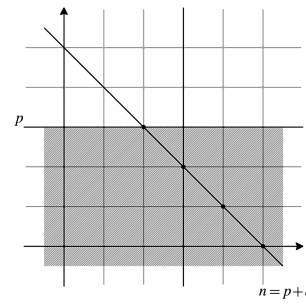



**Figure C-3.2.** Second filtration.

We say that a spectral sequence $E^r_{p,q}$ ***converges*** to a (singly graded) module $H_\bullet$, denoted by

$$E^2_{p,q} \Rightarrow H_n,$$

if each $H_n$ has a filtration with factor modules

$$E_{0,n}, E_{1,n-1}, \ldots, E_{n,0},$$

and, for all $p, q$ with $p + q = n$, the factor module $E_{p,q}$ is a subquotient of $E^2_{p,q}$. There are two steps to establish before using spectral sequences.

**Theorem I**. *If $M_{\bullet\bullet}$ is a first quadrant or third quadrant bicomplex, then*

$$^{\mathrm{I}}E^2_{p,q} \Rightarrow H_n(\mathrm{Tot}(M)) \quad \textit{and} \quad ^{\mathrm{II}}E^2_{p,q} \Rightarrow H_n(\mathrm{Tot}(M)).$$

Thus, for each $n$, there are two filtrations of $\mathrm{Tot}(M)_n$: one whose factor modules are subquotients of $^{\mathrm{I}}E^2_{p,q}$ and another whose factor modules are subquotients of $^{\mathrm{II}}E^2_{p,q}$ (as usual, $p + q = n$ in this context); and both converge to the same thing.

**Theorem II**. *If $M_{\bullet\bullet}$ is a first quadrant or third quadrant bicomplex, then there are formulas for $^{\mathrm{I}}E^2_{p,q}$ and $^{\mathrm{II}}E^2_{p,q}$ for every $p, q$. In more detail,*

$$^{\mathrm{I}}E^2_{p,q} = H'_p H''_q(M) \quad \textit{and} \quad ^{\mathrm{II}}E^2_{p,q} = H''_p H'_q(M).$$

The modules $H'_p H''_q(M)$ and $H''_p H'_q(M)$ are called ***iterated homology***. Each row and column of a double complex $M$ is a complex. In particular, for fixed $p$, the $p$th column $M_{p,\bullet}$ is a complex; its homology gives a (horizontal) complex whose $(p, q)$ term is $H_q(M_{p,\bullet})$ and whose horizontal maps are induced from the differentiations in the original bicomplex $M_{\bullet,\bullet}$. (In fact, $^{\mathrm{I}}E^1_{p,q} = H_q(M_{p,\bullet})$.) Informally, this iterated homology takes homology of the columns, then homology of the rows. Similarly, the other iterated homology takes homology of rows and then homology of columns.

We illustrate the technique by sketching a proof that $\mathrm{Tor}_n(A, B)$ does not depend on the variable resolved; that is, the value of $\mathrm{Tor}_n(A, B)$, defined as $H_n(\mathbf{P}_A \otimes B)$, where $\mathbf{P}_A$ is a deleted projective resolution of $A$, coincides with $\mathrm{Tor}_n(A, B)$, defined as $H_n(A \otimes \mathbf{Q}_B)$, where $\mathbf{Q}_B$ is a deleted projective resolution of $B$. The idea is to resolve both variables simultaneously, using resolutions of each. Define a first quadrant bigraded module $M = \mathbf{P}_A \otimes \mathbf{Q}_B$ whose $p, q$ term is $P_p \otimes Q_q$;

make this into a bicomplex by defining vertical arrows $d''_{p,q} = (-1)^p 1 \otimes \partial_q \colon P_p \otimes Q_q \to P_p \otimes Q_{q-1}$ and horizontal arrows $d'_{p,q} = \Delta_p \otimes 1 \colon P_p \otimes Q_q \to P_{p-1} \otimes Q_q$, where the $\partial_n$ are the differentiations in $\mathbf{Q}_B$ and the $\Delta_n$ are the differentiations in $\mathbf{P}_A$ (the signs force anticommutativity). The formula whose existence is stated in Theorem II for the first spectral sequence ${}^{\mathrm{I}}E^2_{p,q}$ gives, in this case,

$$ {}^{\mathrm{I}}E^2_{p,q} = \begin{cases} \{0\} & \text{if } q > 0, \\ H_p(\mathbf{P}_A \otimes B) & \text{if } q = 0. \end{cases} $$

Since a subquotient of $\{0\}$ must be $\{0\}$, all but one of the factor modules of a filtration of $H_n(\mathrm{Tot}(M))$ are zero, and so

$$ H_n(\mathrm{Tot}(M)) \cong H_n(\mathbf{P}_A \otimes B). $$

Similarly, the formula alluded to in Theorem II for the second spectral sequence gives

$$ {}^{\mathrm{II}}E^2_{p,q} = \begin{cases} \{0\} & \text{if } p > 0, \\ H_q(A \otimes \mathbf{Q}_B) & \text{if } p = 0. \end{cases} $$

Again, there is a filtration of $H_n(\mathrm{Tot}(M))$ with only one possible nonzero factor module, and so

$$ H_n(\mathrm{Tot}(M)) \cong H_n(A \otimes \mathbf{Q}_B). $$

Therefore,

$$ H_n(\mathbf{P}_A \otimes B) \cong H_n(\mathrm{Tot}(M)) \cong H_n(A \otimes \mathbf{Q}_B). $$

We have shown that Tor is independent of the variable resolved.

Here is a cohomology result illustrating how spectral sequences can be used to compute composite functors. The index raising convention extends here, so that one denotes the modules in a third quadrant bicomplex by $M^{p,q}$ instead of by $M_{-p,-q}$.

**Theorem C-3.145 (Grothendieck).** *Let $F \colon \mathcal{B} \to \mathcal{C}$ and $G \colon \mathcal{A} \to \mathcal{B}$ be additive functors, where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are module categories. If $F$ is left exact and if $E$ injective in $\mathcal{A}$ implies $(R^m F)(GE) = \{0\}$ for all $m > 0$ (where $R^m F$ are the right derived functors of $F$), then for every module $A \in \mathcal{A}$, there is a third quadrant spectral sequence*

$$ E_2^{p,q} = (R^p F)(R^q G(A)) \Rightarrow R^n(FG)(A). $$

For a proof, see Rotman [**187**], p. 350.

The next result shows that if $N$ is a normal subgroup of a group $\Pi$, then the cohomology groups of $N$ and of $\Pi/N$ can be used to compute the cohomology groups of $\Pi$.

**Theorem C-3.146 (Lyndon–Hochschild–Serre).** *Let $\Pi$ be a group with normal subgroup $N$. For each $\Pi$-module $A$, there is a third quadrant spectral sequence with*

$$ E_2^{p,q} = H^p(\Pi/N, H^q(N, A)) \Rightarrow H^n(\Pi, A). $$

**Proof.** Define functors $G \colon {}_{\mathbb{Z}\Pi}\mathbf{Mod} \to {}_{\mathbb{Z}(\Pi/N)}\mathbf{Mod}$ and $F \colon {}_{\mathbb{Z}(\Pi/N)}\mathbf{Mod} \to \mathbf{Ab}$ by $G = \mathrm{Hom}_N(\mathbb{Z}, \ )$ and $F = \mathrm{Hom}_{\Pi/N}(\mathbb{Z}, \ )$. Of course, $F$ is left exact, and it is easy to see that $FG = \mathrm{Hom}_\Pi(\mathbb{Z}, \ )$. A proof that $H^m(\Pi/N, E) = \{0\}$ whenever $E$ is an

injective Π-module and $m > 0$ can be found in Rotman [**187**], p. 307. The result now follows from Theorem C-3.145. •

This theorem was found by Lyndon in his dissertation in 1948, in order to compute the cohomology groups of finitely generated abelian groups Π. Several years later, Hochschild and Serre put the result into its present form.

Every convergent spectral sequence yields a five term exact sequence; the five-term exact sequence in cohomology of groups (Theorem C-3.125) is a special case (there is a similar theorem in homology).

**Theorem C-3.147.** *If $(E^r, d^r)$ is a third quadrant spectral sequence, so that $E^r_{p,q} \Rightarrow H_n(\mathrm{Tot}(M))$, then there is an exact sequence*

$$E_2^{n,0} \to H^1(\mathrm{Tot}(M)) \to E_2^{0,n} \xrightarrow{d_{n+1}} E_2^{n+1,0} \to H^{n+1}(\mathrm{Tot}(M)).$$

**Proof.** [**187**], p. 643. •

## Exercises

∗ **C-3.66.** Regard a commutative diagram of modules

$$\begin{array}{ccc} C & \xleftarrow{f} & D \\ {\scriptstyle j}\downarrow & & \downarrow{\scriptstyle g} \\ A & \xleftarrow[i]{} & B \end{array}$$

as a first quadrant double complex $M = M_{\bullet\bullet}$ by replacing $g$ by $-g$ and defining

$$M_{p,q} = \begin{cases} A & \text{if } (p,q) = (0,0), \\ B & \text{if } (p,q) = (1,0), \\ C & \text{if } (p,q) = (0,1), \\ D & \text{if } (p,q) = (1,1), \\ \{0\} & \text{otherwise.} \end{cases}$$

(i) Prove that $\mathrm{Tot}(M) = 0 \to D \xrightarrow{(f,-g)} C \oplus B \xrightarrow{j+i} A \to 0$.

(ii) Prove that $H_2(\mathrm{Tot}(M)) = \{d \in D : fd = 0 = gd\} = \ker f \cap \ker g$.

(iii) Prove that $H_0(\mathrm{Tot}(M)) = A/\operatorname{im}(j+i) = A/(B+C)$.

(iv) Prove that $H_1(\mathrm{Tot}(M)) = (\operatorname{im} j \cap \operatorname{im} i)/\operatorname{im}(jf)$.
**Hint.** If we display the nonzero terms of the bigraded module $H_q(M_{p,\bullet})$ as a matrix

$$\begin{bmatrix} h_{01} & h_{11} \\ h_{00} & h_{10} \end{bmatrix},$$

show that

$${}^{\mathrm{I}}E^1_{p,q} = [H_q(M_{p,\bullet})] = \begin{bmatrix} \operatorname{coker} f & \operatorname{coker} g \\ \ker f & \ker g \end{bmatrix}.$$

See Rotman [**187**], pp. 632–633, for further details.

**C-3.67.** Let $R$ be a ring with a right ideal $I$ and left ideal $J$. Use Exercise C-3.66 to prove that

$$\operatorname{Tor}_1^R(R/I, R/J) \cong (I \cap J)/IJ.$$

**Hint.** See Rotman [**187**], Proposition 10.20.

**C-3.68.** (i) If $R$ is a ring having a unique maximal left ideal $\mathfrak{m}$, prove that $\mathfrak{m}$ is a two-sided ideal.

(ii) If $k = R/\mathfrak{m}$, prove that $\operatorname{Tor}_1^R(k, k) \cong \mathfrak{m}/\mathfrak{m}^2$.

# More Categories

When we introduced category theory in Part 1, we pointed out how it organizes thought: it imposes the viewpoint that it is valuable to think, not only in terms of elements, sets, and functions, but also in terms of arrows, commutative diagrams, and universal mapping problems. We have seen various constructions and their duals: products and coproducts; pullbacks and pushouts; inverse limits and direct limits. And we can compare these constructions using functors and natural transformations.

In this chapter, we will consider the following questions, all of whose answers involve *abelian categories*. (1) We will introduce sheaves, and they form a category. Do sheaves behave like abelian groups? (2) What is the proper context for doing homological algebra? In particular, when can we define sheaf cohomology? (3) When is a category (naturally) equivalent to some category of modules? If $R$ and $S$ are rings, when are ${}_R\mathbf{Mod}$ and ${}_S\mathbf{Mod}$ equivalent? Are ${}_R\mathbf{Mod}$ and $\mathbf{Mod}_R$ equivalent?

Another circle of ideas asks when a functor preserves inverse or direct limits. This discussion involves the ***Adjoint Functor Theorem*** (which we will prove in the special case of module categories).

Finally, we will introduce ***algebraic $K$-theory***, which is to homological algebra what homotopy theory is to homology of topological spaces.

## C-4.1. Additive Categories

Representations of a group $G$ are just another way of considering $kG$-modules, so that contemplating all the representations of $G$ is the same as contemplating the category ${}_{kG}\mathbf{Mod}$. It is natural to ask, more generally, to what extent a category ${}_R\mathbf{Mod}$ determines a ring $R$. We now prepare the answer to this question; the answer itself is given later in this chapter.

Recall that an object $A$ in a category $\mathcal{C}$ is an ***initial object*** if there is a unique morphism $A \to X$ for every $X \in \text{obj}(\mathcal{C})$; an object $\Omega$ in $\mathcal{C}$ is a ***terminal object*** if there is a unique morphism $X \to \Omega$ for every $X \in \text{obj}(\mathcal{C})$; and an object is a ***zero object*** if it is both initial and terminal. We now introduce *additive* categories, which generalize pre-additive categories defined on page 446 in Part 1.

**Definition.** A category $\mathcal{A}$ is ***additive*** if it is pre-additive; that is,

(i) $\text{Hom}_{\mathcal{A}}(A, B)$ is an (additive) abelian group for every $A, B \in \text{obj}(\mathcal{A})$;

(ii) the distributive laws hold: given morphisms

$$X \xrightarrow{k} A \underset{g}{\overset{f}{\rightrightarrows}} B \xrightarrow{h} Y,$$

where $X$ and $Y \in \text{obj}(\mathcal{A})$, then

$$h(f+g) = hf + hg \quad \text{and} \quad (f+g)k = fk + gk;$$

and also

(a) $\mathcal{A}$ has a zero object;

(b) $\mathcal{A}$ has finite products and finite coproducts: for all objects $A, B$ in $\mathcal{A}$, both $A \sqcap B$ and $A \sqcup B$ exist in $\text{obj}(\mathcal{A})$.

Let $\mathcal{A}$ and $\mathcal{C}$ be additive categories. A functor $T\colon \mathcal{A} \to \mathcal{C}$ (of either variance) is ***additive*** if, for all $A, B \in \text{obj}(\mathcal{A})$ and all $f, g \in \text{Hom}_{\mathcal{A}}(A, B)$, we have

$$T(f+g) = Tf + Tg;$$

that is, the function $\text{Hom}_{\mathcal{A}}(A, B) \to \text{Hom}_{\mathcal{C}}(TA, TB)$, given by $f \mapsto Tf$, is a homomorphism of abelian groups.

Exercise C-4.6 on page 344 says that neither **Groups** nor **ComRings** (see Example B-4.1 in Part 1) is a pre-additive category; hence, neither is additive.

It is easy to see, when $\mathcal{A}$ is an additive category, that both Hom functors $\text{Hom}_{\mathcal{A}}(X, \quad)\colon \mathcal{A} \to \mathbf{Ab}$ and $\text{Hom}_{\mathcal{A}}(\ , Y)\colon \mathcal{A} \to \mathbf{Ab}$ are additive. Both ${}_R\mathbf{Mod}$ and $\mathbf{Mod}_R$ are additive categories, and Theorem B-4.80 in Part 1 shows that $X \otimes_R -\colon {}_R\mathbf{Mod} \to \mathbf{Ab}$ and $- \otimes_R Y\colon \mathbf{Mod}_R \to \mathbf{Ab}$ are additive functors.

That finite coproducts and products coincide for modules is a special case of a more general fact: finite products and finite coproducts coincide in all additive categories. Note that Exercise C-4.1 on page 344 below says that if $T$ is an additive functor, then $T(0) = 0$, where 0 is either a zero morphism or a zero object.

**Lemma C-4.1.** *Let $\mathcal{C}$ be an additive category, and let $M, A, B \in \text{obj}(\mathcal{C})$. Then $M \cong A \sqcap B$ (their product) if and only if there are morphisms $i\colon A \to M$, $j\colon B \to M$, $p\colon M \to A$, and $q\colon M \to B$ such that*

$$pi = 1_A, \quad qj = 1_B, \quad pj = 0, \quad qi = 0, \quad \text{and} \quad ip + jq = 1_M.$$

*Moreover, $A \sqcap B$ is also a coproduct with injections $i$ and $j$, and so*

$$A \sqcap B \cong A \sqcup B.$$

**Proof.** The proof of the first statement is a variation of the proof of Proposition B-2.17 in Part 1, and the proof of the second statement is a variation of the proof of Proposition B-4.6 in Part 1; both proofs here are left to the reader. The last statement holds because two coproducts of $A$ and $B$ (here $A \sqcup B$ and $A \sqcap B$) must be isomorphic. •

If $A$ and $B$ are objects in an additive category, then $A \sqcap B \cong A \sqcup B$; their common value, denoted by

$$A \oplus B,$$

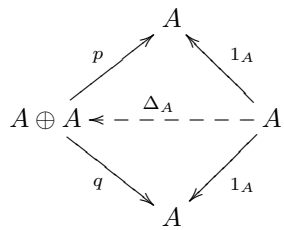
is called their ***direct sum*** (or ***biproduct***).

Addition of homomorphisms in **Ab** can be described without elements, as we saw on page 302. In ${}_R\mathbf{Mod}$, define the ***diagonal*** $\Delta\colon A \to A \oplus A$ by $\Delta\colon a \mapsto (a, a)$; dually, the ***codiagonal*** $\nabla\colon B \oplus B \to B$ is defined by $\nabla\colon (b, b') \mapsto b + b'$. If $f, g\colon A \to B$, then we saw, on page 302 that

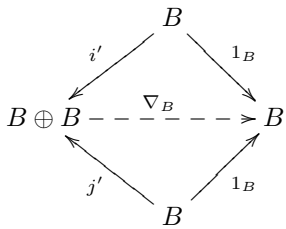
$$\nabla(f \oplus g)\Delta = f + g.$$

As usual, the advantage of definitions given in terms of maps (rather than in terms of elements) is that they can be recognized by functors. Diagonals and codiagonals can be defined and exist in additive categories.

**Definition.** Let $\mathcal{A}$ be an additive category. If $A \in \mathrm{obj}(\mathcal{A})$, then the ***diagonal*** $\Delta_A\colon A \to A \oplus A$ is the unique morphism with $p\Delta_A = 1_A$ and $q\Delta_A = 1_A$, where $p, q$ are projections of the direct product:




If $B \in \mathrm{obj}(\mathcal{A})$, then the ***codiagonal*** $\nabla_B\colon B \oplus B \to B$ is the unique morphism with $\nabla_B i' = 1_B$ and $\nabla_B j' = 1_B$, where $i', j'$ are injections of the direct sum:




The reader should check that these definitions, when specialized to ${}_R\mathbf{Mod}$, give the original diagonal and codiagonal homomorphisms.

**Lemma C-4.2.** *If $\mathcal{A}$ is an additive category and $f, g \in \mathrm{Hom}_{\mathcal{A}}(A, B)$, then*

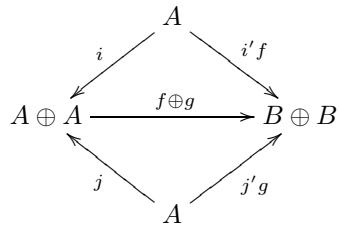
$$\nabla_B(f \oplus g)\Delta_A = f + g.$$

**Proof.** Let $p, q\colon A \oplus A \to A$ be projections, and let $i, j\colon A \to A \oplus A$ and $i', j'\colon B \to B \oplus B$ be injections. We compute

$$\begin{aligned}
\nabla_B(f \oplus g)\Delta_A &= \nabla_B(f \oplus g)(ip + jq)\Delta_A \\
&= \nabla_B(f \oplus g)(ip\Delta_A + jq\Delta_A) \\
&= \nabla_B(f \oplus g)(i + j) \quad \text{(because } p\Delta_A = 1_A = q\Delta_A\text{)} \\
&= \nabla_B(f \oplus g)i + \nabla_B(f \oplus g)j \\
&= \nabla_B i'f + \nabla_B j'g \qquad \text{(Exercise B-4.86 on page 522 in Part 1)} \\
&= f + g \qquad\qquad \text{(because } \nabla_B i' = 1_B = \nabla_B j'\text{).} \quad \bullet
\end{aligned}$$

**Definition.** A functor $T\colon \mathcal{A} \to \mathcal{B}$ between additive categories ***preserves finite direct sums*** if, for all $A, B \in \mathrm{obj}(\mathcal{A})$, whenever $A \oplus B$ is a direct sum with projections $p, q$ and injections $i, j$, then $TA \oplus TB$ is a direct sum with projections $Tp, Tq$ and injections $Ti, Tj$.

Recall Exercise B-4.86 on page 522 in Part 1: if $i, j\colon A \to A \oplus A$ and $i', j'\colon B \to B \oplus B$ are injections and $f, g\colon A \to B$, then $f \oplus g\colon A \oplus A \to B \oplus B$ is the unique map completing the coproduct diagram




It follows that if $T$ preserves finite direct sums, then $T(f \oplus g) = Tf \oplus Tg$.

**Proposition C-4.3.** *A functor $T\colon \mathcal{A} \to \mathcal{B}$ between additive categories is additive if and only if $T$ preserves finite direct sums.*

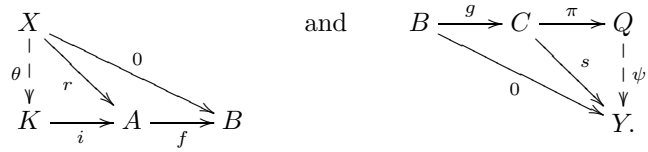
**Proof.** If $T$ is additive, then $T$ preserves finite direct sums, by Lemma C-4.1.

Conversely, let $T$ preserve finite direct sums. If $f, g\colon A \to B$, then

$$\begin{aligned}
T(f + g) &= T\big(\nabla_B(f \oplus g)\Delta_A\big) \qquad \text{(by Lemma C-4.2)} \\
&= (T\nabla_B)T(f \oplus g)(T\Delta_A) \\
&= \nabla_{TB}T(f \oplus g)\Delta_{TA} \\
&= \nabla_{TB}(Tf \oplus Tg)\Delta_{TA} \qquad \text{(Exercise B-4.86 on page 522 in Part 1)} \\
&= Tf + Tg. \quad \bullet
\end{aligned}$$

We can describe kernels and cokernels categorically. We have seen, in Examples B-4.9 and B-4.12 in Part 1, that kernels in **Ab** are pullbacks and cokernels in **Ab** are pushouts. Thus, kernel and cokernel in **Ab** are solutions to universal mapping problems, and we now consider them in additive categories (where zero morphisms exist).

Let $\mathcal{A}$ be an additive category, let $f\colon A \to B$ in $\mathcal{A}$, and consider the universal mapping problem illustrated below on the left: if $fr = 0$, then there exists a unique $\theta\colon X \to K$ with $i\theta = r$:




Denote a solution by $(K, i)$. Dually, if $g\colon B \to C$, consider the universal mapping problem illustrated above on the right: if $sg = 0$, then there exists a unique $\psi\colon Q \to Y$ with $\psi\pi = s$. Denote a solution by $(Q, \pi)$.

As usual, solutions to universal mapping problems, when they exist, are only unique to isomorphism. For example, suppose that $(K', i')$ is another solution posed by the kernel of $f\colon A \to B$. The morphisms $\theta'$ and $\theta$ satisfy $\theta\theta' = 1_{K'}$ and $\theta'\theta = 1_K$

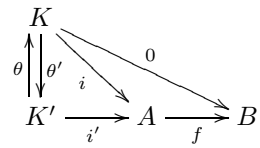



**Figure C-4.1.** Uniqueness of kernel.

(for example, use uniqueness in a similar diagram in which $K' = K$ and $i' = i$), and hence are isomorphisms. Thus, $i' = i\theta$, which defines an equivalence relation on the family of all such morphisms.

**Definition.** Let $\mathcal{A}$ be an additive category. If $f\colon A \to B$ in $\mathcal{A}$, then its ***kernel*** is the equivalence class

$$\ker f = [i]$$

of the equivalence relation illustrated in Figure C-4.1: $i \sim i'$ if $i' = i\theta$, where $\theta\colon \text{domain}(i') \cong \text{domain}(i)$ is an isomorphism.

Similarly, if $g\colon B \to C$, then its ***cokernel*** is the equivalence class $\operatorname{coker} g = [\pi]$, where if $\pi' \in [\pi]$, then $\pi = \psi\pi'$.

We have just described kernels and cokernels as equivalence classes of morphisms, but this is not as strange as it appears. We have written a solution to the universal problem posed by kernels as $(K, i)$, but $K = \text{domain}(i)$ is redundant; Hom sets in a category are pairwise disjoint, and so the morphisms $i$ have unique domains. Similarly, the equivalence class $[\pi]$ determines its target $Q$.

Let us compare the categorical definitions of kernel and cokernel with the familiar definitions in **Ab**. If $f\colon A \to B$ is a homomorphism, then $\ker f$ is usually defined as the subgroup $K = \{a \in A : f(a) = 0\}$ of $A$; if $i\colon K \to A$ is the inclusion, then $[i] = \ker f$ in **Ab** (see Exercise B-4.10 on page 458 in Part 1). Similarly, if

$g\colon B \to C$, then the quotient $C/\operatorname{im} g = \operatorname{target}(\pi)$ and $[\pi] = \operatorname{coker} g$ is the cokernel in **Ab**, where $\pi\colon C \to C/\operatorname{im} g$ is the natural map (we will soon define *image* in $\mathcal{A}$). Choosing a favorite representative of an equivalence class is quite natural. For example, we defined a gcd of two elements in a PID $R$ as a common divisor $d \in R$ such that $c \mid d$ for every common divisor $c$. When $R = \mathbb{Z}$, we chose $d$ to be nonnegative; when $R = k[x]$ for a field $k$, we chose $d$ to be monic. To sum up, kernels and cokernels in additive categories $\mathcal{A}$ are equivalence classes of morphisms. However, a particular additive category (e.g., $\mathcal{A} = \mathbf{Ab}$) may equip its objects with more data, and this extra information may enable us to select a special morphism (e.g., an inclusion or a natural map) to represent a kernel or cokernel. In **Groups**, kernels are certain subgroups, and categorical kernels are defined as (equivalence classes) of inclusions of such subgroups.

## Exercises

∗ **C-4.1.** Let $\mathcal{A}$ and $\mathcal{C}$ be additive categories. If $T\colon \mathcal{A} \to \mathcal{C}$ is an additive functor, prove that $T(0) = 0$, where 0 is either a zero morphism or a zero object.

**Hint.** If $Z$ is a zero object, then its identity $1_Z$ is a zero morphism.

**C-4.2.** If $\mathcal{C}$ is an additive category with zero object 0, prove that the unique morphism $A \to 0$ (where $A \in \operatorname{obj}(\mathcal{C})$) and the unique morphism $0 \to A$ are the identity elements of the abelian groups $\operatorname{Hom}_{\mathcal{C}}(A, 0)$ and $\operatorname{Hom}_{\mathcal{C}}(0, A)$.

∗ **C-4.3.** If $\mathcal{A}$ is an additive category and $A \in \operatorname{obj}(\mathcal{A})$, prove that $\operatorname{End}_{\mathcal{A}}(A) = \operatorname{Hom}_{\mathcal{A}}(A, A)$ is a ring with composition as product.

∗ **C-4.4.** In any category having a zero object, prove that every kernel is a monomorphism and, dually, every cokernel is an epimorphism.

∗ **C-4.5.** Let $\mathcal{C}$ be an additive category and let $\mathcal{S}$ be a subcategory. Prove that $\mathcal{S}$ is an additive category if $\mathcal{S}$ is full, contains a zero object of $\mathcal{C}$, and, for all $A, B \in \operatorname{obj}(\mathcal{S})$, the coproduct $A \sqcup B$ and the product $A \sqcap B$ in $\mathcal{C}$ lie in $\mathcal{S}$.

∗ **C-4.6.** Prove that neither **Groups** nor **ComRings** is an additive category.

**Hint.** Use Lemma C-4.1.

## C-4.2. Abelian Categories

Abelian categories, which generalize the categories ${}_R\mathbf{Mod}$ and $\mathbf{Mod}_R$ of modules, are the proper setting for doing homological algebra.

We have been reluctant to discuss injections and surjections in categories; after all, morphisms in a category need not be functions. On the other hand, it is often convenient to have them.

**Definition.** A morphism $f\colon A \to B$ in a (not necessarily additive) category $\mathcal{C}$ is a ***monomorphism***[1] (or is ***monic***) if $f$ can be canceled from the left; that is, for all

[1] A useful notation for a monomorphism $A \to B$ is $A \rightarrowtail B$, while a notation for an epimorphism $B \to C$ is $B \twoheadrightarrow C$.

objects $X$ and all morphisms $h, k\colon X \to A$, we have $fh = fk$ implies $h = k$,

$$X \underset{k}{\overset{h}{\rightrightarrows}} A \xrightarrow{f} B.$$

It is clear that $f\colon A \to B$ is monic if and only if, for all objects $X$ in $\mathcal{C}$, the induced map $f_*\colon \mathrm{Hom}(X, A) \to \mathrm{Hom}(X, B)$ is an injection in **Sets**. In an additive category, $\mathrm{Hom}(X, A)$ and $\mathrm{Hom}(X, B)$ are abelian groups, and so $f$ is monic if and only if $fh = 0$ implies $h = 0$. Exercise C-4.7 on page 357 shows that monomorphisms and injections coincide in **Sets**, ${}_R\mathbf{Mod}$, and **Groups**. Even in a category whose morphisms are actually functions, however, monomorphisms need not be injections (see Exercise C-4.16 on page 358).

Here is the dual definition.

**Definition.** A morphism $g\colon B \to C$ in a (not necessarily additive) category $\mathcal{C}$ is an ***epimorphism*** (or is ***epic***) if $g$ can be canceled from the right; that is, for all objects $Y$ and all morphisms $u, v\colon C \to Y$, we have $ug = vg$ implies $u = v$,

$$B \xrightarrow{g} C \underset{v}{\overset{u}{\rightrightarrows}} Y.$$

It is clear that $g\colon B \to C$ is epic if and only if, for all objects $Y$ in $\mathcal{C}$, the induced map $g^*\colon \mathrm{Hom}(C, Y) \to \mathrm{Hom}(B, Y)$ is an injection in **Sets**. In an additive category, Hom sets are abelian groups, and so $g$ is epic if and only if $ug = 0$ implies $u = 0$. Exercise C-4.7 on page 357 shows that epimorphisms and surjections coincide in **Sets** and in ${}_R\mathbf{Mod}$. Every surjective homomorphism in **Groups** is epic, but we must be clever to show this (Exercise C-4.11 on page 358). Even in a category whose morphisms are actually functions, epimorphisms need not be surjections. For example, if $R$ is a domain, then the ring homomorphism $g\colon R \to \mathrm{Frac}(R)$, given by $g\colon r \mapsto r/1$, is an epimorphism in **ComRings**: if $A$ is a commutative ring and $u, v\colon \mathrm{Frac}(R) \to A$ are ring homomorphisms agreeing on $R$, then $u = v$. In particular, the inclusion $\mathbb{Z} \to \mathbb{Q}$ is epic in **ComRings**.

**Proposition C-4.4.** *Let $f\colon A \to B$ and $g\colon B \to C$ be morphisms in an additive category $\mathcal{A}$.*

(i) *If* $\ker f$ *exists, then $f$ is monic if and only if* $\ker f = [0]$.

(ii) *Dually, if* $\mathrm{coker}\, g$ *exists, then $g$ is epic if and only if* $\mathrm{coker}\, g = [0]$.

**Proof.**

(i) We refer to the diagrams in the definitions of kernel and cokernel.

Let $\ker f = [i] = [0]$, where $i\colon K \to A$. If $h\colon X \to A$ satisfies $fh = 0$, then the universal property of kernel provides a morphism $\theta\colon X \to K$ with $h = i\theta = 0$ (because $i = 0$). Hence, $f$ is monic.

Conversely, if $f$ is monic, consider

$$K \underset{0}{\overset{i}{\rightrightarrows}} A \xrightarrow{f} B.$$

Since $fi = 0 = f0$, we have $i = 0$. Therefore, $\ker f = [0]$.

(ii) The proof for epimorphism and coker is dual. •

The converse of Proposition C-4.5 is true in *abelian categories*, which are additive categories in which a reasonable notion of exactness can be defined. They are so called because of their resemblance to **Ab**. These are the most interesting categories for homological algebra, for exactness and homology can be defined in them.

**Definition.** A category $\mathcal{A}$ is an ***abelian category*** if it is an additive category such that

(i) every morphism has a kernel and a cokernel;

(ii) every monomorphism is a kernel and every epimorphism is a cokernel.

In more detail, axiom (i) says that if $f$ is a morphism in $\mathcal{A}$, then $\ker f$ and $\operatorname{coker} f$ exist in $\mathcal{A}$. Axiom (ii) says that if $f$ is monic, then there is a morphism $i$ in $\mathcal{A}$ with $[i] = \ker f$; similarly, if $g$ is epic, then there is a morphism $\pi$ in $\mathcal{A}$ with $[\pi] = \operatorname{coker} g$.

**Proposition C-4.5.** *Let $\mathcal{A}$ be an abelian category.*

(i) *A morphism $f$ is a monomorphism if and only if* $\ker f = [0]$*, and a morphism $g$ is an epimorphism if and only if* $\operatorname{coker} f = [0]$.

(ii) *If $[i] = \ker f$, then $i$ is a monomorphism, and if $[\pi] = \operatorname{coker} g$, then $\pi$ is an epimorphism.*

**Proof.**

(i) Immediate from Proposition C-4.4, for kernels and cokernels always exist in abelian categories.

(ii) Suppose that $X \xrightarrow{h} K \xrightarrow{i} A$ and $ih = 0$. Since $ih = 0$, there is a unique $\theta\colon X \to K$ with $i\theta = ih = 0$. Obviously $\theta = 0$ satisfies this equation, and so uniqueness gives $h = \theta = 0$. Therefore, (i) gives $i$ monic.

A dual argument shows that cokernels are comprised of epimorphisms. •

Once we define $\operatorname{im} f$ in an abelian category $\mathcal{A}$, we will be able to define exactness and homology in $\mathcal{A}$. We look first at a homomorphism $f\colon A \to B$ in **Ab**; now $I = \operatorname{im} f \subseteq B$ makes sense in **Ab**. Indeed, if $B/I = Q$, then the natural map $\pi$ is a homomorphism $B \to Q$ whose kernel is $I$. In categorical language, $\operatorname{coker} f = [\pi]$ (if $\pi' \in [\pi]$, then $\pi = \psi\pi'$). Thus, $\ker \pi = I$ in **Ab**; that is, $I = \operatorname{im} f = \ker(\operatorname{coker} f)$. Figure C-4.2 is the picture with $Q = B/I$.

**Definition.** Let $f\colon A \to B$ be a morphism in an abelian category. Then the ***image*** of $f$ is

$$\operatorname{im} f = [j] = \ker(\operatorname{coker} f),$$

where $I = \operatorname{domain}(j) \in \ker(\operatorname{coker} f)$ and $j\colon I \to B$.

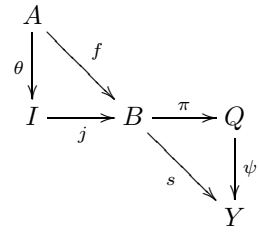



**Figure C-4.2.** Image diagram.

We can now define exactness in an abelian category.

**Definition.** A sequence $A \xrightarrow{f} B \xrightarrow{g} C$ in an abelian category $\mathcal{A}$ is ***exact*** if

$$\ker g = \operatorname{im} f.$$

In terms of morphisms, if $\ker g = [i]$ and $\operatorname{im} f = [j]$, then we are saying $[i] = [j]$. We can picture this. Note that both $j$ and $i\theta$ are morphisms $I \to B$. Note also that $\ker g = [i]$, where $i\colon K \to B$.

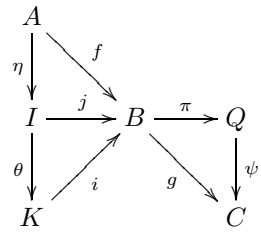



**Figure C-4.3.** Exactness diagram.

Viewing this diagram in **Ab**, we have $\operatorname{im}\theta\eta \subseteq K = \ker g$, so that $\ker g/\operatorname{im} f$ can be regarded as $\operatorname{coker}(\ker\theta\eta)$.

And now homology can be defined in any abelian category.

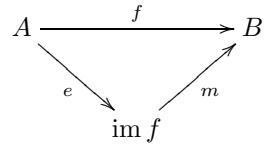
**Definition.** If $A \xrightarrow{f} B \xrightarrow{g} C$ are morphisms in an abelian category $\mathcal{A}$ with $gf = 0$, then ***homology at*** $B$ is

$$H = \operatorname{coker}(\ker\theta\eta),$$

where $\eta\colon A \to I$ and $\theta\colon I \to K$ (see Figure C-4.3).

**Remark.** There is another (equivalent) way to view exactness. If $f\colon A \to B$ is a morphism in an abelian category, then $f = me$, where $m = \ker(\operatorname{coker} f)$ is monic and $e = \operatorname{coker}(\ker f)$ is epic. In down-to-earth language, $f$ is the composite of its image $e$ (epic) followed by its inclusion $m$ (monic). Actually, this is the diagram

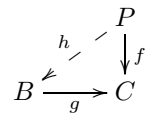
that usually accompanies the statement of the First Isomorphism Theorem,



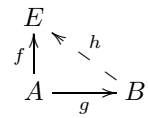

Moreover, this factorization is unique in the following sense. If $f = m'e'$, where $m'$ is monic and $e'$ is epic, then there is equality $[m] = [m']$ and $[e] = [e']$ (see Mac Lane [**144**], Chapter VIII, Sections 1 and 3). In light of this, we may redefine exactness of a sequence in an abelian category. If $f = me$ and $g = m'e'$, where $m, m'$ are monic and $e, e'$ are epic, then $A \xrightarrow{f} B \xrightarrow{g} C$ is exact if and only if $[e] = [m']$. ◀

Projectives and injectives can be defined in any category, but the definitions involve epimorphisms and monomorphisms. Of course, even in abelian categories, projective or injective objects may not exist. But, if there exist enough projective or injective objects in an abelian category, then we can use them to define derived functors. In fact, we will see that the category of sheaves is an abelian category, every sheaf has an injective resolution, and this will allow us to introduce sheaf cohomology.

**Definition.** An object $P$ in a category $\mathcal{C}$ is ***projective*** if, for every epic $g\colon B \to C$ and every $f\colon P \to C$, there exists $h\colon P \to B$ with $f = gh$:




An object $E$ in a category $\mathcal{C}$ is ***injective*** if, for every monic $g\colon A \to B$ and every $f\colon A \to E$, there exists $h\colon B \to E$ with $f = hg$:




Recognizing which morphisms in general categories are monic or epic is too difficult, and so we usually consider projective and injective objects only in abelian categories where Proposition C-4.5 is available to help.

**Remark.** Abelian categories are ***self-dual*** in the sense that the dual of every axiom in its definition is itself an axiom; it follows that if $\mathcal{A}$ is an abelian category, then so is its opposite $\mathcal{A}^{\mathrm{op}}$. A theorem using only these axioms in its proof is true in every abelian category; moreover, its dual is also a theorem in every abelian category, and its proof is dual to the original proof. The categories ${}_R\mathbf{Mod}$ and $\mathbf{Mod}_R$ are abelian categories having extra properties; for example, $R$ is a special type of object. Module categories are not self-dual, and this explains why a theorem and its dual, both of which are true in every module category, may have very different proofs. For example, the statements "every module is a quotient of a

projective module" and "every module can be imbedded in an injective module" are dual and are always true. The proofs are not dual because these statements are not true in every abelian category. Exercise C-4.15 below shows that the abelian category of all torsion abelian groups has no nonzero projectives and the abelian category of all finitely generated abelian groups has no nonzero injectives. ◀

Recall that a category $\mathcal{S}$ is a ***subcategory*** of a category $\mathcal{C}$ if

(i) $\text{obj}(\mathcal{S}) \subseteq \text{obj}(\mathcal{C})$;

(ii) $\text{Hom}_{\mathcal{S}}(A, B) \subseteq \text{Hom}_{\mathcal{C}}(A, B)$ for all $A, B \in \text{obj}(\mathcal{S})$;

(iii) if $f \in \text{Hom}_{\mathcal{S}}(A, B)$ and $g \in \text{Hom}_{\mathcal{S}}(B, C)$, the composite $gf \in \text{Hom}_{\mathcal{S}}(A, C)$ is equal to the composite $gf \in \text{Hom}_{\mathcal{C}}(A, C)$;

(iv) if $A \in \text{obj}(\mathcal{S})$, then $1_A \in \text{Hom}_{\mathcal{S}}(A, A)$ is equal to $1_A \in \text{Hom}_{\mathcal{C}}(A, A)$.

A subcategory $\mathcal{S}$ of a category $\mathcal{C}$ is ***full*** if $\text{Hom}_{\mathcal{S}}(A, B) = \text{Hom}_{\mathcal{C}}(A, B)$ for all $A, B \in \text{obj}(\mathcal{S})$.

It is easy to see that the inclusion of a subcategory is a functor. The subcategory **Ab** is a full subcategory of **Groups**, but if we regard **Top** as a subcategory of **Sets**, then it is not a full subcategory, for there are functions between spaces that are not continuous.

**Example C-4.6.**

(i) For every ring $R$, both ${}_R\mathbf{Mod}$ and $\mathbf{Mod}_R$ are abelian categories. In particular, ${}_{\mathbb{Z}}\mathbf{Mod} = \mathbf{Ab}$ is abelian. If $S$ is a ring, then the category of bimodules ${}_R\mathbf{Mod}_S$ is abelian.

(ii) The full subcategory $\mathcal{G}$ of **Ab** of all finitely generated abelian groups is an abelian category, as is the full subcategory of all torsion abelian groups.

(iii) The full subcategory of **Ab** of all torsion-free abelian groups is *not* an abelian category, for there are morphisms having no cokernel; for example, the inclusion $2\mathbb{Z} \to \mathbb{Z}$ has cokernel $\mathbb{Z}_2$, which is not torsion-free.

(iv) Quillen introduced a more general notion that is adequate for algebraic $K$-theory.

**Definition.** A category $\mathcal{E}$ is an ***exact category*** if $\mathcal{E}$ is a full subcategory of some abelian category $\mathcal{A}$ and if $\mathcal{E}$ is *closed under extensions*; that is, if $0 \to E' \to A \to E'' \to 0$ is an exact sequence in $\mathcal{A}$ and if $E', E'' \in \text{obj}(\mathcal{E})$, then $A \in \text{obj}(\mathcal{E})$.

Every abelian category is an exact category. The full subcategory of **Ab** consisting of all torsion-free abelian groups is an exact category, but it is not an abelian category. There is another, similar definition of *exact category* due to Buchsbaum (see Cartan–Eilenberg [**36**], Appendix).

(v) The category **Groups** is not abelian (it is not even additive). If $S \subseteq G$ is a nonnormal subgroup of a group $G$, then the inclusion $i\colon S \to G$ has no

cokernel. However, if $K$ is a normal subgroup of $G$ with inclusion $j\colon K \to G$, then $\operatorname{coker} j$ does exist. Thus, axiom (ii) in the definition of abelian category essentially says that every "subobject" in an abelian category is normal. ◀

The coming discussion will give several more examples of abelian categories.

**Proposition C-4.7.** *Let $\mathcal{S}$ be a full subcategory of an abelian category $\mathcal{A}$. If*

(i) *a zero object in $\mathcal{A}$ lies in $\mathcal{S}$,*

(ii) *for all $A, B \in \operatorname{obj}(\mathcal{S})$, the direct sum $A \oplus B$ in $\mathcal{A}$ lies in $\mathcal{S}$,*

(iii) *for all $A, B \in \operatorname{obj}(\mathcal{S})$ and all $f\colon A \to B$, both $\ker f$ and $\operatorname{coker} f$ lie in $\mathcal{S}$,*

*then $\mathcal{S}$ is an abelian category.*

**Remark.** When we say that a morphism $i\colon K \to A$ in $\mathcal{A}$ lies in a subcategory $\mathcal{S}$, we assume that its domain $K$ and target $A$ lie in $\operatorname{obj}(\mathcal{S})$. ◀

**Proof.** That $\mathcal{S}$ is a full subcategory of $\mathcal{A}$ satisfying (i) and (ii) gives $\mathcal{S}$ additive, by Exercise C-4.5 on page 344.

If $f\colon A \to B$ is a morphism in $\mathcal{S} \subseteq \mathcal{A}$, then $\ker f$ and $\operatorname{coker} f$ lie in $\mathcal{A}$, and hence they lie in $\mathcal{S}$, by (iii). Thus, axiom (i) in the definition of abelian category holds; it remains to verify axiom (ii).

Let $u$ be a monomorphism in $\mathcal{S}$; we have just seen that $\ker u$ lies in $\mathcal{S}$. Since $u$ is monic, Proposition C-4.4 gives $\ker u = 0$ in $\mathcal{S}$. By hypothesis, $\ker u$ is the same in $\mathcal{A}$ as in $\mathcal{S}$, so that $\ker u = 0$ in $\mathcal{A}$; hence, Proposition C-4.5 says that $u$ is a monomorphism in $\mathcal{A}$. As $\mathcal{A}$ is abelian, we have $u = \ker v$ for some $v\colon A \to B$. By hypothesis, $\ker v$ lies in $\mathcal{S}$; that is, $u = \ker v$. The dual argument shows that epimorphisms in $\mathcal{S}$ are cokernels. Therefore, $\mathcal{S}$ is abelian. •

We defined ***functor categories*** in Example B-4.97 in Part 1. If $\mathcal{C}$ is a category and $\mathcal{S}$ is a *small category* (i.e., $\operatorname{obj}(\mathcal{S})$ is a set), then $\mathcal{C}^{\mathcal{S}}$ is the category whose objects are the (covariant) functors $\mathcal{S} \to \mathcal{C}$ and whose morphisms are natural transformations.

Hom sets in a functor category are often denoted by $\operatorname{Nat}(\quad,\quad)$. The next result shows that $\operatorname{Nat}(F, G)$ is a set when $\mathcal{C}$ is an arbitrary category, $G\colon \mathcal{C} \to \mathbf{Sets}$, $A$ is an object in $\mathcal{C}$, and $F = \operatorname{Hom}_{\mathcal{C}}(A, \quad)$.

**Theorem C-4.8 (Yoneda Lemma).** *Let $\mathcal{C}$ be a category, let $A \in \operatorname{obj}(\mathcal{C})$, and let $G\colon \mathcal{C} \to \mathbf{Sets}$ be a covariant functor. There is a bijection*

$$y\colon \operatorname{Nat}(\operatorname{Hom}_{\mathcal{C}}(A, \quad), G) \to GA$$

*given by $y\colon \tau \mapsto \tau_A(1_A)$. Therefore, $\operatorname{Nat}(\operatorname{Hom}_{\mathcal{C}}(A, \quad), G)$ is a set.*

**Proof.** If $\tau\colon \operatorname{Hom}_{\mathcal{C}}(A, \quad) \to G$ is a natural transformation, then $y(\tau) = \tau_A(1_A)$ lies in the set $GA$, for $\tau_A\colon \operatorname{Hom}_{\mathcal{C}}(A, A) \to G(A)$. Thus, $y$ is a well-defined function.

For each $B \in \text{obj}(\mathcal{C})$ and $\varphi \in \text{Hom}_{\mathcal{C}}(A, B)$, there is a commutative diagram

$$\begin{array}{ccc} \text{Hom}_{\mathcal{C}}(A, A) & \xrightarrow{\tau_A} & GA \\ {\scriptstyle \varphi_*}\downarrow & & \downarrow{\scriptstyle G\varphi} \\ \text{Hom}_{\mathcal{C}}(A, B) & \xrightarrow[\tau_B]{} & GB, \end{array}$$

so that

$$(G\varphi)\tau_A(1_A) = \tau_B\varphi_*(1_A) = \tau_B(\varphi 1_A) = \tau_B(\varphi).$$

To see that $y$ is an injection, suppose that $\sigma\colon \text{Hom}_{\mathcal{C}}(A, \quad) \to G$ is another natural transformation. Then $\sigma_B(\varphi) = (G\varphi)\sigma_A(1_A)$. Hence, if $\sigma_A(1_A) = \tau_A(1_A)$, then $\sigma_B = \tau_B$ for all $B \in \text{obj}(\mathcal{C})$, and hence $\sigma = \tau$.

To see that $y$ is a surjection, take $x \in GA$. For $B \in \text{obj}(\mathcal{C})$ and $\psi \in \text{Hom}_{\mathcal{C}}(A, B)$, define

$$\tau_B(\psi) = (G\psi)(x)$$

(note that $G\psi\colon GA \to GB$, so that $(G\psi)(x) \in GB$). We claim that $\tau$ is a natural transformation; that is, if $\theta\colon B \to C$ is a morphism in $\mathcal{C}$, then the following diagram commutes:

$$\begin{array}{ccc} \text{Hom}_{\mathcal{C}}(A, B) & \xrightarrow{\tau_B} & GB \\ {\scriptstyle \theta_*}\downarrow & & \downarrow{\scriptstyle G\theta} \\ \text{Hom}_{\mathcal{C}}(A, C) & \xrightarrow[\tau_C]{} & GC. \end{array}$$

Going clockwise, we have $(G\theta)\tau_B(\psi) = G\theta G\psi(x)$; going counterclockwise, we have $\tau_C\theta_*(\psi) = \tau_C(\theta\psi) = G(\theta\psi)(x)$. Since $G$ is a functor, however, $G(\theta\psi) = G\theta G\psi$; thus, $\tau$ is a natural transformation. Now $y(\tau) = \tau_A(1_A) = G(1_A)(x) = x$, and so $y$ is a bijection. The last statement follows from $GA \in \text{obj}(\mathbf{Sets})$; that is, $GA$ is a set. •

Certainly the Hom functor is one of the most important functors. We are going to apply the Yoneda Lemma when $G = \text{Hom}_{\mathcal{C}}(B, \quad)$.

**Definition.** If $\mathcal{C}$ is a category, then a covariant functor $F\colon \mathcal{C} \to \mathbf{Sets}$ is ***representable*** if $F$ is naturally isomorphic to $\text{Hom}_{\mathcal{C}}(A, \quad)$ for some object $A \in \text{obj}(\mathcal{C})$.

The next theorem says that if a functor is representable, then the object $A$ is essentially unique. Note first that if $\tau\colon \text{Hom}_{\mathcal{C}}(A, \quad) \to \text{Hom}_{\mathcal{C}}(B, \quad)$ is a natural transformation, then $\tau_A\colon \text{Hom}_{\mathcal{C}}(A, A) \to \text{Hom}_{\mathcal{C}}(B, A)$, so that $\tau_A(1_A) \in \text{Hom}_{\mathcal{C}}(B, A)$ and $\tau_A(1_A)\colon B \to A$.

**Lemma C-4.9.** *Let $\mathcal{C}$ be a category with objects $A$ and $B$, let $\tau\colon \text{Hom}_{\mathcal{C}}(A, \quad) \to \text{Hom}_{\mathcal{C}}(B, \quad)$ be a natural transformation, and write*

$$\psi = \tau_A(1_A)\colon B \to A.$$

(i) *For every $C \in \text{obj}(\mathcal{C})$, we have $\tau_C = \psi^*\colon \text{Hom}_{\mathcal{C}}(A, C) \to \text{Hom}_{\mathcal{C}}(B, C)$; that is, $\psi^*\colon \varphi \mapsto \varphi\psi$.*

(ii) *The morphism $\psi$ is unique: if $\tau_C = \theta^*$, then $\theta = \psi$.*

(iii) *If* $\mathrm{Hom}_{\mathcal{C}}(A, \quad) \xrightarrow{\tau} \mathrm{Hom}_{\mathcal{C}}(B, \quad) \xrightarrow{\sigma} \mathrm{Hom}_{\mathcal{C}}(B', \quad)$ *are natural transformations,* $\tau_C = \psi^*$*, and* $\sigma_C = \eta^*$*, then for all* $C \in \mathrm{obj}(\mathcal{C})$ *we have*

$$(\sigma\tau)_C = (\psi\eta)^*.$$

**Proof.**

(i) The Yoneda Lemma says, for all $C \in \mathrm{obj}(\mathcal{C})$ and all $\varphi \in \mathrm{Hom}_{\mathcal{C}}(A, C)$, that $\tau_C(\varphi) = \varphi_*(\psi)$. But $\varphi_*(\psi) = \varphi\psi = \psi^*(\varphi)$.

(ii) Uniqueness follows from the Yoneda function $y$ being an injection.

(iii) For every $C \in \mathrm{obj}(\mathcal{C})$, there are, by (ii), unique morphisms $\psi \in \mathrm{Hom}_{\mathcal{C}}(B, A)$ and $\eta \in \mathrm{Hom}_{\mathcal{C}}(B', B)$ with

$$\tau_C(\varphi) = \psi^*(\varphi) \quad \text{and} \quad \sigma_C(\varphi') = \eta^*(\varphi')$$

for all $\varphi \in \mathrm{Hom}_{\mathcal{C}}(A, C)$ and $\varphi' \in \mathrm{Hom}_{\mathcal{C}}(B, C)$. By definiton, $(\sigma\tau)_C = \sigma_C\tau_C$, and so

$$(\sigma\tau)_C(\varphi) = \sigma_C(\psi^*(\varphi)) = \eta^*\psi^*(\varphi) = (\psi\eta)^*(\varphi). \quad \bullet$$

**Theorem C-4.10.** *Let* $\mathcal{C}$ *be a category with objects* $A$ *and* $B$*. If* $\tau\colon \mathrm{Hom}_{\mathcal{C}}(A, \quad) \to \mathrm{Hom}_{\mathcal{C}}(B, \quad)$ *is a natural isomorphism, then* $A \cong B$*.*

**Proof.** Since $\tau$ is a natural isomorphism, there exists a natural transformation $\sigma\colon \mathrm{Hom}_{\mathcal{C}}(B, \quad) \to \mathrm{Hom}_{\mathcal{C}}(A, \quad)$ with $\sigma\tau$ and $\tau\sigma$ the identity natural transformations of $\mathrm{Hom}_{\mathcal{C}}(B, \quad)$ and $\mathrm{Hom}_{\mathcal{C}}(B, \quad)$, respectively. By Lemma C-4.9, there are morphisms $\psi\colon B \to A$ and $\eta\colon A \to B$ with $\tau_C = \psi^*$ and $\sigma_C = \eta^*$ for all $C \in \mathrm{obj}(\mathcal{C})$. Moreover, part (iii) of the lemma gives

$$\tau\sigma = (\eta\psi)^* = 1_B^* \quad \text{and} \quad \sigma\tau = (\psi\eta)^* = 1_A^*.$$

Part (ii) of Lemma C-4.9, uniqueness, gives $\eta\psi = 1_B$ and $\psi\eta = 1_A$, so that $\eta\colon A \to B$ is an isomorphism. $\bullet$

Here are more examples of abelian categories.

**Proposition C-4.11.** *If* $\mathcal{A}$ *is an abelian category and* $\mathcal{S}$ *is a small category, then the functor category* $\mathcal{A}^{\mathcal{S}}$ *is an abelian category.*

**Proof.** We assume that $\mathcal{S}$ is small to guarantee that $\mathcal{A}^{\mathcal{S}}$ is a category (Example B-4.97 in Part 1).

It is straightforward to check that $\mathcal{A}^{\mathcal{S}}$, with the following definitions, is an additive category. The zero object in $\mathcal{A}^{\mathcal{S}}$ is the constant functor with value 0, where 0 is a zero object in $\mathcal{A}$. If $\tau, \sigma \in \mathrm{Hom}(F, G) = \mathrm{Nat}(F, G)$, where $F, G\colon \mathcal{S} \to \mathcal{A}$ are functors, define $\tau + \sigma\colon F \to G$ by $(\tau + \sigma)_S = \tau_S + \sigma_S\colon FS \to GS$ for all $S \in \mathrm{obj}(\mathcal{S})$. Finally, define $F \oplus G$ by $(F \oplus G)S = FS \oplus GS$.

Let $\tau\colon F \to G$ be a natural transformation. For each object $S$ in $\mathcal{S}$, there is a morphism $\tau_S\colon F \to GS$ in $\mathcal{A}$. Since $\mathcal{A}$ is abelian, $\ker(\tau_S) = [\iota_S]$ exists; there is an object $K_S$ and a morphism $\iota_S\colon K_S \to FS$ solving the universal mapping problem for kernel, part of which says there is a commutative diagram involving $\tau_S$ and $\iota_S$.

In the following commutative diagram with exact rows, where $f\colon S \to S'$ in $\mathcal{S}$, there is a unique $Kf\colon KS \to KS'$ making the augmented diagram commute:

$$\begin{array}{ccccccc} 0 & \longrightarrow & KS & \xrightarrow{\iota_S} & FS & \xrightarrow{\tau_S} & GS \\ & & \downarrow{\scriptstyle Kf} & & \downarrow{\scriptstyle Ff} & & \downarrow{\scriptstyle Gf} \\ 0 & \longrightarrow & KS' & \xrightarrow[\iota_{S'}]{} & FS' & \xrightarrow[\tau_{S'}]{} & GS'. \end{array}$$

It follows that $[i] = (i_S)_{S\in\mathrm{obj}(\mathcal{A})}$ determines a functor $K \in \mathrm{obj}(\mathcal{A}^{\mathcal{S}})$ with $[i] = \ker(\tau)$. The dual construction displays $\mathrm{coker}(\tau)$. The reader may check that $K$ is a functor, $\iota\colon K \to F$ is a natural transformation, and $\ker\tau = [\iota]$; dually, cokernels exist in $\mathcal{A}^{\mathcal{S}}$. Verification of the various details is routine. •

**Corollary C-4.12.** *If $\mathcal{A}$ is an abelian category and $\mathcal{U}$ is a small category, then a sequence*

$$0 \to F' \xrightarrow{\tau} F \xrightarrow{\sigma} F'' \to 0$$

*is exact in $\mathcal{A}^{\mathcal{U}}$ if and only if*

$$0 \to F'(U) \xrightarrow{\tau_U} F(U) \xrightarrow{\sigma_U} F''(U) \to 0$$

*is exact in $\mathcal{A}$ for every $U \in \mathrm{obj}(\mathcal{U})$.*

**Proof.** The proof of Proposition C-4.11 describes $\ker\sigma$ and $\mathrm{im}\,\tau$. •

Here is a special case of Corollary C-4.12. Recall that the topology $\mathcal{U}$ on a space $X$, being a poset under inclusion, can be viewed as a category. Following Mac Lane, we denote $\mathcal{U}$ by

$$\mathbf{Open}(X).$$

**Definition.** A ***presheaf over $X$ in* Ab** is a contravariant functor $\mathbf{Open}(X) \to \mathbf{Ab}$. The category of presheaves of abelian groups over a space $X$ is denoted by

$$\mathbf{pSh}(X, \mathbf{Ab}) = \mathbf{Ab}^{\mathbf{Open}(X)^{\mathrm{op}}}.$$

Let $\mathcal{P}$ and $\mathcal{P}'$ be presheaves (of abelian groups) over a space $X$, with restriction maps $\rho_U^V\colon \mathcal{P}(V) \to \mathcal{P}(U)$ and $\tau_U^V\colon \mathcal{P}'(V) \to \mathcal{P}'(U)$ whenever $U \subseteq V$ are open. A ***presheaf map*** $\varphi\colon \mathcal{P} \to \mathcal{P}'$ is a natural transformation; that is, $\varphi$ is a one-parameter family of morphisms $\varphi_U\colon \mathcal{P}(U) \to \mathcal{P}'(U)$, indexed by $U \in \mathbf{Open}(X)$, such that there is a commutative diagram whenever $U \subseteq V$:

$$\begin{array}{ccc} \mathcal{P}(V) & \xrightarrow{\varphi_V} & \mathcal{P}'(V) \\ {\scriptstyle \rho_U^V}\downarrow & & \downarrow{\scriptstyle \tau_U^V} \\ \mathcal{P}(U) & \xrightarrow[\varphi_U]{} & \mathcal{P}'(U). \end{array}$$

**Corollary C-4.13.** *For every topological space $X$, the category*

$$\mathbf{pSh}(X, \mathbf{Ab})$$

*of all presheaves over $X$ is an abelian category.*

**Proof.** This is a special case of Proposition C-4.11. •

**Corollary C-4.14.** *A sequence of presheaves over $X$ in* **Ab**

$$0 \to \mathcal{P}' \xrightarrow{\tau} \mathcal{P} \xrightarrow{\sigma} \mathcal{P}'' \to 0$$

*is exact if and only if*

$$0 \to \mathcal{P}'(U) \xrightarrow{\tau_U} \mathcal{P}(U) \xrightarrow{\sigma_U} \mathcal{P}''(U) \to 0$$

*is exact in* **Ab** *for every* $U \in \mathrm{obj}(\mathbf{Open}(X))$.

**Proof.** Corollary C-4.12 applies, for contravariant functors become covariant when we exchange $\mathbf{Open}(X)$ with its opposite category. •

For any ring $R$, we have defined the category ${}_R\mathbf{Comp}$ having objects all complexes, that is, sequences of left $R$-modules

$$\cdots \to A_{n+1} \xrightarrow{d_{n+1}} A_n \xrightarrow{d_n} A_{n-1} \to \cdots$$

for which the composite of adjacent morphisms is 0,

$$d_n d_{n+1} = 0 \quad \text{for all } n \in \mathbb{Z},$$

and whose morphisms are chain maps.

We generalize this definition by replacing ${}_R\mathbf{Mod}$ by any abelian category $\mathcal{A}$. Define the category of ***complexes over*** $\mathcal{A}$,

$$\mathbf{Comp}(\mathcal{A}),$$

to be the category whose objects are complexes (whose terms $A_n$ and differentiations $d_n$ lie in $\mathcal{A}$) and whose morphisms are chain maps.

We can realize $\mathbf{Comp}(\mathcal{A})$ as a full subcategory of the functor category $\mathcal{A}^{\mathcal{S}}$ (where $\mathcal{S} = \mathbf{PO}(\mathbb{Z})^{\mathrm{op}}$ is the opposite category of $\mathbf{PO}(\mathbb{Z})$, the partially ordered set $\mathbb{Z}$ with the usual order).[2] We saw, in Exercise C-3.21 on page 261, that the objects of $\mathcal{A}^{\mathbf{PO}(\mathbb{Z})^{\mathrm{op}}}$ are contravariant functors, namely, sequences $\mathfrak{A} = \cdots A_{n+1} \xrightarrow{d_{n+1}} A_n \xrightarrow{d_n} A_{n-1} \cdots$ in $\mathcal{A}$, and morphisms $\mathfrak{A} \to \mathfrak{B}$, where $\mathfrak{B} = \cdots B_{n+1} \xrightarrow{\delta_{n+1}} B_n \xrightarrow{\delta_n} B_{n-1} \cdots$, are natural transformations, namely, sequences $(f_n \colon A_n \to B_n)_{n\in\mathbb{Z}}$ giving commutative diagrams, i.e., chain maps. Thus, every complex is an object in $\mathcal{A}^{\mathcal{S}}$, every chain map is a morphism, and $\mathbf{Comp}(\mathcal{A})$ is the full subcategory of $\mathcal{A}^{\mathcal{S}}$ generated by the complexes over $\mathcal{A}$.

**Theorem C-4.15.** *If $\mathcal{A}$ is an abelian category, then* $\mathbf{Comp}(\mathcal{A})$ *is also an abelian category.*

**Proof.** Since $\mathbf{PO}(\mathbb{Z})^{\mathrm{op}}$ is a small category (i.e., $\mathrm{obj}(\mathbf{PO}(\mathbb{Z})^{\mathrm{op}})$ is a set), Proposition C-4.11 says that the functor category $\mathcal{A}^{\mathbf{PO}(\mathbb{Z})^{\mathrm{op}}}$ is abelian. Proposition C-4.7 now says that the full subcategory $\mathbf{Comp}(\mathcal{A})$ is abelian if it satisfies several conditions.

(i) The *zero complex* is the complex each of whose terms is 0.

(ii) The *direct sum* $(\mathbf{C}, d) \oplus (\mathbf{C}', d')$ is the complex whose $n$th term is $C_n \oplus C'_n$ and whose $n$th differential is $d_n \oplus d'_n$.

[2] A contravariant functor $\mathcal{C} \to \mathcal{D}$ is the same thing as a covariant functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$. Of course, $n - 1 \leq n$ in $\mathbb{Z}$, so that a contravariant functor $d_\bullet \colon \mathbb{Z} \to \mathcal{A}$ has $d_n \colon A_n \to A_{n-1}$ for all $n \in \mathbb{Z}$.

(iii) If $f = (f_n)\colon (\mathbf{C}, d) \to (\mathbf{C}', d')$ is a chain map, define

$$\mathbf{ker}\, f = \to \ker f_{n+1} \xrightarrow{\delta_{n+1}} \ker f_n \xrightarrow{\delta_n} \ker f_{n-1} \to,$$

where $\delta_n = d_n | \ker f_n$, and

$$\mathbf{im}\, f = \to \operatorname{im} f_{n+1} \xrightarrow{k_{n+1}} \operatorname{im} f_n \xrightarrow{k_n} \operatorname{im} f_{n-1} \to,$$

where $k_n = d'_n | \operatorname{im} f_n$.

Since these complexes lie in the full subcategory $\mathbf{Comp}(\mathcal{A})$, Proposition C-4.7 applies to prove the theorem. •

Sheaves will be introduced in the next sections, and we shall prove that they, too, form an abelian category.

We end this section by describing the Full Imbedding Theorem, which says, for all intents and purposes, that working in abelian categories is the same as working in **Ab**. The thrust of the next theorem is that it allows us to do diagram chasing in abelian categories.

**Definition.** Let $\mathcal{C}, \mathcal{D}$ be categories, and let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor. Then $F$ is ***faithful*** if, for all $A, B \in \operatorname{obj}(\mathcal{C})$, the functions $\operatorname{Hom}_{\mathcal{C}}(A, B) \to \operatorname{Hom}_{\mathcal{D}}(FA, FB)$, given by $f \mapsto Ff$, are injections; $F$ is ***full*** if these functions are surjective.

If $\mathcal{A}$ is an abelian category, then a functor $F\colon \mathcal{A} \to \mathbf{Ab}$ is ***exact*** if $A' \to A \to A''$ exact in $\mathcal{A}$ implies $FA' \to FA \to FA''$ exact in **Ab**.

**Theorem C-4.16 (Freyd–Haron[3]–Lubkin).**

*If $\mathcal{A}$ is a small abelian category, then there is a covariant faithful exact functor $F\colon \mathcal{A} \to \mathbf{Ab}$.*

**Proof.** See Freyd [**69**], Chapter 7, or Mitchell [**157**], p. 101. •

This imbedding theorem can be improved so that its image is a *full* subcategory of **Ab**.

**Theorem C-4.17 (Mitchell's Full Imbedding Theorem).** *If $\mathcal{A}$ is a small abelian category, then there is a covariant full faithful exact functor $F\colon \mathcal{A} \to \mathbf{Ab}$.*

**Proof.** Mitchell [**157**], p. 151. •

In [**157**], Mitchell writes, "Let us say that a statement about a diagram in an abelian category is ***categorical*** if it states that certain parts of the diagram are or are not commutative, that certain sequences in the diagram are or are not exact, and that certain parts of the diagram are or are not (inverse) limits or (direct) limits. Then we have the following metatheorem."

[3]Katie Guest, Karen Langdon, and Sophie Quantrell, librarians at the University of Oxford, were able to locate the (unpublished) 1960 Oxford thesis of Haron, *Problems in Homological Algebra.*

**Metatheorem.** *Let $\mathcal{A}$ be a (not necessarily small) abelian category.*

(i) *Let $\Sigma$ be a statement of the form p implies q, where p and q are categorical statements about a diagram in $\mathcal{A}$. If $\Sigma$ is true in* **Ab**, *then $\Sigma$ is true in $\mathcal{A}$.*

(ii) *Let $\Sigma'$ be a statement of the form p implies q, where p is a categorical statement concerning a diagram in $\mathcal{A}$, while q states that additional morphisms exist between certain objects in the diagram and that some categorical statement is true of the extended diagram. If the statement can be proved in* **Ab** *by constructing the additional morphisms through diagram chasing, then the statement is true in $\mathcal{A}$.*

**Proof.** See Mitchell [**157**], p. 97. •

Part (i) follows from the Freyd–Haron–Lubkin Imbedding Theorem. To illustrate, the Five Lemma is true in **Ab**, as is the $3 \times 3$ Lemma (Exercise B-1.58 on page 310 in Part 1), and so they are true in every abelian category.

Part (ii) follows from Mitchell's Full Imbedding Theorem. To illustrate, recall Proposition B-1.46 in Part 1: given a commutative diagram of abelian groups with exact rows,

$$\begin{array}{ccccccc} A' & \xrightarrow{i} & A & \xrightarrow{p} & A'' & \longrightarrow & 0 \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ B' & \xrightarrow[j]{} & B & \xrightarrow[q]{} & B'' & \longrightarrow & 0, \end{array}$$

there exists a unique map $h\colon A'' \to B''$ making the augmented diagram commute. Suppose now that the diagram lies in an abelian category $\mathcal{A}$. Applying the imbedding functor $F\colon \mathcal{A} \to \mathbf{Ab}$ of the Full Imbedding Theorem, we have a diagram in **Ab** as above, and so there is a homomorphism in **Ab**, say, $h\colon FA'' \to FB''$, making the diagram commute: $F(q)F(g) = hF(p)$. Since $F$ is a full imbedding, there exists $\eta \in \mathrm{Hom}_{\mathcal{A}}(A'', B'')$ with $h = F(\eta)$; hence, $F(qg) = F(q)F(g) = hF(p) = F(\eta)F(p) = F(\eta p)$. But $F$ is faithful, so that $qg = \eta p$.

The following definition allows us to say when two categories are the same; it will be discussed more thoroughly in Section C-4.6.

**Definition.** A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an ***equivalence*** if there is a functor $G\colon \mathcal{D} \to \mathcal{C}$, called its ***inverse***, such that $GF$ and $FG$ are naturally isomorphic to the identity functors $1_{\mathcal{C}}$ and $1_{\mathcal{D}}$, respectively. When $\mathcal{C}$ and $\mathcal{D}$ are additive categories, we will further assume that an equivalence $F\colon \mathcal{C} \to \mathcal{D}$ is an additive functor.

**Proposition C-4.18.** *A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an equivalence if and only if*

(i) *$F$ is full and faithful: i.e., the function $\mathrm{Hom}_{\mathcal{C}}(C, C') \to \mathrm{Hom}_{\mathcal{D}}(FC, FC')$, given by $f \mapsto Ff$, is a bijection for all $C, C' \in \mathrm{obj}(\mathcal{C})$;*

(ii) *every $D \in \mathrm{obj}(\mathcal{D})$ is isomorphic to $FC$ for some $C \in \mathrm{obj}(\mathcal{C})$.*

**Proof.** Assume that $F$ is an equivalence. Given a morphism $f\colon C \to C'$ in $\mathcal{C}$, there is a commutative diagram

$$\begin{array}{ccc} GFC & \xrightarrow{\tau_C} & C \\ {\scriptstyle GFf}\downarrow & & \downarrow{\scriptstyle f} \\ GFC' & \xrightarrow[\tau_{C'}]{} & C'. \end{array}$$

Since $\tau$ is a natural isomorphism, each $\tau_C$ is an isomorphism; hence,

$$f = \tau_{C'}(GFf)\tau_C^{-1}. \tag{1}$$

$F$ is faithful: if $f, f' \in \mathrm{Hom}_{\mathcal{C}}(C, C')$ and $Ff' = Ff$ in $\mathrm{Hom}_{\mathcal{D}}(FC, FC')$, then

$$f' = \tau_{C'}(GFf')\tau_C^{-1} = \tau_{C'}(GFf)\tau_C^{-1} = f.$$

Similarly, $FG \cong 1_{\mathcal{D}}$ implies that $G$ is faithful. We have proved (i).

We claim that $F$ is full. If $g\colon FC \to FC'$, define a morphism $f = \tau_{C'}(Gg)\tau_C^{-1}$. Now $f = \tau_{C'}(GFf)\tau_C^{-1}$, by Eq. (1), so that $GFf = Gg$. Since $G$ is faithful, we have $Ff = g$. If $F$ is an equivalence, then there exists a functor $G\colon \mathcal{D} \to \mathcal{C}$ with $GF \cong 1_{\mathcal{C}}$ and $FG \cong 1_{\mathcal{D}}$; let $\tau\colon GF \to 1_{\mathcal{C}}$ and $\sigma\colon FG \to 1_{\mathcal{D}}$ be natural isomorphisms. For each $D \in \mathrm{obj}(\mathcal{D})$, there is an isomorphism $\sigma_D\colon FGD \to D$. Thus, if we define $C = GD$, then $FC \cong D$, which proves (ii).

Conversely, assume that $F\colon \mathcal{C} \to \mathcal{D}$ satisfies (i) and (ii). For each $D \in \mathrm{obj}(\mathcal{D})$, part (ii) gives a unique $C = C_D \in \mathrm{obj}(\mathcal{C})$ with $D \cong FC_D$; choose an isomorphism $h_D\colon D \to FC_D$. Define a functor $G\colon \mathcal{D} \to \mathcal{C}$ on objects by $GD = C_D$. If $g\colon D \to D'$ is a morphism in $\mathcal{D}$, (i) gives a unique morphism $f\colon C_D \to C_{D'}$ with $Ff = h_{D'}gh_D^{-1}$. It is routine to check that $G$ is a functor, $GF \cong 1_{\mathcal{C}}$, and $FG \cong 1_{\mathcal{D}}$. Therefore, $F$ is an equivalence. •

Note Exercise C-4.19 below. If $\mathcal{A}$ and $\mathcal{B}$ are equivalent categories, then $\mathcal{A}$ abelian implies $\mathcal{B}$ abelian.

## Exercises

∗ **C-4.7.** (i) Prove that a function is epic in **Sets** if and only if it is surjective and that a function is monic in **Sets** if and only if it is injective.

(ii) Prove that an $R$-map is epic in ${}_R\mathbf{Mod}$ if and only if it is surjective and that an $R$-map is monic in ${}_R\mathbf{Mod}$ if and only if it is injective.

(iii) Prove that every object in **Sets** is projective and injective.

∗ **C-4.8.** Let $\mathcal{C}$ be the category of all divisible abelian groups.

(i) Prove that the natural map $\mathbb{Q} \to \mathbb{Q}/\mathbb{Z}$ is monic in $\mathcal{C}$.

(ii) Conclude that $\mathcal{C}$ is a category in which monomorphisms and injections do not coincide.

∗ **C-4.9.** Prove, in every abelian category, that the injections of a coproduct are monic and the projections of a product are epic.

∗ **C-4.10.** (i) Prove that every isomorphism in an additive category is both monic and epic.

(ii) Let $R$ be a domain that is not a field, and let $\varphi\colon R \to \mathrm{Frac}(R)$ be given by $r \mapsto r/1$. In **ComRings**, prove that $\varphi$ is both monic and epic but that $\varphi$ is not an isomorphism.

(iii) Prove that a morphism in an abelian category is an isomorphism if and only if it is both monic and epic.

∗ **C-4.11. (Eilenberg–Moore)** Let $G$ be a (possibly nonabelian) group.

(i) If $H$ is a proper subgroup of a group $G$, prove that there exist a group $L$ and distinct homomorphisms $f, g\colon G \to L$ with $f|H = g|H$.
**Hint.** Define $L = S_X$, where $X$ denotes the family of all the left cosets of $H$ in $G$ together with an additional element, denoted $\infty$. If $a \in G$, define $f(a) = f_a \in S_X$ by $f_a(\infty) = \infty$ and $f_a(bH) = abH$. Define $g\colon G \to S_X$ by $g = \gamma f$, where $\gamma \in S_X$ is conjugation by the transposition $(H, \infty)$.

(ii) Prove that a homomorphism $\varphi\colon A \to G$, where $A$ and $G$ are groups, is surjective if and only if it is an epimorphism in **Groups**.

**C-4.12.** Prove that every abelian category is an exact category in the sense of Quillen (see Example C-4.6(iv)).

**C-4.13.** State and prove the ***First Isomorphism Theorem*** in an abelian category $\mathcal{A}$.

∗ **C-4.14.** If $\mathcal{A}$ is an abelian category, prove that a morphism $f = (f_n)$ in $\mathrm{Comp}(\mathcal{A})$ (i.e., a chain map) is monic (or epic) if and only if each $f_n$ is monic (or epic) in $\mathcal{A}$.

∗ **C-4.15.** Let $\mathcal{T}$ be the full subcategory of **Ab** consisting of all torsion abelian groups.

(i) Prove that $\mathcal{T}$ is an abelian category having no nonzero projective objects, and conclude that it is not equivalent to a category of modules.

(ii) If $(T_i)_{i\in I}$ is a (possibly infinite) family of torsion abelian groups, prove that their categorical product $\prod_{i\in I} T_i$ exists in $\mathcal{T}$.
**Hint.** Try the torsion subgroup of the product in **Ab**.

(iii) Prove that $\mathcal{T}'$, the abelian category of all finitely generated abelian groups, is an abelian category that has no nonzero injectives.

∗ **C-4.16.** (i) Let $\mathcal{C}$ be the category of all divisible abelian groups. Prove that the natural map $\mathbb{Q} \to \mathbb{Q}/\mathbb{Z}$ is monic in $\mathcal{C}$. Conclude that $\mathcal{C}$ is a category whose morphisms are functions and in which monomorphisms and injections do not coincide.

(ii) Let $\mathbf{Top}_2$ be the category of all Hausdorff spaces. If $D \subsetneq X$ is a dense subspace of a space $X$, prove that the inclusion $i\colon D \to X$ is an epimorphism. Conclude that $\mathbf{Top}_2$ is a category whose morphisms are functions and in which epimorphisms and surjections do not coincide.
**Hint.** Two continuous functions agreeing on a dense subspace of a Hausdorff space must be equal.

**C-4.17.** Let $\mathcal{S}$ be a full subcategory of an abelian category $\mathcal{A}$ which satisfies the hypotheses of Proposition C-4.7. Prove that if $A \xrightarrow{f} B \xrightarrow{g} C$ is an exact sequence in $\mathcal{S}$, then it is an exact sequence in $\mathcal{A}$.

**C-4.18.** If $F\colon \mathcal{A} \to \mathcal{B}$ is an equivalence of abelian categories, prove that if $0 \to A' \to A \to A'' \to 0$ is an exact sequence in $\mathcal{A}$, then $0 \to FA' \to FA \to FA'' \to 0$ is an exact sequence in $\mathcal{B}$.

∗ **C-4.19.** Let $\mathcal{A}$ and $\mathcal{B}$ be additive categories, and let $F\colon \mathcal{A} \to \mathcal{B}$ be an additive functor. Prove that if $F$ is an equivalence and $\mathcal{A}$ is abelian, then $\mathcal{B}$ is abelian.

## C-4.3. g-Sheaves

One of the main aims of this book is to prepare the reader to do more sophisticated work in algebra, and sheaves are a fundamental tool in algebraic geometry. For example, *schemes* are the modern generalization of *varieties*, and sheaves are an essential ingredient in the very definition of scheme: just as topological manifolds are constructed by gluing together coordinate patches, schemes are constructed by gluing together sheaves.

At the beginning of his book [**217**], *The Theory of Sheaves*, Swan writes, "What are sheaves good for? The obvious answer is that sheaves are very useful in proving theorems." He then lists interesting applications of sheaves to algebraic topology, complex variables, and algebraic geometry, and concludes, "the importance of the theory of sheaves is simply that it gives relations (quite strong relations, in fact) between the local and global properties of a space."

Covering spaces are geometric constructs in topology (we used an algebraic version of them in Chapter C-1 to prove some nice theorems in group theory). Recall that a ***covering space*** is a triple $(E, p, X)$, where $p\colon E \to X$ is a continuous surjection between topological spaces $E$ and $X$, in which each $x \in X$ has an open neighborhood $U$ such that $p^{-1}(U) = \bigcup_i \Sigma_i$, a disjoint union of open subsets $\Sigma_i$ of $E$ with $p|\Sigma_i\colon \Sigma_i \to U$ a homeomorphism for each $i$. For example, $(\mathbb{R}, p, S^1)$, where $S^1$ is the unit circle, is a covering space, where $p\colon \mathbb{R} \to S^1$ is defined by $x \mapsto e^{2\pi i x}$. Sheaves, which generalize covering spaces, share the same picture (see Figure C-4.4).

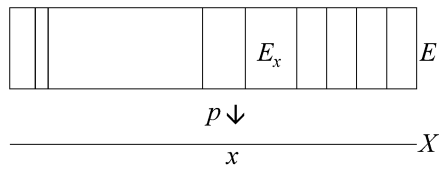



**Figure C-4.4.** g-Sheaf.

**Definition.** A continuous surjective map $p\colon E \to X$ between topological spaces $E$ and $X$ is called a ***local homeomorphism*** if, for each $e \in E$, there is an open neighborhood $\Sigma$ of $e$, called a ***sheet***, with $p(\Sigma)$ open in $X$ and $p|\Sigma\colon \Sigma \to p(\Sigma)$ a homeomorphism.

A triple

$$\mathcal{G} = (E, p, X),$$

where $p\colon E \to X$ is a local homeomorphism, is called a ***g-sheaf***[4] ***of abelian groups over*** $X$ if

(i) the *fiber* $p^{-1}(x)$, denoted by $E_x$ and called the ***stalk over*** $x$, is an abelian group for each $x \in X$;

(ii) inversion $\iota\colon E \to E$, given by $e \mapsto -e$, and addition $\alpha\colon E + E \to E$, given by $(e, e') \mapsto e + e'$, are continuous, where $e, e'$ lie in the same stalk $E_x$; that is,

$$(e, e') \in E + E = \{(e, e') \in E \times E : p(e) = p(e')\} = \bigcup_{x \in X} (E_x \times E_x) \subseteq E \times E.$$

Each of the ingredients of a g-sheaf has a name. The space $E$ is called the ***sheaf space*** (or étale space), $X$ is called the ***base space***, and the map $p\colon E \to X$ is called the ***projection***.

The meaning of continuity of inversion $e \mapsto -e$ is clear, but we elaborate on the definition of continuity of addition. Let $(e, e') \in E + E$; if $W \subseteq E$ is an open set containing $e + e'$, then there exists an open set $O \subseteq E \times E$ containing $(e, e')$, with $\alpha(O \cap (E + E)) \subseteq W$.

The definition of g-sheaf can be modified so that its stalks lie in algebraic categories other than **Ab**, such as ${}_R\mathbf{Mod}$ or **ComRings**. Of course, axiom (ii) is modified so that all the algebraic operations are continuous. Even though most results in this section hold in more generality, we assume throughout that stalks are merely abelian groups.

**Remark.** A g-sheaf is reminiscent of a manifold. For example, the 2-sphere $S^2$ can be viewed as a union of "patches", each homeomorphic to $\mathbb{R}^2$. More formally, there is a continuous map $p\colon S^2 \to \mathbb{R}^2$ with each patch homeomorphic to an open disk in $\mathbb{R}^2$; that is, we may think of patches as sheets. The main difference between manifolds and g-sheaves is that the fibers $p^{-1}(x)$ of a manifold need not have any algebraic structure; in particular, they need not be abelian groups. ◀

Here are some examples of g-sheaves.

**Example C-4.19.**

(i) If $X$ is a topological space and $G$ is a discrete topological abelian group, define $E = X \times G$, and define $p\colon E \to X$ by $p\colon (x, y) \mapsto x$. If $e = (x, y) \in E$ and $V$ is an open neighborhood of $x$, then $\Sigma = V \times \{y\}$ is an open neighborhood of $e$ (because $\{y\}$ is open in $G$) and $p|\Sigma\colon \Sigma \to V = p(\Sigma)$ is a homeomorphism. The triple $\mathcal{G} = (E, p, X)$ is called the ***constant* g-*sheaf at*** $G$. We denote

[4] In my homological algebra book [**187**], I used the term *etale-sheaf*, which is confusing, for *etale-sheaf* (or *sheaf space*) is a standard term in algebraic geometry having a different definition. I now use the term g-*sheaf*, where the letter g is to remind the reader of the adjective *geometric*.

the constant g-sheaf at $G$ by

$$\mathcal{G} = G^{\mathrm{g}}.$$

In particular, if $G = \{0\}$, then the constant g-sheaf $\{0\}^{\mathrm{g}}$ is called the ***zero g-sheaf***.

(ii) If $G$ is an abelian topological group and $H$ is a discrete subgroup of $G$, then $(G, p, G/H)$ is a covering space, where $p$ is the natural map.

(iii) Every covering space is a g-sheaf if its fibers are abelian groups.

(iv) The covering space $(\mathbb{R}, \exp, S^1)$ gives rise to an example of a g-sheaf which is not a covering space, for its fibers are not abelian groups; for example, $(\exp)^{-1}(\frac{1}{2})$ is not a group because it does not contain the identity element 0 of $\mathbb{R}$.[5]

(v) A ***vector bundle*** is a g-sheaf $(E, p, X)$ with extra properties: there is a field $k$ and every stalk $E_x$ is a finite-dimensional vector space over $k$; for each sheet $\Sigma_U$ over an open set $U$, there is a commutative diagram

$$\begin{array}{ccc} U \times E_x & \xrightarrow{\varphi_U} & p^{-1}(U) \\ & {\scriptstyle\pi}\searrow \quad \swarrow{\scriptstyle p} & \\ & U & \end{array}$$

where $\varphi_U \colon U \times E_x \to p^{-1}(U)$ is a homeomorphism and $\pi$ is the projection onto the first factor. It follows that all the stalks are isomorphic vector spaces; their common dimension is call the ***rank*** of the vector bundle. A ***line bundle*** is a vector bundle of rank 1.

(vi) If $R$ is a commutative ring, its ***structure sheaf*** is $(E, p, \mathrm{Spec}(R))$ whose stalk $E_{\mathfrak{p}}$ for a prime ideal $\mathfrak{p}$ is the *localization* $R_{\mathfrak{p}}$, and $p \colon E \to \mathrm{Spec}(R)$ sends $R_{\mathfrak{p}} \mapsto \mathfrak{p} \in \mathrm{Spec}(R)$ (we discuss this example, along with localization, in Chapter C-5 on commutative rings). In contrast to vector bundles, the stalks are not isomorphic. The structure sheaf of $R$ has base space $\mathrm{Spec}(R)$ with the Zariski topology, sheaf space $E = \bigcup_{\mathfrak{p} \in \mathrm{Spec}(R)} R_{\mathfrak{p}}$ suitably topologized, and projection $p \colon E \to \mathrm{Spec}(R)$ defined by $p(e) = \mathfrak{p}$ for all $e \in R_{\mathfrak{p}}$. ◀

Here are some properties of g-sheaves.

**Proposition C-4.20.** *Let $\mathcal{G} = (E, p, X)$ be a g-sheaf.*

(i) *The sheets form a base of open sets for $E$.*

(ii) *$p$ is an open map.*

(iii) *Each stalk $E_x$ is discrete.*

(iv) (**Gluing**) *Let $(U_i)_{i \in I}$ be a family of open subsets of $X$ and, for $(i, j) \in I \times I$, define $U_{(i,j)} = U_i \cap U_j$. If $(s_i \colon U_i \to E)_{i \in I}$ are continuous maps satisfying $s_i | U_{(i,j)} = s_j | U_{(i,j)}$ for all $(i, j) \in I \times I$, then there exists a unique continuous $s \colon U \to E$ with $s | U_i = f_i$ for all $i \in I$.*

[5] Every nonempty set is the underlying set of an abelian group, but the natural candidate here would be a subgroup of $\mathbb{R}$.

(v) (**Uniqueness**) *Let* $(U_i)_{i\in I}$ *be a family of open subsets of* $X$*, and let* $U = \bigcup_{i\in I} U_i$*. If* $s, s' \colon U \to E$ *and* $s|U_i = s'|U_i$ *for all* $i \in I$*, then* $s = s'$*.*

**Proof.**

(i) Since, for each $e \in E$, there is a sheet $\Sigma$ containing $e$, the sheaf space $E$ is the union of all the sheets: $E = \bigcup_\Sigma \Sigma$. If $U \subseteq E$ is open, then $U \cap \Sigma$ is open for every sheet $\Sigma$, and so $U = \bigcup_\Sigma (U \cap \Sigma)$. But every open subset of a sheet is also a sheet, and so $U$ is a union of sheets; that is, the sheets comprise a base for the topology of $E$.

(ii) If $U \subseteq E$ is open, then $p(U) = \bigcup_\Sigma p(U \cap \Sigma)$. But $p(U \cap \Sigma)$ is open in $X$, because $p$ is a local homeomorphism; thus, $p(U)$ is open, for it is a union of open sets.

(iii) Let $e \in E_x$, and let $\Sigma$ be a sheet containing $e$. If $e' \in E_x$ and $e' \neq e$, then $e' \notin \Sigma$, for $p|\Sigma$ is injective and $p(e') = x = p(e)$. Therefore, $\Sigma \cap E_x = \{e\}$, and so $E_x$ is discrete.

(iv) If $x \in U$, then $x \in U_i$ for some $i$; define $s \colon U \to E$ by $s(x) = s_i(x)$. The condition on overlaps $U_{(i,j)}$ shows that $s$ is a well-defined function; it is obviously the unique function $U \to E$ satisfying $s|U_i = s_i$ for all $i \in I$.

We prove continuity of $s$. If $V$ is an open subset of $E$, then $s^{-1}(V) = U \cap s^{-1}(V) = (\bigcup_i U_i) \cap s^{-1}(V) = \bigcup_i (U_i \cap s^{-1}(V)) = \bigcup_i f_i^{-1}(V)$. Continuity of $s_i$ says that $s_i^{-1}(V)$ is open in $U_i$ for all $i$, hence is open in $U$; thus, $s^{-1}(V)$ is open in $U$, and $s$ is continuous.

(v) If $x \in U$, then $x \in U_i$ for some $i$, and $s(x) = (s|U_i)x = (s'|U_i)x = s'(x)$. Hence, $s = s'$. •

The last two parts of Proposition C-4.20 suggest the following definition.

**Definition.** If $\mathcal{G} = (E, p, X)$ is a g-sheaf of abelian groups and $U \subseteq X$ is a nonempty open set, then a ***section over*** $U$ is a continuous map $s \colon U \to E$ such that $ps = 1_U$; call $s$ a ***global section*** if $U = X$.

Since $p$ is a local homeomorphism and $\Sigma$ is a sheet over $U$, the function $s = (p|\Sigma)^{-1}$ is a section over $U$. Exercise C-4.21 on page 368 shows that $z \colon X \to E$, given by $z(x) = 0_x$ (the identity of the abelian group $E_x$), is a global section; $z$ is called the ***zero section***.

**Notation.** Given a g-sheaf $\mathcal{G} = (E, p, X)$ and a nonempty open set $U$ in $X$, the set of all sections over $U$ is denoted by

$$\Gamma(U, \mathcal{G}) = \{\text{sections } s \colon U \to E\}.$$

We define $\Gamma(\varnothing, \mathcal{G}) = \{0\}$, the trivial group.

Sections $\Gamma(U, \mathcal{G})$ may be viewed as describing local properties of a base space $X$ (they remind us of patches in a manifold), while $\Gamma(X, \mathcal{G})$ describes the corresponding global properties.

If $\mathcal{U}$ and $\mathcal{C}$ are any categories, then a ***presheaf over*** $\mathcal{U}$ ***in*** $\mathcal{C}$ is a contravariant functor $\mathcal{P} \colon \mathcal{U} \to \mathcal{C}$. In particular, if $\mathcal{U}$ is a poset (for example, if $\mathcal{U} = \mathbf{Open}(X)$

is the topology of a space $X$ viewed as a poset under inclusion (Example B-4.1 in Part 1)), then $\mathbf{Open}(X)$ is a (small) category.

**Proposition C-4.21.** *Let $\mathcal{G} = (E, p, X)$ be a g-sheaf of abelian groups.*

(i) *$\Gamma(U, \mathcal{G})$ is an abelian group for each open $U \subseteq X$.*

(ii) *$\Gamma(\quad, \mathcal{G})$ is a presheaf of abelian groups on $X$ (called the **(pre)sheaf of sections** of $\mathcal{G}$); that is, $\Gamma$ is a contravariant functor $\mathbf{Open}(X) \to \mathbf{Ab}$.*

**Proof.**

(i) Let us show that $\Gamma(U, \mathcal{G}) \neq \varnothing$ for every open set $U \subseteq X$. If $U = \varnothing$, then $\Gamma(\varnothing, \mathcal{G})$ is the trivial group $\{0\}$. If $U \neq \varnothing$, take $x \in U$, and choose $e \in E_x$ and a sheet $\Sigma$ containing $e$. Since $p$ is an open map, $p(\Sigma) \cap U$ is an open neighborhood of $x$. Now $(p|\Sigma)^{-1}\colon p(\Sigma) \to \Sigma \subseteq E$ is a section; define $\sigma_\Sigma$ to be its restriction to $p(\Sigma) \cap U$. The family of all such $p(\Sigma) \cap U$ is an open cover of $U$; since the maps $\sigma_\Sigma$ agree on overlaps, Proposition C-4.20(v) shows that they may be glued together to give a section in $\Gamma(U, \mathcal{G})$.

If $s, s' \in \Gamma(U, \mathcal{G})$, then $(s, s')\colon x \mapsto (sx, s'x)$ is a continuous map $U \to E + E$; composing with the continuous map $(sx, s'x) \mapsto sx + s'x$ shows that $s + s'\colon x \mapsto sx + s'x$ is a section over $U$. That $\Gamma(U, \mathcal{G})$ is an abelian group now follows from inversion $E \to E$ being continuous, for $s \in \Gamma(U, \mathcal{G})$ implies $-s \in \Gamma(U, \mathcal{G})$.

(ii) Recall that posets may be viewed as categories (Example B-4.1(viii) in Part 1). We have defined $\Gamma(\quad, \mathcal{G})$ on objects $U$ in $\mathbf{Open}(X)$. To see that it is a presheaf, we define it on morphisms: that is, on the inclusions $\lambda_V^U\colon U \subseteq V$ of open sets. Define $\Gamma(\lambda_V^U)\colon \Gamma(V, \mathcal{G}) \to \Gamma(U, \mathcal{G})$ to be the restriction $\rho_U^V\colon s \mapsto s|U$ (we will usually use this simpler notation):

$$\rho_U^V = \Gamma(\lambda_V^U) \quad \text{and} \quad \rho_U^V(s) = \Gamma(\lambda_V^U)(s) = s|U.$$

It is clear that $\rho_U^V \rho_V^W = \rho_U^W$ when $U \subseteq V \subseteq W$. •

We can recapture stalks from sections, but let us first review direct limits (see Proposition B-7.7 in Part 1). Let $I$ be a poset and let $\{M_i, \varphi_j^i\}$ be a direct system of abelian groups. For each $i \in I$, let $\lambda_i$ be the injection $M_i \to \bigoplus_i M_i$, and define $N$ to be the subgroup of the direct sum generated by all $\lambda_j \varphi_j^i m_i - \lambda_i m_i$, where $m_i \in M_i$ and $i \leq j$. Then

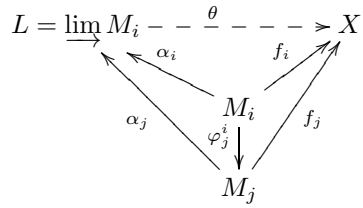
$$L = \Big(\bigoplus_i M_i\Big)/N,$$

together with maps $\alpha_i\colon M_i \to L$ defined by

$$\alpha_i\colon m_i \mapsto \lambda_i m_i + N,$$

is the direct limit: $L \cong \varinjlim M_i$. Thus, every element of $L$ has the form $\sum_i \lambda_i m_i + N$. In the special case when $I$ is directed, then all elements of $L$ have the simpler form $\lambda_i m_i + N$ for some $i$ (Proposition B-7.12 in Part 1). Here is the diagram for the

universal mapping problem:




**Example C-4.22.** Let $\mathcal{G} = (E, p, X)$ be a g-sheaf. For each $x \in X$, consider the family $I = (U)_{U \ni x}$ of all open sets in $X$ containing $x$ (we have written $U \ni x$ instead of $x \in U$). Define an equivalence relation on $\bigcup_{U \ni x} \Gamma(U, \mathcal{G})$ by

$$s \sim s' \text{ if there exists } W \ni x \text{ with } s|W = s'|W.$$

We call the equivalence class of $s$, denoted by

$$[s, x],$$

a ***germ*** of $x$. ◀

**Theorem C-4.23.** *Let $\mathcal{G} = (E, p, X)$ be a* g*-sheaf over a topological space $X$.*

(i) *The family $I = (U)_{U \ni x}$, viewed as a poset under reverse inclusion, is directed.*

(ii) *For each $x \in X$, $\{\Gamma(U, \mathcal{G}), \rho^V_U\}$ is a direct system, where $\rho^V_U \colon \Gamma(V, \mathcal{G}) \to \Gamma(U, \mathcal{G})$ is the restriction $s \mapsto s|U$ when $U \subseteq V$.*

(iii) *Each stalk $E_x$ is a direct limit in* **Ab** :

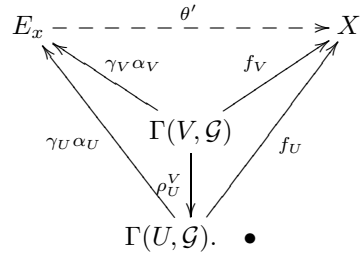
$$E_x = \varinjlim_{U \ni x} \Gamma(U, \mathcal{G}).$$

**Proof.**

(i) If $U$ and $V$ are open sets containing $x$, then $U \cap V$ is also an open set containing $x$.

(ii) This follows from $\rho^V_U \rho^W_V = \rho^W_U$ whenever $U \subseteq V \subseteq W$.

(iii) If $L = \varinjlim_{U \ni x} \Gamma(U, \mathcal{G})$, first define $\alpha_U \colon \Gamma(U, \mathcal{G}) \to L$ by $s_U \mapsto [s_U, x]$, and then define $\gamma \colon L \to E_x$ by $[s_U, x] \mapsto s_U(x)$; $\gamma$ is the desired isomorphism $L \cong E_x$. If $\theta \colon L \to X$ is the map completing the direct limit diagram, then $\theta' = \theta\gamma^{-1} \colon E_x \to X$ is the desired map here,




•

Not only can we construct the stalks of a g-sheaf $\mathcal{G}$ from its presheaf of sections $\Gamma(\ \ ,\mathcal{G})$, we can almost recapture $\mathcal{G}$ itself. We now construct a g-sheaf $\mathcal{P}^g$ from any presheaf $\mathcal{P}$, not necessarily the sheaf of sections of some g-sheaf.

**Proposition C-4.24.** *Every presheaf $\mathcal{P}$ of abelian groups over a space $X$ determines a* g*-sheaf $\mathcal{P}^{\mathrm{g}}$ (called its* ***sheafification****)*

$$\mathcal{P}^{\mathrm{g}} = (E^{\mathrm{g}}, p^{\mathrm{g}}, X)$$

*whose stalks $E^{\mathrm{g}}_x = \varinjlim_{U \ni x} \mathcal{P}(U)$.*

**Proof.** For each $x \in X$, define $E^{\mathrm{g}}_x = \varinjlim_{U\ni x} \mathcal{P}(U)$, define $E^{\mathrm{g}} = \bigcup_{x\in X} E^{\mathrm{g}}_x$, and define a surjection $p^{\mathrm{g}}\colon E^{\mathrm{g}} \to X$ by $e^{\mathrm{g}}_x \mapsto x$ (we denote the elements of $E^{\mathrm{g}}_x$ by $e^{\mathrm{g}}_x$).

If $U \subseteq X$ is a nonempty open set and $s \in \mathcal{P}(U)$, define

$$\langle s, U\rangle = \{[\rho^U_x(s)] : x \in U\}.$$

We claim that $\langle s, U\rangle \cap \langle s', U'\rangle$ either is empty or contains a subset of the same form. If $e \in \langle s, U\rangle \cap \langle s', U'\rangle$, then $e = [\rho^U_x(s)] = [\rho^{U'}_y(s')]$, where $x \in U$, $s \in \mathcal{P}(U)$, and $y \in U'$, $s' \in \mathcal{P}(U')$. But $x = p^{\mathrm{g}}[\rho^U_x(s)] = p^{\mathrm{g}}[\rho^{U'}_y(s')] = y$, so that $x \in U \cap U'$. By Lemma B-7.12 in Part 1, there is an open $W \subseteq U \cap U'$ with $W \ni x$ and $[\rho^U_W \rho^W_x(s)] = [\rho^{U'}_W \rho^W_x(s')]$; call this element $[t]$; note that $\langle t, W\rangle \subseteq \langle s, U\rangle \cap \langle s', U'\rangle$, as desired. Equip $E^{\mathrm{g}}$ with the topology[6] generated by all $\langle s, U\rangle$; it follows that these sets form a base for the topology; that is, every open set is a union of $\langle s, U\rangle$'s.

To see that $(E^{\mathrm{g}}, p^{\mathrm{g}}, X)$ is a g-sheaf, we must show that the surjection $p^{\mathrm{g}}$ is a local homeomorphism. If $e \in E^{\mathrm{g}}$, then $e = [\rho^U_x(s)]$ for some $x \in X$, where $U$ is an open neighborhood of $x$ and $s \in \mathcal{P}(U)$. If $\Sigma = \langle s, U\rangle$, then $\Sigma$ is an open neighborhood of $e$, and it is routine to see that $p^{\mathrm{g}}|\Sigma\colon \Sigma \to U$ is a homeomorphism.

Now each stalk $E^{\mathrm{g}}_x$ is an abelian group. To see that addition is continuous, take $(e, e') \in E^{\mathrm{g}} + E^{\mathrm{g}}$; that is, $e = [\rho^U_x(s)]$ and $e' = [\rho^{U'}_x(s')]$. We may assume the representatives have been chosen so that $s, s' \in \mathcal{P}(U)$ for some $U$, so that $e + e' = [\rho^U_x(s + s')]$. Let $V^{\mathrm{g}} = \langle s + s', V\rangle$ be a basic open neighborhood of $e + e'$. If $\alpha\colon E^{\mathrm{g}} + E^{\mathrm{g}} \to E^{\mathrm{g}}$ is addition, then it is easy to see that if $U^{\mathrm{g}} = [\langle t, W\rangle \times \langle t', W\rangle] \cap (E^{\mathrm{g}} + E^{\mathrm{g}})$, then $\alpha(U^{\mathrm{g}}) \subseteq V^{\mathrm{g}}$. Thus, $\alpha$ is continuous. As inversion $E^{\mathrm{g}} \to E^{\mathrm{g}}$ is also continuous, $\mathcal{P}^{\mathrm{g}} = (E^{\mathrm{g}}, p^{\mathrm{G}}, X)$ is a g-sheaf. •

There are morphisms of g-sheaves.

**Definition.** Let $\mathcal{G} = (E, p, X)$ and $\mathcal{G}' = (E', p', X)$ be g-sheaves over a space $X$. A ***g-map*** $\varphi\colon \mathcal{G} \to \mathcal{G}'$ is a continuous map $\varphi\colon E \to E'$ such that $p'\varphi = p$ (so that $\varphi|E_x\colon E_x \to E'_x$ for all $x \in X$), and each $\varphi|E_x$ is a homomorphism. We write

$$\mathrm{Hom}_{\mathrm{g}}(\mathcal{G}, \mathcal{G}')$$

for the set of all g-maps.

It is easy to check that all g-sheaves of abelian groups over a topological space $X$ and all g-maps form a category, which we denote by

$$\mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab}).$$

[6] This is the coarsest topology on $E$ that makes all sections continuous.

**Proposition C-4.25.** *Let $\mathcal{G} = (E, p, X)$ be a* g*-sheaf, and let $\mathcal{G}^{\mathrm{g}} = (E^{\mathrm{g}}, p^{\mathrm{g}}, X)$ be its sheafification. If $\varphi\colon \mathcal{G} \to \mathcal{G}^{\mathrm{g}}$ is a* g*-map which is an isomorphism on each stalk (i.e., $\varphi_x\colon E_x \to E_x^{\mathrm{g}}$ is an isomorphism for all $x \in X$), then $\varphi\colon \Gamma(\quad, \mathcal{G}) \to \Gamma(\quad, \mathcal{G}^{\mathrm{g}})$ is an isomorphism of presheaves.*

**Proof.** We must show that $\varphi_U\colon \Gamma(U, \mathcal{G}) \to \Gamma(U, \mathcal{G}^{\mathrm{g}})$ is an isomorphism for every open $U$ in $X$. This suffices, for then $(\varphi_U)^{-1}\colon \Gamma(U, \mathcal{G}^{\mathrm{g}}) \to \Gamma(U, \mathcal{G})$ determines its inverse.

We show $\varphi_U$ is injective. If $\varphi_U(s) = 0$ for some $s \in \Gamma(U, \mathcal{G})$, then $\varphi(s)_x = 0$ for each $x \in X$. Since $\varphi_x$ is injective, it follows that $\varphi_x(s) = 0$ for all $x$. As $\varphi_x(s)$ is a direct limit, there is some open neighborhood $V_x$ of $x$ with $V_x \subseteq U$ and the restriction $s|V_x = 0$. But $U$ is covered by all the $V_x$, so that the gluing condition gives $s = 0$. Hence, $\varphi_U$ is injective.

We now show $\varphi_U$ is surjective. Take $t \in \Gamma(U, \mathcal{G}^{\mathrm{g}})$ and, for each $x \in X$, let $t(x)$ lie in the stalk of $x$ in $\mathcal{G}^{\mathrm{g}}$. Since $\varphi_x$ is surjective, there is some $s_x \in \Gamma(U, \mathcal{G})$ with $\varphi_x s_x(x) = t(x)$. Now $s_x(x) = [W_x, x]$ for some open neighborhood $W_x \subseteq U$; as $\varphi$ is continuous, there is some open $V_x \subset U$ with $\varphi^{-1}(V_x) \subseteq W_x$ and $t(x) = [V_x, x]$; note that $U$ is covered by all $W_x$. Suppose that $x, x' \in X$ satisfy $[W_x, x] = [W'_{x'}, x']$ and both are sent by $\varphi$ to $t(x)$. Then both $[W_x \cap W'_{x'}, x] = [W_x \cap W'_{x'}, x']$ are sent by $\varphi$ to $t(x) = [W_x \cap W'_{x'}, x] = [W, x]$. As $\varphi$ is injective on stalks, as we just proved above, $s_x = s_{x'}$ for all $x, x' \in U$. By gluing, there is $s \in \Gamma(U, \mathcal{G})$ with $s|(W_x) = s|(W'_{x'})$ for all $x, x' \in X$. Thus, $\varphi(s) = \varphi_x(s) = t$ for all $x \in U$. Now $\varphi(s)$ and $t$ satisfy $(\varphi s)(x) = [W_x, x] = t(x)$. Gluing all $(\varphi s)_W - t_W$, we have $\varphi s = t$, as desired. •

Exercise C-4.22 on page 368 illustrates a difference between arbitrary presheaves and the special presheaves which arise from sections of a g-sheaf: Proposition C-4.25 says that a presheaf arising from a g-sheaf and its sheafification have the same stalks, while Example C-4.35 on page 372 shows that these two presheaves can be distinct.

The hypotheses of Proposition C-4.25 can be weakened.

**Corollary C-4.26.** *Let $\mathcal{P}$ and $\mathcal{P}'$ be presheaves over a space $X$ satisfying the gluing and uniqueness conditions in* Proposition C-4.20. *If $\varphi\colon \mathcal{P} \to \mathcal{P}'$ is a presheaf map with $\varphi_x$ an isomorphism of stalks for every $x \in X$, then $\varphi$ is an isomorphism of presheaves.*

**Proof.** The proof of Proposition C-4.25 used only the stated properties. •

We are going to prove that $\mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab})$ is an abelian category, and the notion of sub-g-sheaf will be used to describe kernels. Since we are working in a specific category, we do not need the abstract version of subobject (namely, the equivalence class $[i]$ of a monic morphism $i$) required in general additive categories.

**Definition.** A g-sheaf $\mathcal{G}' = (E', p', X)$ is a ***sub-g-sheaf*** of a g-sheaf $\mathcal{G} = (E, p, X)$ if $E'$ is a subspace of $E$, $p' = p|E'$, and the inclusion $\iota\colon \mathcal{G}' \to \mathcal{G}$ is a g-map.

The stalks $E'_x$ of a sub-g-sheaf are subgroups of $E_x$ for all $x \in X$.

**Proposition C-4.27.** $\mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab})$ *is an abelian category.*

**Sketch of Proof.** One first proves that $\mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab})$ is an additive category. If $\mathcal{G} = (E, p, X)$ and $\mathcal{G}' = (E', p', X)$ are g-sheaves and $\varphi, \psi\colon E \to E'$ are g-maps, then $\mathrm{Hom}_{\mathrm{g}}(\mathcal{G}, \mathcal{G}')$ is an additive abelian group if we define $\varphi + \psi\colon E \to E'$ by $\varphi + \psi\colon e \mapsto \varphi(e) + \psi(e)$. Verification that the distributive laws hold and that the zero g-sheaf $0 = \{0\}^{\mathrm{g}}$ (see Example C-4.19(i)) is a g-sheaf are routine. Given g-sheaves $\mathcal{G} = (E, p, X)$ and $\mathcal{G}' = (E', p', X)$, define

$$\mathcal{G} \oplus \mathcal{G}' = (E + E', p^+, X),$$

where $E + E'$ is the subspace of $E \times E'$:

$$E + E' = \bigcup_{x \in X} (E_x \times E'_x),$$

and $p^+\colon E + E' \to X$ is given by $p^+\colon (e_x, e'_x) \mapsto x$. It is straightforward to prove that $\mathcal{G} \oplus \mathcal{G}'$ is a g-sheaf that is both a product and coproduct of $\mathcal{G}$ and $\mathcal{G}'$.

It remains to display kernels and cokernels of g-maps $\varphi\colon \mathcal{G}' \to \mathcal{G}$ and to prove the remaining axioms in the definition of abelian category. Since $\varphi\colon E' \to E$ is a g-map, its restriction $\varphi|E'_x\colon E'_x \to E_x$ is a homomorphism. Define $\varphi_x = \varphi|E'_x$, and let $e'_x \in \ker \varphi_x \subseteq E'_x$. For every open $U_x$ in $X$ containing $x$, let $\Sigma'_x$ be the sheet over $U_x$ containing $e'_x$, and define $K' = \bigcup_{x \in X} \Sigma'_x \subseteq E'$. Finally, define $\mathcal{K}' = (K', q', X)$, where $q' = p'|K'$. Then $\mathcal{K}' = (K', q', X)$ is a g-sheaf and the inclusion $\iota\colon K' \to E'$ is a g-map $\mathcal{K}' \to \mathcal{G}'$.

Similarly, if $\varphi\colon \mathcal{G}' \to \mathcal{G}$, define $\mathcal{Q} = (E^*, p^*, X)$, where the stalk $E^*_x = E_x / \operatorname{im} \varphi_x$ for all $x \in X$ and $p^*\colon E^* \to X$ maps each coset $E^*_x = E_x / \operatorname{im} \varphi_x \mapsto x$; define $\operatorname{coker}(\varphi) = \mathcal{Q}$. Again, the straightforward details are left to the reader. •

A sequence of g-sheaves is exact if and only if the sequence of stalks is exact.

**Corollary C-4.28.** *A sequence of* g*-sheaves over a space* $X$

$$0 \to \mathcal{G}' \xrightarrow{\varphi} \mathcal{G} \xrightarrow{\psi} \mathcal{G}'' \to 0$$

*is exact if and only if the sequence of stalks*

$$0 \to E'_x \xrightarrow{\varphi|E'_x} E_x \xrightarrow{\psi|E_x} E''_x \to 0$$

*is exact in* **Ab** *for every* $x \in X$.

**Proof.** Exercise C-4.22 below says that $\operatorname{im} \varphi$ is the sub-g-sheaf of $\mathcal{G}$ whose stalks are $E_x = \operatorname{im} \varphi|E'_x$ for all $x \in X$. The result follows from $\mathbf{Sh}_{\mathrm{g}}(X)$ being an abelian category. •

Exactness of presheaves over a space $X$ can be contrasted with exactness of g-sheaves. Corollary C-4.14 says that a sequence of presheaves over $X$

$$\mathcal{P}' \xrightarrow{\sigma} \mathcal{P} \xrightarrow{\tau} \mathcal{P}''$$

is exact if and only if

$$\mathcal{P}'(U) \xrightarrow{\sigma|U} \mathcal{P}(U) \xrightarrow{\tau|U} \mathcal{P}''(U)$$

is exact in **Ab** for every open set $U$ in $X$.

## Exercises

∗ **C-4.20.** (i) Prove that every local homeomorphism $p\colon E \to X$ is an open map; that is, if $V$ is an open set in $E$, then $p(V)$ is an open set in $X$.

(ii) Prove that if $\mathcal{G} = (E, p, X)$ and $\mathcal{G}' = (E', p', X)$ are g-sheaves over a topological space $X$, then every g-map $\varphi\colon \mathcal{G} \to \mathcal{G}'$ is an open map $E \to E'$.
**Hint.** This follows from Proposition C-4.20(i), which says that the sheets form a base of open sets for $E$.

∗ **C-4.21.** Let $\mathcal{G} = (E, p, X)$ be a g-sheaf. Prove that $z\colon X \to E$, defined by $z\colon x \mapsto 0_x$, the identity element of the abelian group $E_x$, is continuous. Conclude that the ***zero section*** $z$ is a global section.

∗ **C-4.22.** (i) Prove that two g-sheaves over a space $X$, say, $\mathcal{G} = (E, p, X)$ and $\mathcal{G}' = (E', p', X)$, are equal if and only if the stalks $E_x = E'_x$ for all $x \in X$.

(ii) Prove that a g-sheaf is the zero sheaf $0 = \{0\}^{\mathrm{g}}$ if and only if all its stalks are $\{0\}$.

∗ **C-4.23.** Let $\varphi\colon \mathcal{G}' \to \mathcal{G}$ be a g-map, where $\mathcal{G} = (E, p, X)$ and $\mathcal{G}' = (E', p', X)$ are g-sheaves over a space $X$. Prove that $\operatorname{im}\varphi$ is the sub-g-sheaf of $\mathcal{G}$ having stalks $\operatorname{im}\varphi_x$.

**C-4.24.** Prove that there are equivalences of categories $\mathbf{pSh}(X, \mathbf{Ab}) \cong \mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab}) \cong \mathbf{Ab}$ if $X$ is a one-point space.

**C-4.25.** Prove that sub-g-sheaves $(E, p, X)$ and $(E', p', X)$ of a g-sheaf are equal if and only if they have the same stalks; that is, $E_x = E'_x$ for all $x \in X$.

∗ **C-4.26.** (i) Prove that $\mathcal{P} \mapsto \mathcal{P}^{\mathrm{g}}$ is an additive functor $\Sigma\colon \mathbf{pSh}(X, \mathbf{Ab}) \to \mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab})$.

(ii) Use Proposition C-4.25 to prove that if $\mathcal{P} = \Gamma(\ \ , \mathcal{G})$, where $\mathcal{G}$ is a g-sheaf, then $\mathcal{P} \cong \Sigma(\mathcal{P}) = \mathcal{P}^{\mathrm{g}}$.

## C-4.4. Sheaves

There are two equivalent versions of *sheaf*: the first, more visual, is a g-sheaf; the second, more algebraic, arises from a special kind of presheaf, and it is the version serious users accept. We now begin discussing this second viewpoint.

Recall that if $X$ is a topological space, then $\mathbf{Open}(X)$ is the topology of $X$ viewed as a poset under inclusion (for later use, we note that $\mathbf{Open}(X)$ is a directed set, for the union of open sets is open).

In particular, if $\mathcal{G}$ is a g-sheaf over $X$, then its presheaf of sections $\Gamma(\ \ , \mathcal{G})$ (see Proposition C-4.21) satisfies a special property not shared by arbitrary presheaves.

**Proposition C-4.29.** *Let $\mathcal{G} = (E, p, X)$ be a* g*-sheaf with presheaf of sections $\Gamma(\ \ , \mathcal{G})$, let $U$ be an open set in $X$, and let $(U_i)_{i\in I}$ be an open cover of $U$: $U = \bigcup_{i\in I} U_i$.*

(i) *If a family* $(s_i \in \Gamma(U_i, \mathcal{G}))_{i\in I}$ *satisfies* $s_i|(U_i \cap U_j) = s_j|(U_i \cap U_j)$ *for all* $(i,j) \in I \times I$, *then there exists a unique* $s \in \Gamma(U, \mathcal{G})$ *with* $s|U_i = s_i$ *for all* $i \in I$.

(ii) *If* $s, s' \in \Gamma(U, \mathcal{G})$ *and* $s|U_i = s'|U_i$ *for all* $i \in I$, *then* $s = s'$.

**Proof.** Proposition C-4.20(iv) and (v). •

**Definition.** A presheaf $\mathcal{P}$ of abelian groups on a space $X$ satisfies the ***equalizer condition*** if the following two conditions hold for every open cover $(U_i)_{i\in I}$ of open sets $U \subseteq X$.

(i) (**Gluing**) If $s_i|(U_i \cap U_j) = s_j|(U_i \cap U_j)$ for all $s_i \in \mathcal{P}(U_i)$ and $s_j \in \mathcal{P}(U_j)$, then there exists $s \in \mathcal{P}(U)$ with $s|U_i = s_i$ for all $i \in I$.

(ii) (**Uniqueness**) If $s, s' \in \mathcal{P}(U)$ satisfy $s|U_i = s'|U_i$ for all $i \in I$, then $s = s'$.

The equalizer condition can be restated in a more categorical way.

**Corollary C-4.30.** *Let* $\mathcal{P}$ *be a presheaf of abelian groups over a topological space* $X$.

(i) $\mathcal{P}$ *satisfies the equalizer condition if and only if, for every open cover* $(U_i)_{i\in I}$ *of an open set* $U$ *in* $X$, *there is an exact sequence of abelian groups*

$$0 \to \mathcal{P}(U) \xrightarrow{\alpha} \prod_{i\in I} \mathcal{P}(U_i) \xrightarrow{\beta} \prod_{(i,j)\in I\times I} \mathcal{P}(U_{(i,j)}),$$

*where* $U_{(i,j)} = U_i \cap U_j$ *for* $i, j \in I$ *and* $\alpha$ *and* $\beta$ *are defined as follows. If* $s \in \mathcal{P}(U)$, *then the* $i$*th coordinate of* $\alpha(s)$ *is* $s|U_i$; *if* $(s_i) \in \prod_{i\in I} \mathcal{P}(U_i)$, *then the* $(i,j)$*th coordinate of* $\beta((s_i))$ *is* $s_i|U_{(i,j)} - s_j|U_{(i,j)}$.

(ii) *If* $\mathcal{P}$ *satisfies the equalizer condition and* $s = (s_i) \in \prod_{i\in I} \mathcal{P}(U_i)$ *satisfies* $s_i|(U_i \cap U_j) = s_j|(U_i \cap U_j)$, *then* $s$ *corresponds to a unique global section in* $\mathcal{P}(X)$.

**Proof.**

(i) Proposition C-4.29(i) shows that $\alpha$ is an injection. Now $\operatorname{im}\alpha \subseteq \ker\beta$, for $\beta\alpha(s)$ has $(i,j)$ coordinate $s|U_{(i,j)} - s|(i,j) = 0$. The reverse inclusion follows from Proposition C-4.29(ii).

(ii) Such an element $s$ lies in $\ker\beta$, where $\beta = \beta' - \beta''$, where $\beta, \beta'$ are the maps in the equalizer diagram above. •

Here is an example of a presheaf which does not satisfy the equalizer condition.

**Example C-4.31.** Let $G$ be an abelian group and let $X$ be a topological space. The ***constant presheaf at*** $G$ over $X$ is defined on a nonempty open set $U \subseteq X$ by

$$\mathcal{G}(U) = \{f : U \to G : f \text{ is constant}\};$$

define $\mathcal{G}(\varnothing) = \{0\}$ and, if $U \subseteq V$, define $\rho_U^V : \mathcal{G}(V) \to \mathcal{G}(U)$ by $f \mapsto f|U$. If $U = U_1 \cup U_2$, where $U_1, U_2$ are disjoint nonempty open sets, define $s_1 \in \mathcal{P}(U_1)$ by $s_1(u_1) = 0$ for all $u_1 \in U_1$, and define $s_2 \in \mathcal{P}(U_2)$ by $s_2(u_2) = 5$ for all $u_2 \in U_2$. The overlap condition here is vacuous, because $U_1 \cap U_2 = \varnothing$, but there is no *constant*

function $s \in \mathcal{P}(U)$ with $s|U_i = s_i$ for $i = 1, 2$. Hence, $\mathcal{G}$ is not a sheaf, for it does not satisfy the equalizer condition. ◀

The following definition is the one preferred by every serious user of sheaves.

**Definition.** A ***sheaf of abelian groups*** over a space $X$ is a presheaf[7]

$$\mathcal{F}\colon \mathbf{Open}(X)^{\mathrm{op}} \to \mathbf{Ab}$$

that satisfies the equalizer condition. We shall always assume that $\mathcal{F}(\varnothing) = \{0\}$.

Note that sheaves, defined as contravariant functors, are much simpler than g-sheaves, for they avoid fussy point-set topology. As with g-sheaves, sheaves can be defined with values in categories other than **Ab**. A function $f\colon X \to Y$ of topological spaces is ***locally constant*** if each $x \in X$ has an open neighorhood $U$ with $f|U$ constant.

**Example C-4.32.**

(i) Let $G$ be a (discrete) abelian group. Define the ***constant sheaf at*** $G$ over $X$ to be the sheaf $\Gamma_G$ where, for all $U \in \mathbf{Open}(X)$,

$$\Gamma_G(U) = \{\text{all locally constant } U \to G\}.$$

Compare this definition with that of the constant g-sheaf $G^{\mathrm{g}}$ in Example C-4.19(i); also see Example C-4.31.

(ii) Let $G$ be an abelian group, $X$ a topological space, and $x \in X$. Define a presheaf by

$$x_*G(U) = \begin{cases} G & \text{if } x \in U, \\ \{0\} & \text{otherwise.} \end{cases}$$

If $U \subseteq V$, then the restriction map $\rho_U^V$ is either $1_G$ or 0. It is easy to check that $x_*G$ is a sheaf, called a ***skyscraper sheaf***; it is so called because if $x$ is a closed point,[8] then all the stalks of $x_*G$ are $\{0\}$ except $(x_*G)_x$, which is $G$.

(iii) Let $X$ be the unit circle, which we view as $\{z \in \mathbb{C} : |z| = 1\}$, and let $p\colon X \to X$ be defined by $p\colon z \mapsto z^2$. If we set $E = X$, then we have defined a g-sheaf $\mathcal{S} = (E, p, X)$, which we call the ***double cover***. See Figure C-4.5.

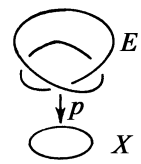



**Figure C-4.5.** Double cover.

[7] We denote a sheaf by $\mathcal{F}$ because F is the initial letter of the French term *faisceau*.

[8] If $X = \mathrm{Spec}(R)$ for some commutative ring $R$, then the closure of a point $\mathfrak{p} \in X$ consists of all the prime ideals in $R$ that contain $\mathfrak{p}$. Thus, $\mathfrak{p}$ is a closed point in $X$ if and only if it is a maximal ideal in $R$.

All the stalks of $\mathcal{S}$ are isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$. An interesting feature of the sheaf of sections $\Gamma(\quad, \mathcal{S})$ is that it has the same stalks as the constant sheaf at $\mathbb{Z} \oplus \mathbb{Z}$, yet the two sheaves are not isomorphic. This merely reflects the obvious fact that different spaces can be the same locally.

(iv)

**Definition.** Let $\mathcal{G}'$ and $\mathcal{G}$ be presheaves of abelian groups on a topological space $X$ such that $\mathcal{G}'(U) \subseteq \mathcal{G}(U)$ for every open set $U$ in $X$; that is, there are inclusions $\iota_U \colon \mathcal{G}'(U) \to \mathcal{G}(U)$. Then $\mathcal{G}'$ is a ***subpresheaf*** of $\mathcal{G}$ if the ***inclusion*** $\iota \colon \mathcal{G}' \to \mathcal{G}$ is a presheaf map.

If $\mathcal{F}$ is a sheaf, then $\mathcal{G}'$ is a ***subsheaf*** of $\mathcal{F}$ if $\mathcal{G}'$ is a subpresheaf that is also a sheaf.

(v) The zero sheaf (see Example C-4.32) is a subsheaf of every sheaf.

(vi) Let $\mathcal{F}$ be the sheaf of germs of continuous functions on a space $X$. Define $\mathcal{G}$ by setting $\mathcal{G}(U) = \mathcal{F}(U)$ for all open sets $U$ and by setting restrictions $\psi_U^V$ to be identically 0. Then $\mathcal{G}$ is a presheaf, but $\mathcal{G}$ is not a subpresheaf of $\mathcal{F}$ (for the inclusion is not a presheaf map).

It is clear that subpresheaves $\mathcal{F}$ and $\mathcal{F}'$ of a presheaf $\mathcal{G}$ are equal if and only if $\mathcal{F}(U) = \mathcal{F}'(U)$ for all open $U$. This simplifies for sheaves. ◀

**Proposition C-4.33.** *If $\mathcal{G} = (E, p, X)$ is a* g*-sheaf, then its presheaf of sections $\Gamma(\quad, \mathcal{G})$ is a sheaf of abelian groups over $X$.*

**Proof.** Proposition C-4.29 says that $\Gamma(\quad, \mathcal{G})$ satisfies the equalizer condition. •

From now on, we shall call $\Gamma(\quad, \mathcal{P}^{\mathrm{g}})$ the ***sheaf of sections*** instead of the *presheaf of sections*

**Example C-4.34.** For each open set $U$ of a topological space $X$, define

$$\mathcal{P}(U) = \{\text{continuous } f \colon U \to \mathbb{R}\};$$

$\mathcal{P}(U)$ is an abelian group under pointwise addition: $f + g \colon x \mapsto f(x) + g(x)$ and $\mathcal{P}$ is a presheaf over $X$. For each $x \in X$, define an equivalence relation on $\bigcup_{U \ni x} \mathcal{F}(U)$ by $f \sim g$ if there is some open set $W$ containing $x$ with $f|W = g|W$. The equivalence class of $f$, denoted by $[x, f]$, is called a ***germ*** at $x$ (see Example C-4.22). Define $E_x$ to be the family of all germs at $x$, define $E = \bigcup_{x \in X} E_x$, and define $p \colon E \to X$ by $p \colon [x, f] \mapsto x$. Proposition C-4.24 shows that the sheafification $\mathcal{P}^{\mathrm{g}} = (E^{\mathrm{g}}, p, X)$ is a g-sheaf, and so Proposition C-4.33 says that $\Gamma(\quad, \mathcal{P}^{\mathrm{g}})$ is a sheaf (it is called the ***sheaf of germs of continuous functions over*** $X$). ◀

**Definition.** Let $\mathcal{F}$ and $\mathcal{G}$ be sheaves (of abelian groups) over a space $X$, with restriction maps $\rho_U^V \colon \mathcal{F}(V) \to \mathcal{F}(U)$ and $\tau_U^V \colon \mathcal{G}(V) \to \mathcal{G}(U)$ whenever $U \subseteq V$ are open. A ***sheaf map*** $\varphi \colon \mathcal{F} \to \mathcal{G}$ is a natural transformation; that is, $\varphi$ is a one-parameter family of morphisms $\varphi_U \colon \mathcal{F}(U) \to \mathcal{G}(U)$, indexed by $U \in \mathbf{Open}(X)$,

such that there is a commutative diagram whenever $U \subseteq V$:

$$\begin{array}{ccc} \mathcal{F}(V) & \xrightarrow{\varphi_V} & \mathcal{G}(V) \\ {\scriptstyle \rho_U^V}\downarrow & & \downarrow{\scriptstyle \tau_U^V} \\ \mathcal{F}(U) & \xrightarrow[\varphi_U]{} & \mathcal{G}(U). \end{array}$$

In the previous section, we defined the sheafification of a presheaf $\mathcal{P}$ to be the g-sheaf $\mathcal{P}^{\mathrm{g}}$; we now modify the definition so that the sheafification of $\mathcal{P}$ is a sheaf.

**Definition.** If $\mathcal{P}$ is a presheaf of abelian groups, then its ***sheafification*** is the sheaf $\Gamma(\quad, \mathcal{P}^{\mathrm{g}})$, where $\Gamma$ is the sheaf of sections of $\mathcal{P}^{\mathrm{g}}$ (see Proposition C-4.24 and Corollary C-4.37). Henceforth, we shall also denote the sheafification of $\mathcal{P}$ by $\mathcal{P}^{\mathrm{g}}$.

**Remark.** There is a construction of the sheafification $\mathcal{P}^{\mathrm{g}}$ of a presheaf $\mathcal{P}$ that does not use g-sheaves. We quote from Hartshorne [**95**], p. 64.

> For any open set $U$, let $\mathcal{P}^{\mathrm{g}}(U)$ be the set of functions $s$ from $U$ to the union $\bigcup_{x \in U} \mathcal{P}_x$ of the stalks of $\mathcal{P}$ over points of $U$, such that
> (i) for each $x \in U$, $s(x) \in \mathcal{P}_x$ and
> (ii) for each $x \in U$, there is a neighborhood $V$ of $x$, contained in $U$, and an element $t \in \mathcal{P}(V)$, such that for all $y \in V$, the germ $t_y$ of $t$ at $y$ is equal to $s(y)$. ◀

Recall Corollary C-4.13: all presheaves over a topological space $X$ form an abelian category $\mathbf{pSh}(X, \mathbf{Ab})$, whose morphisms, *presheaf maps*, are natural transformations: $\mathrm{Hom}_{\mathrm{g}}(\mathcal{P}, \mathcal{P}') = \mathbf{Nat}(\mathcal{P}, \mathcal{P}')$. It follows that if $\mathcal{F}$ and $\mathcal{G}$ are sheaves, then every presheaf map $\mathcal{F} \to \mathcal{G}$ is a sheaf map.

**Notation.** Define $\mathbf{Sh}(X, \mathbf{Ab})$ to be the full subcategory of $\mathbf{pSh}(X, \mathbf{Ab})$ generated by all sheaves over a space $X$.

It is easy to see that $\mathbf{Sh}(X, \mathbf{Ab})$ is an additive category.

In Example C-4.31, we saw that there exist presheaves that do not arise from g-sheaves, which explains why we invoked Proposition C-4.24 in Example C-4.34. But if we focus on sheaves, there is a more natural example: the cokernel of a sheaf map, though always a presheaf, may not be a sheaf.

**Example C-4.35.** There exists an exact sequence of presheaves,

$$0 \to \mathcal{F}' \xrightarrow{\varphi} \mathcal{F} \xrightarrow{\psi} \mathcal{G} \to 0,$$

where both $\mathcal{F}'$ and $\mathcal{F}$ are sheaves but $\mathcal{G} = \operatorname{coker} \varphi$ is not a sheaf.

Let $\mathcal{F} = \mathcal{O}$ be the sheaf of germs of complex holomorphic functions on the punctured plane $X = \mathbb{C} - \{0\}$; thus,

$$\mathcal{O}(U) = \{\text{holomorphic } f \colon U \to \mathbb{C}\}$$

is an additive abelian group for all open $U$. Let $\mathcal{G} = \mathcal{O}^{\times}$ be the sheaf on $X$ defined by $\mathcal{O}^{\times}(U) = \{\text{holomorphic } f \colon U \to \mathbb{C}^{\times}\}$; that is, $f(z) \neq 0$ for all $z \in U$.

If $\varphi\colon \mathcal{O} \to \mathcal{O}^\times$ is the sheaf map defined by $\varphi_U\colon f \mapsto e^{2\pi i f}$, then $\ker\varphi \cong \mathbb{Z}$, the constant sheaf at $\mathbb{Z}$ on $X$, and, for each $U$, there is an exact sequence of presheaves

$$0 \to \mathbb{Z} \to \mathcal{O} \xrightarrow{\varphi} \mathcal{O}^\times \to 0.$$

We claim that $\operatorname{im}\varphi$ is not a sheaf. Let $(U_i)_{i\in I}$ be an open cover of $X$ by disks. Define $f_i \in \mathcal{O}^\times(U_i)$ by $f_i(z) = z$ for all $z \in U_i$. Of course, this family agrees on overlaps, and the unique global section they determine is $f = 1_X$. Now each $f_i \in \varphi(U_i)$, for there is a logarithm $\ell_i(z)$ defined on $U_i$ with $e^{\ell_i(z)} = z$ (because the disk $U_i$ is simply connected). However, it is well known from a complex analysis course that logarithm cannot be defined as a holomorphic function on all of $X$, and so $1_X$ is not a global section of $\operatorname{im}\varphi$. Therefore, $\operatorname{im}\varphi$ is not a sheaf. ◀

Sheaves arise naturally when encoding local information, and we have just seen an example for which local data cannot be globalized. This phenomenon will be "measured" by sheaf cohomology.

The next result shows that there is no essential difference between sheaves and g-sheaves.

**Theorem C-4.36.** *Let $X$ be a topological space.*

(i) *There is an equivalence of categories*:

$$\Phi\colon \mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab}) \to \mathbf{Sh}(X, \mathbf{Ab}).$$

(ii) $\mathbf{Sh}(X, \mathbf{Ab})$ *is an abelian category.*

(iii) *A sequence of sheaves*

$$0 \to \mathcal{F}' \xrightarrow{\tau} \mathcal{F} \xrightarrow{\sigma} \mathcal{F}'' \to 0$$

*is exact if and only if the sequence of stalks is exact in* $\mathbf{Ab}$:

$$0 \to \mathcal{F}'_x \xrightarrow{\tau_U} \mathcal{F}_x \xrightarrow{\sigma_U} \mathcal{F}''_x \to 0.$$

**Sketch of Proof.**

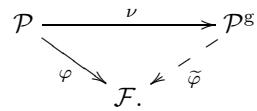
(i) Define $\Phi$ on objects by $\Phi\mathcal{G} = \Gamma(\quad, \mathcal{G})$. If $\mathcal{G} = (E, p, X)$ and $\mathcal{G}' = (E', p', X)$, then a g-map $\mathcal{G} \to \mathcal{G}'$ is a continuous map $\tau\colon E \to E'$ such that $p'\tau = p$. Hence, if $s \in \Gamma(U, \mathcal{G})$, then $\tau s \in \Gamma(U, \mathcal{G}')$, and we define $\Phi\tau$ to be the natural transformation $(\Phi\tau)_U\colon \Gamma(U, \mathcal{G}) \to \Gamma(U, \mathcal{G}')$ sending $s \mapsto \tau s$. It is routine to check that $\Phi$ is an additive functor.

By Exercise C-4.26 on page 368, there is a functor $\Sigma\colon \mathbf{pSh}(X, \mathbf{Ab}) \to \mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab})$; namely, $\mathcal{P} \mapsto \mathcal{P}^{\mathrm{g}}$, in the reverse direction. If $\Sigma'$ is the restriction of $\Sigma$ to $\mathbf{Sh}(X, \mathbf{Ab})$, then the reader may use Exercise C-4.26(ii) (essentially Proposition C-4.25, which says that a presheaf $\mathcal{G}$ and its sheafification $\mathcal{G}^*$ have the same stalks) to show that both composites $\Phi\Sigma'$ and $\Sigma'\Phi$ are naturally equivalent to identity functors.

(ii) As $\mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab})$ is abelian, Exercise C-4.19 on page 359 says that $\mathbf{Sh}(X, \mathbf{Ab})$ is abelian.

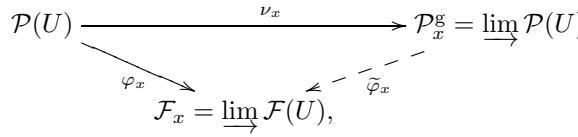
(iii) Corollary C-4.26. •

**Corollary C-4.37.** *Let $\mathcal{P}$ be a presheaf of abelian groups over a space $X$, let $\mathcal{P}^{\mathrm{g}} = (E^{\mathrm{g}}, p^{\mathrm{g}}, X)$ be its associated* g*-sheaf, and let $\mathcal{P}^{\mathrm{g}} = \Gamma(\ \ , \mathcal{P}^{\mathrm{g}})$ be the sheaf of sections of $\mathcal{P}^{\mathrm{g}}$. There exists a presheaf map $\nu\colon \mathcal{P} \to \mathcal{P}^{\mathrm{g}}$ that solves the following universal mapping problem:*




*For every presheaf map $\varphi\colon \mathcal{P} \to \mathcal{F}$, where $\mathcal{F}$ is a sheaf over $X$, there exists a unique sheaf map $\widetilde{\varphi}\colon \mathcal{P}^{\mathrm{g}} \to \mathcal{F}$ with $\widetilde{\varphi}\nu = \varphi$.*

**Proof.** The functor $\Phi\colon \mathbf{Sh}_{\mathrm{g}}(X, \mathbf{Ab}) \to \mathbf{Sh}(X, \mathbf{Ab})$ in Theorem C-4.36 gives a sheaf map $\Phi(\varphi)\colon \Gamma(\ \ , \mathcal{P}) \to \Gamma(\ \ , \mathcal{F})$ making the diagram commute. By Proposition C-4.25, it suffices to see that $\widetilde{\varphi}_x = \psi_x$ for all $x \in X$. But $\mathcal{P}^{\mathrm{g}}_x = \varinjlim_{U \ni x} \mathcal{P}(U)$,




and the universal property of direct limit gives a unique map making the diagram commute. •

Both g-sheaf and presheaf views of a sheaf are useful. Here are two important constructions, called ***change of base***, arising from a continuous map $f\colon X \to Y$. Proofs in the rest of this section will be less detailed, for they tend to be lengthy. Of course, you know that a long proof need not be difficult (writing out all the details of a diagram chase is tedious). The reader may either look at the references cited in the proofs or regard such proofs as exercises.

**Definition.** Given a continuous $f\colon X \to Y$ and a presheaf $\mathcal{F}$ over $X$, define the ***direct image*** $f_*\mathcal{F}$ to be the presheaf over $Y$ with

$$f_*\mathcal{F}(V) = \mathcal{F}(f^{-1}V)$$

for every open $V \subseteq Y$.

Note that $f^{-1}V$ is open in $X$ because $f$ is continuous. It is easy to see that if $\mathcal{F}$ is a sheaf, then $f_*\mathcal{F}$ is a sheaf as well. (We note that there is a map of stalks $\mathcal{F}_x \to (f_*\mathcal{F})_{f(x)}$, but it may not be an isomorphism because the direct system indexed by $\{f^{-1}V \ni f(x)\}$ may not be cofinal in the poset $\{V \ni f(x)\}$.)

Even more is true: $f_*\colon \mathbf{pSh}(Y; \mathbf{Ab}) \to \mathbf{pSh}(X, \mathbf{Ab})$ is an additive functor whose restriction to the subcategory of sheaves

$$f_*\colon \mathbf{Sh}(X, \mathbf{Ab}) \to \mathbf{Sh}(Y, \mathbf{Ab})$$

is also an additive functor (some authors denote $f_*$ by $f_!$).

As an example, let $i\colon X \to Y$ be the inclusion of a *closed* subspace, and let $\mathcal{F}$ be a constant sheaf at $G$ over $X$. Then the stalks of $i_*\mathcal{F}$ are $G$, while the stalks over $y \notin X$ are $\{0\}$(see Tennison [**219**], p. 54).

**Definition.** Let $\mathcal{F}$ be a sheaf over a space $Y$. If $i\colon X \to Y$ is the inclusion of a subspace, then the ***restriction*** $\mathcal{F}|X$ is $i_*\mathcal{F}$. The sheaf $\mathcal{F}$ is called an ***extension*** of $\mathcal{F}|X$.

The restriction $\mathcal{F}|X$ is a sheaf over $X$.

**Definition.** Let $X$ be a subspace of a space $Y$, and let $\mathcal{F}'$ be a sheaf over $X$. We say that a sheaf $\mathcal{F}$ over $Y$ is the ***extension of $\mathcal{F}$ by zero*** if $\mathcal{F}|X = \mathcal{F}'$ and the stalk $\mathcal{F}_y = \{0\}$ for all $y \in Y$ with $y \notin X$.

**Proposition C-4.38.** *If $X$ is a locally closed subspace*[9] *of a space $Y$, then every sheaf over $X$ has an extension by zero over $Y$.*

**Proof.** Tennison [**219**], pp. 63–64. •

There is a second change of base construction, called *inverse image*, that constructs a sheaf over $X$ from a sheaf over $Y$. It is simplest to define inverse image in terms of g-sheaves. Afterward, we will give the sheaf construction.

**Definition.** Let $\mathcal{G}$ be a presheaf over $Y$, and let $\mathcal{G}^{\mathrm{g}} = (E, p, Y)$ be its sheafification. Construct the pullback (in **Top**)

$$\begin{array}{ccc} E' & \dashrightarrow & E \\ {\scriptstyle p'}\downarrow & & \downarrow{\scriptstyle p} \\ X & \xrightarrow[f]{} & Y. \end{array}$$

Then $\mathcal{G}^{\mathrm{g}} = (E', p', X)$ is a g-sheaf over $X$. Then $f^*\mathcal{G} = \Gamma(\ \ , \mathcal{G}^{\mathrm{g}})$ is called the ***inverse image*** of $\mathcal{G}$.

In more detail, the pullback $E' = \{(e, x) \in E \times X : p(e) = f(x)\}$; its topology is that of a subspace of $E \times X$. It is easy to see that if $\mathcal{G}$ is a sheaf over $Y$, then $f^*\mathcal{G}$ is a sheaf over $X$, and its stalks are

$$(f^*\mathcal{G})_x = \mathcal{G}_{f(x)}$$

(see Tennison [**219**], p. 58).

**Example C-4.39.**

(i) If $f\colon U \to X$ is the inclusion of an open subset, then $f^*\mathcal{F}$ is the restriction sheaf $\mathcal{F}|U$.

(ii) If $\mathcal{G}$ is a presheaf over a space $X$, then $(1_X)^*\mathcal{G}$ is the sheafification of $\mathcal{G}$ (see the remark on p. 61 of [**219**]).

(iii) Let $G$ be an abelian group and, if $x \in X$, let $i\colon \{x\} \to X$ be the inclusion. If $\mathcal{G}(x)$ is the sheaf over $\{x\}$ with stalk $G$ (this is just a fancy way of viewing $G$), then $i_*\mathcal{G} = x_*G$, the skyscraper sheaf (Example C-4.32). Moreover, if $\mathcal{F}$ is a sheaf over $X$, then $i^*\mathcal{F} = \mathcal{G}(x)$. ◄

[9] A subspace $X$ of a space $Y$ is ***locally closed*** if there are an open $U$ and a closed $V$ in $Y$ such that $X = U \cap V$. Of course, every closed set is locally closed, as is every open set.

**Remark.** Here is a second construction of an inverse image sheaf. Let $f\colon X \to Y$ be continuous, and let $\mathcal{F}$ be a sheaf over $Y$. To define a presheaf $\mathcal{F}^+$ over $X$, we must define $\mathcal{F}^+(U)$ for every open $U$ in $X$. Since $f(U)$ need not be open in $Y$, we define

$$\mathcal{F}^+(U) = \varinjlim_{W \supseteq f(U)} \mathcal{F}(W),$$

where $W$ is open in $Y$. Then $f^*\mathcal{F}$ is equal to the sheafification of $\mathcal{F}^+$. Note that the remark on page 372 constructs the sheafification without g-sheaves. ◀

Inverse image gives an additive functor $f^*\colon \mathbf{pSh}(Y, \mathbf{Ab}) \to \mathbf{pSh}(X, \mathbf{Ab})$, whose restriction gives an additive functor

$$f^*\colon \mathbf{Sh}(Y, \mathbf{Ab}) \to \mathbf{Sh}(X, \mathbf{Ab})$$

(some authors denote $f^*$ by $f^!$).

**Theorem C-4.40.** *If $f\colon X \to Y$ is continuous, then $(f^*, f_*)$ is an adjoint pair of functors between $\mathbf{Sh}(X, \mathbf{Ab})$ and $\mathbf{Sh}(Y, \mathbf{Ab})$.*

**Proof.** For details, see Tennison [**219**], pp. 57–61. •

## Exercises

∗ **C-4.27.**

(i) Prove that the zero sheaf is a zero object in $\mathbf{Sh}(X, \mathbf{Ab})$ and in $\mathbf{pSh}(X, \mathbf{Ab})$.

(ii) Prove that $\mathrm{Hom}(\mathcal{P}, \mathcal{P}')$ is an additive abelian group when $\mathcal{P}, \mathcal{P}'$ are presheaves or when $\mathcal{P}, \mathcal{P}'$ are sheaves.

(iii) The distributive laws hold: given presheaf maps

$$\mathcal{X} \xrightarrow{\alpha} \mathcal{P} \underset{\psi}{\overset{\varphi}{\rightrightarrows}} \mathcal{Q} \xrightarrow{\beta} \mathcal{Y},$$

where $\mathcal{X}$ and $\mathcal{Y}$ are presheaves over a space $X$, prove that

$$\beta(\varphi + \psi) = \beta\varphi + \beta\psi \quad \text{and} \quad (\varphi + \psi)\alpha = \varphi\alpha + \psi\alpha.$$

∗ **C-4.28.** Let $(E, p, X)$ be a g-sheaf, and let $\mathcal{F}$ be its sheaf of sections.

(i) Prove that a *subset* $G \subseteq E$ is a sheet if and only if $G = \sigma(U)$ for some open $U \subseteq X$ and $\sigma \in \mathcal{F}(U)$.

(ii) Prove that $G \subseteq E$ is a sheet if and only if $G$ is an open subset of $E$ and $p|G$ is a homeomorphism.

(iii) If $G = \sigma(U)$ and $H = \tau(V)$ are sheets, where $\sigma \in \mathcal{F}(U)$ and $\tau \in \mathcal{F}(V)$, prove that $G \cap H$ is a sheet.

(iv) If $\sigma \in \mathcal{F}(U)$, prove that

$$\mathrm{Supp}(\sigma) = \{x \in X : \sigma(x) \neq 0_x \in E_x\}$$

is a closed subset of $X$.
**Hint.** Consider $\sigma(U) \cap z(U)$, where $z \in \mathcal{F}(U)$ is the zero section.

**C-4.29.** Prove that a g-map $\varphi\colon (E,p,X)\to(E',p',X)$ is an isomorphism in $\mathbf{Sh}_{\mathrm{g}}(X,\mathbf{Ab})$ if and only if $\varphi\colon E\to E'$ is a homeomorphism.

∗ **C-4.30.** Let $\varphi\colon \mathcal{P}\to\mathcal{P}'$ be a presheaf map. Prove that the following statements are equivalent:

(i) $\varphi$ is an isomorphism;

(ii) $\varphi|\mathcal{P}(U)\colon \mathcal{P}(U)\to\mathcal{P}'(U)$ is an isomorphism for every open set $U$;

(iii) $\varphi|\mathcal{P}(U)\colon \mathcal{P}(U)\to\mathcal{P}'(U)$ is a bijection for every open set $U$.

**C-4.31.** Prove that every presheaf of abelian groups $\mathcal{P}$ over a discrete space $X$ is a sheaf.

∗ **C-4.32.** Prove that $\mathbf{Sh}(X,\mathbf{Ab})$ has (infinite) products.

∗ **C-4.33.** Let $x_*G$ be a skyscraper sheaf (see Example C-4.32), where $G$ is an abelian group.

(i) Prove, for every sheaf $\mathcal{G}$, that there is a natural (in $\mathcal{G}$) isomorphism

$$\iota\colon \mathrm{Hom}_{\mathbf{Sh}(X,\mathbf{Ab})}(\mathcal{G},x_*G)\to\mathrm{Hom}_{\mathbb{Z}}(\mathcal{G}_x,G).$$

**Hint.** Use Theorem C-4.40, adjointness of the pair $(f^*,f_*)$.

(ii) Prove, for every abelian group $G$, that $\mathrm{Hom}_{\mathbb{Z}}(\quad,G)$ and $\mathrm{Hom}_{\mathbf{Sh}(X,\mathcal{A})}(\quad,x_*G)$ are naturally equivalent functors.

∗ **C-4.34.** Let $\mathcal{F}$ be a sheaf over a space $X$. Let $x\in X$, let $\overline{\{x\}}$ be its closure, let $G=\mathcal{F}_x$ be the stalk over $x$, and let $i\colon \overline{\{x\}}\to X$ be the inclusion.

(i) Prove that the direct image $i_*\mathcal{F}$ is isomorphic to the skyscraper sheaf $x_*G$.

(ii) Prove that there is a sheaf map $\mu\colon \mathcal{F}\to x_*G$ with $\mu_x\colon \mathcal{F}_x\to(x_*G)_x\cong G$ a monomorphism of abelian groups.

∗ **C-4.35.** Let $\nu\colon \mathcal{P}\to\Gamma(\quad,\mathcal{P}^{\mathrm{g}})$ be the natural map in Theorem C-4.36: in the notation of this theorem, if $U$ is an open set in $X$, then $\nu_U\colon \mathcal{P}(U)\to\Gamma(U,\mathcal{P}^{\mathrm{g}})$ is given by $\sigma\mapsto\sigma^{\mathrm{g}}$. If $x\in X$, prove that $\nu_x\colon \sigma(x)\mapsto\sigma^{\mathrm{g}}(x)=\sigma(x)$.

∗ **C-4.36.** Let $\mathcal{F}$ be a sheaf over a space $X$, and let $f\colon X\to Y$ be continuous. Prove, for all $x\in X$, that

$$(f^*\mathcal{F})_x\cong\mathcal{F}_{f(x)}.$$

Is the analogous isomorphism true for direct image $f_*$?

∗ **C-4.37.** Let $X$ be a topological space and let $\mathcal{B}$ be a base for the topology $\mathcal{U}$ on $X$. Viewing $\mathcal{B}$ as a partially ordered set, we may define a presheaf on $\mathcal{B}$ to be a contravariant functor $\mathcal{Q}\colon \mathcal{B}\to\mathbf{Ab}$. Prove that $\mathcal{Q}$ can be extended to a presheaf $\widetilde{\mathcal{Q}}\colon \mathcal{U}\to\mathbf{Ab}$ by defining

$$\widetilde{\mathcal{Q}}(U)=\varprojlim_{V\in\mathcal{B},\,V\subseteq U}\mathcal{Q}(V).$$

If $U\in\mathcal{B}$, prove that $\widetilde{\mathcal{Q}}(U)$ is canonically isomorphic to $\mathcal{Q}(U)$.

**C-4.38.** Let $\mathcal{S}=(E,p,X)$ be a g-sheaf and let $\mathcal{G}=(G,p|G,X)$, where $G\subseteq E$. Prove that $\Gamma(\quad,\mathcal{G})$ is a sub-g-sheaf of $\Gamma(\quad,\mathcal{S})$ if and only if $G$ is open in $E$ and $G_x=G\cap E_x$ is a subgroup for all $x\in X$.

**C-4.39.** Denote the sheafification functor $\mathbf{pSh}(X,\mathbf{Ab})\to\mathbf{Sh}(X,\mathbf{Ab})$ by $\mathcal{P}\mapsto\mathcal{P}^{\mathrm{g}}$. Prove that ${}^{\mathrm{g}}$ is left adjoint to the inclusion functor $\mathbf{Sh}(X,\mathbf{Ab})\to\mathbf{pSh}(X,\mathbf{Ab})$.

## C-4.5. Sheaf Cohomology

In a word, sheaf cohomology arises as the right derived functors of global sections. We restrict our discussion to sheaves of abelian groups, but the reader should have no problem extending it to sheaves having values in other abelian categories.

If $X$ is a space, global sections defines functors $\Gamma'\colon \mathbf{pSh}(X, \mathbf{Ab}) \to \mathbf{Ab}$ and its restriction $\Gamma\colon \mathbf{Sh}(X, \mathbf{Ab}) \to \mathbf{Ab}$. In either case, the functor is defined on objects by

$$\Gamma(\mathcal{F}) = \Gamma(X, \mathcal{F}) = \mathcal{F}(X)$$

and on (pre)sheaf maps $\varphi = \{\varphi_U\}\colon \mathcal{F} \to \mathcal{G}$ by

$$\Gamma(\varphi)\colon s \mapsto \ \varphi_X(s),$$

where $s \in \Gamma(X, \mathcal{F})$ is a global section. It is clear that each $\Gamma$ is a (covariant) additive functor.

**Lemma C-4.41.** *The functors* $\Gamma'\colon \mathbf{pSh}(X, \mathbf{Ab}) \to \mathbf{Ab}$ *and* $\Gamma\colon \mathbf{Sh}(X, \mathbf{Ab}) \to \mathbf{Ab}$ *are left exact.*

**Proof.** Exactness of presheaves $0 \to \mathcal{P}' \xrightarrow{\varphi} \mathcal{P} \xrightarrow{\psi} \mathcal{P}'' \to 0$ is defined as exactness of the abelian groups $0 \to \mathcal{P}'(U) \xrightarrow{\varphi_U} \mathcal{P}(U) \xrightarrow{\psi_U} \mathcal{P}''(U) \to 0$ for every open $U \subseteq X$. In particular, the sequence is exact when $U = X$, and so $\Gamma$ is even an exact functor on presheaves.

Exactness of sheaves means exactness of stalks, which is usually different than exactness of presheaves. However, if $0 \to \mathcal{F}' \to \mathcal{F} \xrightarrow{\psi} \mathcal{F}''$ is an exact sequence of sheaves, then $\psi$ is a presheaf map, and it follows from Theorem C-4.36 that $\ker\psi$ computed in $\mathbf{Sh}(X, \mathbf{Ab})$ is the same as $\ker\psi$ computed in $\mathbf{pSh}(X, \mathbf{Ab})$. Hence, $0 \to \mathcal{F}' \to \mathcal{F} \to \mathcal{F}''$ is exact in $\mathbf{pSh}(X, \mathbf{Ab})$, and the first paragraph applies. •

The next example shows that the global section functor $\Gamma\colon \mathbf{Sh}(X, \mathbf{Ab}) \to \mathbf{Ab}$ need not be exact.

**Example C-4.42.** In Example C-4.35, we saw that there is an exact sequence of sheaves over the punctured plane $X = \mathbb{C} - \{0\}$,

$$0 \to \mathbb{Z} \to \mathcal{O} \xrightarrow{\varphi} \mathcal{O}^\times \to 0,$$

where $\mathbb{Z}$ is the constant sheaf, $\mathcal{O}$ is the sheaf of germs of analytic functions, $\mathcal{O}^\times$ is the sheaf of nonzero analytic functions, and $\varphi_U\colon \mathcal{O}(U) \to \mathcal{O}^\times(U)$ is given by $f \mapsto e^{2\pi i f}$. For every open set $U$, we have $\mathcal{O}(U)$ the additive group of all analytic $f\colon U \to \mathbb{C}$ and $\mathcal{O}^\times(U)$ the multiplicative group of all never zero analytic $f\colon U \to \mathbb{C}^\times$. If the function $s(z) = z$ in $\Gamma(X, \mathcal{C}^\times)$ is in $\operatorname{im}\varphi^*$ (where $\varphi^*\colon \Gamma(\mathcal{O}) \to \Gamma(\mathcal{O}^\times)$ is the induced map), then $z = e^{2\pi i f(z)}$; that is, $f(z) = \frac{1}{2\pi i}\log(z)$. This is a contradiction, for no branch of $\log(z)$ on the punctured plane is single-valued. Therefore, $\Gamma$ is not an exact functor. ◀

**Definition.** An abelian category $\mathcal{A}$ has ***enough injectives*** if, for every $A \in \operatorname{obj}(\mathcal{A})$, there exist an injective $E$ and a monic $A \to E$. Dually, $\mathcal{A}$ has ***enough projectives*** if, for every $A \in \operatorname{obj}(\mathcal{A})$, there exist a projective $P$ and an epic $P \to A$.

We have seen, in Theorem B-4.64 in Part 1, that for every ring $R$, the module categories ${}_R\mathbf{Mod}$ and $\mathbf{Mod}_R$ have enough injectives (that every module is a quotient of a free module shows that these categories also have enough projectives).

**Proposition C-4.43.** *If $G$ is an injective (i.e., divisible) abelian group and $X$ is a space, then for every $x \in X$, the skyscraper sheaf $x_*G$ is injective in $\mathbf{Sh}(X, \mathbf{Ab})$.*

**Proof.** Recall (Example C-4.32(ii)) that a *skyscraper sheaf* $x_*G$ is defined by

$$(x_*G)(U) = \begin{cases} G & \text{if } x \in U, \\ \{0\} & \text{if } x \notin U. \end{cases}$$

By part (ii) of Exercise C-4.33 on page 377, for every abelian group $G$, the functors $\mathrm{Hom}_{\mathbb{Z}}(\quad, G)$ and $\mathrm{Hom}_{\mathbf{Sh}}(\quad, x_*G)$ are naturally equivalent. Hence, if $G$ is injective, then $\mathrm{Hom}_{\mathbb{Z}}(\quad, G)$ is an exact functor. It follows that $\mathrm{Hom}_{\mathbf{Sh}(X,\mathbf{Ab})}(\quad, x_*G)$ is also an exact functor; that is, $x_*G$ is an injective sheaf. •

**Theorem C-4.44.** *For every space $X$, $\mathbf{Sh}(X, \mathbf{Ab})$ has enough injectives.*[10]

**Proof.** Let $\mathcal{F}$ be a sheaf over $X$. Since $\mathbf{Ab}$ has enough injectives, there is an injective abelian group $G(x)$ and a monic homomorphism $\lambda(x)\colon \mathcal{F}_x \to G(x)$ for each $x \in X$. Now each homomorphism $\lambda(x)$ gives a sheaf map $\Lambda(x)\colon \mathcal{F} \to x_*G(x)$ which, when restricted to the stalk over $x$, has $\Lambda(x) = \lambda(x)$, by Exercise C-4.34 on page 377. By the universal property of products, the sheaf maps $\Lambda(x)$ can be assembled into a sheaf map $\Lambda\colon \mathcal{F} \to \prod_{x\in X} x_*G(x)$ with $\Lambda_x = \Lambda(x)$. But $\Lambda$ is monic, for it is monic on stalks. This completes the proof, for any product of injectives is injective. •

We now define sheaf cohomology as right derived functors of global sections $\Gamma$; this is possible because $\mathbf{Sh}(X)$ has enough injectives, by Theorem C-4.44. Note that taking derived functors of $\Gamma\colon \mathbf{pSh}(X) \to \mathbf{Ab}$ is uninteresting, for the higher derived functors of an exact functor are trivial.

**Definition.** If $X$ is a topological space, then ***sheaf cohomology*** is defined, for every sheaf $\mathcal{F}$ over $X$ and every $q \geq 0$, by

$$H^q(\mathcal{F}) = (R^q\Gamma)(\mathcal{F}).$$

In short, take an injective resolution $\mathcal{E}^\bullet$ of $\mathcal{F}$, delete $\mathcal{F}$ to obtain $\mathcal{E}^\bullet_{\mathcal{F}}$, apply $\Gamma$ to the complex $\mathcal{E}^\bullet_{\mathcal{F}}$, and take homology:

$$H^q(\mathcal{F}) = H^q(\Gamma\mathcal{E}^\bullet_{\mathcal{F}}).$$

As usual, $H^0(\mathcal{F})$ can be computed.

[10] In *The Theory of Sheaves* [**217**], 1964, Swan wrote "... if the base space $X$ is not discrete, I know of no examples of projective sheaves except the zero sheaf." Later, the following was noticed. A generalization of discrete space is an *Alexandrov space*: every possibly infinite intersection of open sets is open. If $X$ is a locally connected space (each $x \in X$ has a connected open neighborhood), then $\mathbf{Sh}(X, \mathbf{Ab})$ has enough projectives if and only if $X$ is an Alexandrov space.

**Proposition C-4.45.** *If $X$ is a topological space, then*

$$H^0(\mathcal{F}) \cong \Gamma(\mathcal{F})$$

*for every sheaf $\mathcal{F}$ over $X$.*

**Proof.** Since $\Gamma$ is a left exact functor, the result follows from Theorem C-3.67. •

Thus, $H^1(\mathcal{F})$ repairs the loss of exactness arising from $\Gamma\colon \mathbf{Sh}(X) \to \mathbf{Ab}$ not being exact; in other words, we may interpret $H^1$ as obstructions, and Example C-4.42 shows that this is interesting. Indeed, sheaf cohomology globalizes the local information in the data of a sheaf.

**Remark.** The global section functor $\Gamma = \Gamma(X, \quad)$ is often modified. A ***family of supports*** $\Phi$ is a family of closed subsets of $X$ such that

(i) whenever $A \in \Phi$ and $B \subseteq A$ is closed, then $B \in \Phi$;

(ii) whenever $A, A' \in \Phi$, then $A \cup A' \in \Phi$.

Define $\Gamma_\Phi(\mathcal{F}) = \{s \in \Gamma(X, \mathcal{F}) : \{x \in X : s(x) \neq 0_x \in E_x\} \in \Phi\}$, where $\mathcal{F}$ has g-sheaf $(E, p, X)$ and $\Phi$ is a family of supports. It is easy to see that $\Gamma_\Phi\colon \mathbf{Sh}(X) \to \mathbf{Ab}$ is a covariant left exact additive functor. One defines ***sheaf cohomology*** $H^q_\Phi$ ***with supports*** $\Phi$ as the right derived functors of $\Gamma_\Phi$. If $\Phi$ is the family of all closed subsets, it is a family of supports, and so $H^q_\Phi$ generalizes sheaf cohomology. ◄

Sheaf cohomology can be computed in the usual ways: one can use various long exact sequences; there is a spectral sequence, due to Leray: if $f\colon X \to Y$ is continuous, then for each sheaf $\mathcal{F}$ over $X$,

$$E_2^{pq} = H^p(R^q f_\bullet \mathcal{F}) \Rightarrow H^{p+q}(\mathcal{F}),$$

where $f_\bullet$ abbreviates iterated application of direct image $f_*$.

There is another construction of cohomology of sheaves, called Čech cohomology (see Rotman [**187**], §6.3.1). Although its definition is complicated, Čech cohomology is more amenable to computation than is sheaf cohomology. Let $\mathcal{F}$ be a sheaf over a space $X$. One first defines, for every open cover $\mathcal{U}$, cohomology groups $\check{H}^\bullet(\mathcal{U}, \mathcal{F})$; second, the family of all open covers of $X$ can be partially ordered (by *refinement*); third, the family of all $\check{H}^\bullet(\mathcal{U}, \mathcal{F})$ can be made into a direct system. Finally,

$$\check{H}^q(\mathcal{F}) = \varinjlim_{\mathcal{U}} \check{H}^q(\mathcal{U}, \mathcal{F}).$$

It is true that $\check{H}^0(\mathcal{F}) = \Gamma(\mathcal{F})$ for every sheaf $\mathcal{F}$, and so $\check{H}^0(\mathcal{F}) \cong H^0(\mathcal{F})$; that is, Čech cohomology and sheaf cohomology agree in degree 0. It is also true that they agree in degree 1: $\check{H}^1(\mathcal{F}) \cong H^1(\mathcal{F})$ (Tennison [**219**], p. 147); however, they can disagree for $q \geq 2$.

Here are some instances when Čech cohomology coincides with sheaf cohomology over an arbitrary (not necessarily Hausdorff) space.

**Theorem C-4.46 (Cartan).** *Let $\mathcal{F}$ be a sheaf over a space $X$. Assume that $\mathcal{U}$ is an open cover of $X$ which contains arbitrarily small open sets and which is closed*

*under finite intersections. If* $\check{H}^q(\mathcal{U}, \mathcal{F}) = \{0\}$ *for all* $U \in \mathcal{U}$ *and all* $q \geq 1$*, then there are isomorphisms for all* $q \geq 0$

$$\check{H}^q(\mathcal{F}) \cong H^q(\mathcal{F}).$$

**Proof.** Godement [**80**], p. 227. •

There are theorems which say that $\check{H}^q(\mathcal{F}) \cong \check{H}^q(\mathcal{U}, \mathcal{F})$ for some open cover $\mathcal{U}$, so that one can avoid direct limits.

Recall the definition of the restriction sheaf: if $\mathcal{F}$ is a sheaf over a space $X$ and if $Y \subseteq X$ is an open subset, then $\mathcal{F}|Y$ is the sheaf over $Y$ defined on open sets $V$ in $Y$ by

$$(\mathcal{F}|Y)(V) = \mathcal{F}(V).$$

**Theorem C-4.47 (Leray).** *Let* $\mathcal{F}$ *be a sheaf over a space* $X$*, and let* $\mathcal{U}$ *be an open cover of* $X$*. If, for every intersection* $Y$ *of finitely many terms of* $\mathcal{U}$*, we have* $\check{H}^q(\mathcal{F}|Y) = \{0\}$ *for all* $q \geq 1$*, then there are isomorphisms for all* $q \geq 0$ *:*

$$\check{H}^q(\mathcal{U}, \mathcal{F}) \cong H^q(\mathcal{F}).$$

**Proof.** Godement [**80**], p. 209. •

**Definition.** A sheaf $\mathcal{L}$ over a space $X$ is ***acyclic*** if $H^q(\mathcal{L}) = \{0\}$ for all $q \geq 1$.

We know that injective sheaves are acyclic, but there are other examples. Acyclic sheaves become especially interesting when there are enough of them; that is, when every sheaf $\mathcal{F}$ can be imbedded in an acyclic sheaf $\mathcal{L}$. The most popular acyclic sheaves are *flabby sheaves*.

**Definition.** A sheaf $\mathcal{L}$ over a space $X$ is ***flabby*** (or *flasque*) if, for each open $U \subseteq X$, every section $s \in \mathcal{L}(U)$ can be extended to a global section.

A ***flabby resolution*** of a sheaf $\mathcal{F}$ is an exact sequence

$$0 \to \mathcal{F} \to \mathcal{L}^0 \to \mathcal{L}^1 \to \cdots$$

in which $\mathcal{L}^q$ is flabby for all $q \geq 0$.

A sheaf $\mathcal{L}$ is flabby if and only if the restriction maps $\Gamma(X, \mathcal{L}) \to \Gamma(U, \mathcal{L})$ are all epic; it follows that the restriction maps $\rho_U^V \colon \Gamma(V, \mathcal{L}) \to \Gamma(U, \mathcal{L})$ are epic for all open sets $U \subseteq V$, because $\rho_V^X \rho_U^V = \rho_U^X$. Hence, if $U \subseteq X$ is open, then $\mathcal{L}$ flabby implies $\mathcal{L}|U$ is also flabby.

**Proposition C-4.48.** *The functor* $\mathcal{L}^0\mathcal{F} \colon U \mapsto \prod_{x \in U} \mathcal{F}_x$ *is a flabby sheaf.*

**Proof.** If $U \subseteq V$, define $\mathcal{L}^0 \colon \mathcal{L}^0(V) \to \mathcal{L}^0(U)$ by $s \mapsto s|U$ (if $s \in \mathcal{L}^0(V) = \prod_{x \in V} \mathcal{F}_x$, then $s \colon V \to \bigcup_{x \in V} \mathcal{F}_x$ is a function with $s(x) \in \mathcal{F}_x$). It is obvious that $\mathcal{L}^0\mathcal{F}$ is a presheaf satisfying the equalizer condition. Moreover, $\mathcal{L}^0\mathcal{F}$ is flabby: since global sections here are merely functions $X \to \prod_{x \in X} \mathcal{F}_x$, every section $s$ over $U$ extends to a global section $s'$; for example, define $s'|U = s$ and, if $x \notin U$, define $s'(x) = 0$. •

**Proposition C-4.49 (Godement).** *Let $\mathcal{F}$ be a sheaf over a space $X$.*

(i) *There is an imbedding $0 \to \mathcal{F} \to \mathcal{L}^0\mathcal{F}$.*

(ii) *There is a flabby resolution*

$$\mathcal{L}^\bullet\mathcal{F} = \ 0 \to \mathcal{F} \to \mathcal{L}^0\mathcal{F} \to \mathcal{L}^1\mathcal{F} \to \cdots .$$

**Proof.**

(i) If $U \subseteq X$ is open, define $\mathcal{F}(U) \to (\mathcal{L}^0\mathcal{F})(U)$ by

$$s \mapsto (s(x)) \in \prod_{x\in U} \mathcal{F}_x = (\mathcal{L}^0\mathcal{F})(U).$$

It is routine to check that this is a sheaf monomorphism.

(ii) We prove, by induction on $q$, that there are flabby sheaves $\mathcal{L}^i\mathcal{F}$ for all $i \le q$ and sheaf maps $d^i\colon \mathcal{L}^i\mathcal{F} \to \mathcal{L}^{i+1}\mathcal{F}$ for $i \le q-1$ such that

$$0 \to \mathcal{F} \to \mathcal{L}^0\mathcal{F} \xrightarrow{d^0} \mathcal{L}^1\mathcal{F} \to \cdots \to \mathcal{L}^{q-1}\mathcal{F} \xrightarrow{d^{q-1}} \mathcal{L}^q\mathcal{F}$$

is exact. We have already defined $\mathcal{L}^0\mathcal{F}$. Define

$$\mathcal{L}^{q+1}\mathcal{F} = \mathcal{L}^0(\operatorname{coker} d^{q-1}),$$

and define $d^q\colon \mathcal{L}^q\mathcal{F} \to \mathcal{L}^{q+1}\mathcal{F}$ as the composite

$$\mathcal{L}^q\mathcal{F} \to \operatorname{coker} d^{q-1} \to \mathcal{L}^0(\operatorname{coker} d^{q-1}) = \mathcal{L}^{q+1}\mathcal{F}.$$

Now $\mathcal{L}^{q+1}\mathcal{F}$ is flabby because it is $\mathcal{L}^0$ of some sheaf, and the sequence is exact because $\operatorname{coker} d^{q-1} \to \mathcal{L}^{q+1}\mathcal{F}$ is monic. •

**Corollary C-4.50.** *Every injective sheaf $\mathcal{E}$ over a space $X$ is flabby.*

**Proof.** It is easy to see that every direct summand of a flabby sheaf is flabby. By Proposition C-4.49(i), there is an exact sequence $0 \to \mathcal{E} \to \mathcal{L} \to \mathcal{L}/\mathcal{E} \to 0$, where $\mathcal{L}$ is flabby. But this sequence splits, because $\mathcal{E}$ is injective; thus, $\mathcal{E}$ is a direct summand of $\mathcal{L}$ and, hence, it is flabby. •

Flabby sheaves give yet another construction of sheaf cohomology.

**Definition.** The flabby resolution $\mathcal{L}^\bullet\mathcal{F}$ in Proposition C-4.49(ii) is called the ***Godement resolution*** of $\mathcal{F}$.

**Proposition C-4.51.** *Let $\mathcal{F}$ be a sheaf over a space $X$.*

(i) *If $0 \to \mathcal{F}' \xrightarrow{\iota} \mathcal{F} \xrightarrow{\varphi} \mathcal{F}'' \to 0$ is an exact sequence of sheaves with $\mathcal{F}'$ flabby, then $0 \to \Gamma(\mathcal{F}') \to \Gamma(\mathcal{F}) \to \Gamma(\mathcal{F}'') \to 0$ is an exact sequence of abelian groups.*

(ii) *Let $0 \to \mathcal{L}' \to \mathcal{L} \to \mathcal{Q} \to 0$ be an exact sequence of sheaves. If $\mathcal{L}'$ and $\mathcal{L}$ are flabby, then $\mathcal{Q}$ is flabby.*

(iii) *Flabby sheaves $\mathcal{L}$ are acyclic.*

(iv) *$H^q(\Gamma(\mathcal{L}^\bullet\mathcal{F})_\mathcal{F}) \cong H^q(\mathcal{F})$ for all $q \ge 0$, where $(\mathcal{L}^\bullet\mathcal{F})_\mathcal{F}$ is the complex obtained from $\mathcal{L}^\bullet\mathcal{F}$ by deleting $\mathcal{F}$.*

**Proof.**

(i) It suffices to prove that $\varphi_X\colon \Gamma(\mathcal{F}) \to \Gamma(\mathcal{F}'')$, given by $\varphi_X\colon s \mapsto \varphi s$, is epic. Let $s'' \in \mathcal{F}''(X) = \Gamma(\mathcal{F}'')$. Define

$$\mathcal{X} = \big\{(U, s) : U \subseteq X \text{ is open}, s \in \mathcal{F}(U), \varphi s = s''|U\big\}.$$

Partially order $\mathcal{X}$ by $(U, s) \preceq (U_1, s_1)$ if $U \subseteq U_1$ and $s_1|U = s$. It is routine to see that chains in $\mathcal{X}$ have upper bounds, and so Zorn's Lemma provides a maximal element $(U_0, s_0)$. If $U_0 = X$, then $s_0$ is a global section and $\varphi_X$ is epic. Otherwise, choose $x \in X$ with $x \notin U_0$. Since $\varphi\colon \mathcal{F} \to \mathcal{F}''$ is an epic sheaf map, it is epic on stalks, and so there are an open $V \subseteq X$ with $V \ni x$ and a section $t \in \mathcal{F}(V)$ with $\varphi t = s''|V$. Now $s - t \in \mathcal{F}'(U \cap V)$ (we regard $\iota\colon \mathcal{F}' \to \mathcal{F}$ as the inclusion), so that $\mathcal{F}'$ flabby provides $r \in \mathcal{F}'(X)$ extending $s - t$. Hence, $s = (t + r)|(U \cap V)$ in $\mathcal{F}(U \cap V)$. Therefore, these sections may be glued: there is $\tilde{s} \in \mathcal{F}(U \cup V)$ with $\tilde{s}|U = s$ and $\tilde{s}|V = (t + r)|(U \cap V)$. But $\varphi(\tilde{s}) = s''$, and this contradicts the maximality of $(U_0, s_0)$.

(ii) Let $U \subseteq X$ be open, and consider the commutative diagram

$$\begin{array}{ccc} \mathcal{F}(X) & \xrightarrow{\varphi_X} & \mathcal{F}''(X) \\ {\scriptstyle\rho}\downarrow & & \downarrow{\scriptstyle\rho''} \\ \mathcal{F}(U) & \xrightarrow[\varphi_U]{} & \mathcal{F}''(U) \end{array}$$

where $\rho, \rho''$ are restriction maps. Since $\mathcal{F}$ is flabby, $\rho$ is epic. We have exactness of $0 \to \mathcal{F}'|U \to \mathcal{F}|U \to \mathcal{F}''|U \to 0$, for exactness of sheaves is stalkwise. As mentioned earlier, $\mathcal{F}'$ flabby implies $\mathcal{F}'|U$ flabby, so that part (i) gives $\varphi_U$ epic. Therefore, the composite $\varphi_U \rho = \rho'' \varphi_X$ is epic, and hence $\rho''$ is epic; that is, $\mathcal{F}''$ is flabby.

(iii) Let $\mathcal{L}$ be flabby. Since there are enough injective sheaves, there is an exact sequence $0 \to \mathcal{L} \to \mathcal{E} \to \mathcal{Q} \to 0$ with $\mathcal{E}$ injective. Now $\mathcal{E}$ is flabby, by Corollary C-4.50, and so $\mathcal{Q}$ is flabby, by part (ii). We prove that $H^q(\mathcal{L}) = \{0\}$ by induction on $q \geq 1$. If $q = 1$, the long exact cohomology sequence contains the fragment

$$H^0(\mathcal{E}) \to H^0(\mathcal{Q}) \to H^1(\mathcal{L}) \to H^1(\mathcal{E}).$$

Since $H^1(\mathcal{E}) = \{0\}$, we have $H^1(\mathcal{L}) = \operatorname{coker}(\Gamma(\mathcal{E})) \to \Gamma(\mathcal{Q})$. But this cokernel is $\{0\}$, by part (i), and so $H^1(\mathcal{L}) = \{0\}$. For the inductive step, consider the fragment

$$H^q(\mathcal{Q}) \to H^{q+1}(\mathcal{L}) \to H^{q+1}(\mathcal{E}).$$

Now $H^{q+1}(\mathcal{E}) = \{0\}$, because $\mathcal{E}$ is injective, while $H^q(\mathcal{Q}) = \{0\}$, by the inductive hypothesis (which applies because $\mathcal{Q}$ is flabby). Therefore, exactness gives $H^{q+1}(\mathcal{L}) = \{0\}$.

(iv) Since the homology functors $H^q(\mathcal{F})$ defined from flabby resolutions are 0 for $q \geq 1$, by part (iii), the result follows from uniqueness (adapt the proof of Theorem C-3.45, replacing $\text{EXT}^n$ by $H^n$). •

There are other interesting classes of acyclic sheaves: for example, *fine sheaves*, defined in terms of locally finite open covers and *partitions of unity*, are acyclic (see Godement [**80**], Chapter II, §3, or Gunning [**87**], p. 36).

The article of Serre [**198**] developed the theory of sheaves over spaces $X$ which need not be Hausdorff. This enabled him to apply sheaves in algebraic geometry; for example, the structure sheaf of a commutative ring $R$ is a sheaf over $X = \operatorname{Spec}(R)$, and $\operatorname{Spec}(R)$ is rarely Hausdorff. Because of the importance of this paper, it has acquired a nickname; it is usually referred to as FAC. We shall say a bit more about structure sheaves in the next chapter.

## C-4.6. Module Categories

When is a category isomorphic to a module category $\mathbf{Mod}_R$?

**Definition.** A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an ***isomorphism*** if there exists a functor $G\colon \mathcal{D} \to \mathcal{C}$ with both composites $GF = 1_{\mathcal{C}}$ and $FG = 1_{\mathcal{D}}$ being identity functors.

Every category is isomorphic to itself; Exercise C-4.43 on page 391 shows that if $R$ is a ring with opposite ring $R^{\mathrm{op}}$, then $\mathbf{Mod}_R$ is isomorphic to ${}_{R^{\mathrm{op}}}\mathbf{Mod}$.

Empirically, isomorphism of functors turns out to be uninteresting. Consider the category $\mathcal{V}$ of all finite-dimensional vector spaces over a field $k$ and its full subcategory $\mathcal{W}$ generated by all vector spaces equal to $k^n$ for $n \in \mathbb{N}$. Since functors take identity morphisms to identity morphisms, an isomorphism $F\colon \mathcal{V} \to \mathcal{W}$ would give a bijection $\operatorname{obj}(\mathcal{V}) \to \operatorname{obj}(\mathcal{W})$. But $\mathcal{W}$ is a small category ($|\operatorname{obj}(\mathcal{W})| = \aleph_0$) while $\mathcal{V}$ is not small, and so these categories are not isomorphic. Carefully distinguishing between two such categories does not seem to be a worthy enterprise.

Recall the following weaker but more useful definition.

**Definition.** A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an ***equivalence*** if there is a functor $G\colon \mathcal{D} \to \mathcal{C}$, called its ***inverse***, such that $GF$ and $FG$ are naturally isomorphic to the identity functors $1_{\mathcal{C}}$ and $1_{\mathcal{D}}$, respectively. When $\mathcal{C}$ and $\mathcal{D}$ are abelian categories, we will further assume that an equivalence $F\colon \mathcal{C} \to \mathcal{D}$ is an additive functor.

Recall Proposition C-4.18:

*A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an equivalence if and only if*

(i) *$F$ is full and faithful: i.e., the function $\operatorname{Hom}_{\mathcal{C}}(C, C') \to \operatorname{Hom}_{\mathcal{D}}(FC, FC')$, given by $f \mapsto Ff$, is a bijection for all $C, C' \in \operatorname{obj}(\mathcal{C})$;*

(ii) *every $D \in \operatorname{obj}(\mathcal{D})$ is isomorphic to $FC$ for some $C \in \operatorname{obj}(\mathcal{C})$.*

**Example C-4.52.**

(i) If $\mathcal{C}$ is a category, let $\mathcal{S} \subseteq \operatorname{obj}(\mathcal{C})$ consist of one object from each isomorphism class of objects. The full subcategory generated by $\mathcal{S}$ (also denoted by $\mathcal{S}$) is called a ***skeletal subcategory*** of $\mathcal{C}$. The inclusion functor $\mathcal{S} \to \mathcal{C}$ is an equivalence, by Proposition C-4.18; thus, every category $\mathcal{C}$ is equivalent to a skeletal subcategory. For example, if $\mathcal{V}$ is the category of all finite-dimensional vector spaces over a field $k$, then the full category $\mathcal{W}$ of $\mathcal{V}$ generated by all

$k^n$ for $n \in \mathbb{N}$ is a skeletal subcategory. Hence, $\mathcal{V}$ and $\mathcal{W}$ are equivalent, but they are not isomorphic.

(ii) If $R$ is a ring with opposite ring $R^{\mathrm{op}}$, then $\mathbf{Mod}_R$ is equivalent to ${}_{R^{\mathrm{op}}}\mathbf{Mod}$, for we have already observed that these categories are isomorphic.

(iii) If $R$ is a ring, then $\mathbf{Mod}_R$ and ${}_R\mathbf{Mod}$ need *not* be equivalent (Exercise C-4.45 on page 391).

(iv) If $\mathcal{V}$ is the category of all finite-dimensional vector spaces over a field $k$, then ***double dual*** $F\colon \mathcal{V} \to \mathcal{V}$, sending $V \mapsto V^{**}$, is an equivalence ($V^* = \mathrm{Hom}_k(V, k)$ is the dual space), for $F$ satisfies the conditions in Proposition C-4.18. ◀

Let us rephrase our original question. When is a category *equivalent* to a module category $\mathbf{Mod}_R$? We shall see that the Wedderburn–Artin Theorems can be better understood in the context of determining those abstract categories that are isomorphic to module categories.

We know that $\mathbf{Mod}_R$, for any ring $R$, is an abelian category; it also has arbitrary direct sums (coproducts).

**Definition.** An abelian category $\mathcal{A}$ is ***cocomplete*** if it contains the coproduct $\bigoplus_{i\in I} A_i$ for every set $(A_i)_{i\in I}$ of objects, where the index set $I$ may be infinite.

We now generalize some categorical properties of the object $R$ in $\mathbf{Mod}_R$.

**Definition.** An object $P$ in an abelian category $\mathcal{A}$ is ***small*** if the covariant Hom functor $\mathrm{Hom}_{\mathcal{A}}(P, \quad)$ preserves (possibly infinite) coproducts.

In more detail, if $P$ is small and $B = \bigoplus_{i\in I} B_i$ has injections $\lambda_i\colon B_i \to B$, then $\mathrm{Hom}\big(P, \bigoplus_{i\in I} B_i\big) = \bigoplus_{i\in I} \mathrm{Hom}(P, B_i)$ has as injections the induced morphisms $(\lambda_i)_*\colon \mathrm{Hom}(P, B_i) \to \mathrm{Hom}(P, B)$.

**Example C-4.53.**

(i) Any finite coproduct of small objects is small, and any direct summand of a small object is small.

(ii) Since every ring $R$ is a small $R$-module, by Corollary B-4.27 in Part 1, it follows from (i) that every finitely generated projective $R$-module is small. ◀

**Definition.** An object $P$ in an abelian category $\mathcal{A}$ is a ***generator*** of $\mathcal{A}$ if every $M \in \mathrm{obj}(\mathcal{A})$ is a quotient of a (possibly infinite) coproduct of copies of $P$.

It is clear that $R$ is a generator of $\mathbf{Mod}_R$, as is any free right $R$-module. However, a projective right $R$-module might not be a generator. For example, if $R = \mathbb{Z}_6$, then $R = P \oplus Q$, where $P \cong \mathbb{Z}_3$ and $Q \cong \mathbb{Z}_2$. The projective module $P$ is not a generator, for $Q \cong \mathbb{Z}_2$ is not a quotient of a direct sum of copies of $P$.

Recall that a functor $F\colon \mathcal{A} \to \mathcal{B}$ is *faithful* if, for all $A, A' \in \mathrm{obj}(\mathcal{A})$, the function $\mathrm{Hom}_{\mathcal{A}}(A, A') \to \mathrm{Hom}_{\mathcal{B}}(FA, FA')$, given by $\varphi \mapsto F\varphi$, is injective.

**Proposition C-4.54.** *An object $P$ in an abelian category $\mathcal{A}$ is a generator of $\mathcal{A}$ if and only if* $\mathrm{Hom}_{\mathcal{A}}(P, \quad)$ *is a faithful functor.*

*In particular, a right $R$-module $P$ is a generator of* $\mathbf{Mod}_R$ *if and only if* $\mathrm{Hom}_R(P, \quad)$ *is a faithful functor.*

**Proof.** We give the proof for $R$-modules, leaving that for abelian categories as a routine exercise.

Assume that $\mathrm{Hom}_R(P, \quad)$ is faithful. Given a right $R$-module $A$ and a map $f\colon P \to A$, let $P_f$ be an isomorphic copy of $P$, and let $Y = \bigoplus_{f \in \mathrm{Hom}_R(P,A)} P_f$. Define $\varphi\colon Y \to A$ by $(g_f) \mapsto \sum_f f(g_f)$. If $\varphi$ is not surjective, then the natural map $\nu\colon A \to A/\operatorname{im}\varphi$ is nonzero. Since $\mathrm{Hom}_R(P, \quad)$ is faithful, $\nu_*\colon \mathrm{Hom}_R(P, A) \to \mathrm{Hom}_R(P, A/\operatorname{im}\varphi)$ is also nonzero. Thus, there is $f \in \mathrm{Hom}_R(P, A)$ such that $\nu_*(f) = \nu f \neq 0$. But $f(A) \subseteq \operatorname{im}\varphi$ so that $\nu f(A) = \{0\}$; that is, $\nu_* f = \nu f = 0$, a contradiction. Therefore, $P$ is a generator.

Conversely, assume that every module $A$ is a quotient of a direct sum $Y = \bigoplus_{i\in I} P_i$, where $P_i \cong P$ for all $i$; say, there is a surjective $\varphi\colon Y \to A$. Now $\varphi = \sum_i \varphi\lambda_i$, where $(\lambda_i\colon P_i \to Y)_{i\in I}$ are the injections. If $\alpha\colon A \to A'$ is nonzero, then $\alpha\varphi = \alpha\sum_i \varphi\lambda_i = \sum_i(\alpha\varphi\lambda_i)$. Now $\alpha\varphi \neq 0$, because $\varphi$ is surjective. Hence, $\alpha\varphi\lambda_i \neq 0$ for some $i$. But $P_i \cong P$, so that $0 \neq \alpha\varphi\lambda_i = \alpha_*(\varphi\lambda_i)$; that is, $\alpha_* \neq 0$, and so $\mathrm{Hom}_R(P, \quad)$ is faithful. •

Here is the characterization of module categories.

**Theorem C-4.55 (Gabriel–Mitchell).** *A category $\mathcal{A}$ is equivalent to a module category* $\mathbf{Mod}_R$ *if and only if $\mathcal{A}$ is a cocomplete abelian category having a small projective generator*[11] *$P$. Moreover, $R \cong \mathrm{End}_{\mathcal{A}}(P)$ in this case.*

**Proof.** The proof of necessity is easy: $\mathbf{Mod}_R$ is a cocomplete abelian category and $R$ is a small projective generator.

For the converse, define $F = \mathrm{Hom}_{\mathcal{A}}(P, \quad)\colon \mathcal{A} \to \mathbf{Ab}$. Note that $F$ is additive and that $R = \mathrm{End}_{\mathcal{A}}(P)$ is a ring, by Exercise C-4.43 on page 391. For each $A \in \mathrm{obj}(\mathcal{A})$, we claim that $FA$ is a right $R$-module. If $f \in FA = \mathrm{Hom}_{\mathcal{A}}(P, A)$ and $\varphi\colon P \to P$ lies in $R = \mathrm{End}(P)$, define scalar multiplication $f\varphi$ to be the composite $P \xrightarrow{\varphi} P \xrightarrow{f} A$. It is routine to check that $F$ actually takes values in $\mathbf{Mod}_R$.

Let us prove that $F$ is an equivalence. Now $F = \mathrm{Hom}_{\mathcal{A}}(P, \quad)\colon \mathcal{A} \to \mathbf{Mod}_R$ (where $R = \mathrm{End}(P)$) is faithful, by Proposition C-4.54. It remains to prove, by Proposition C-4.18, that $F$ is full (i.e., the maps $\mathrm{Hom}_{\mathcal{A}}(Y, X) \to \mathrm{Hom}_R(FY, FX)$, given by $\varphi \mapsto F\varphi$, are all surjective) and that every $M \in \mathrm{obj}(\mathbf{Mod}_R)$ is isomorphic to $FA$ for some $A \in \mathrm{obj}(\mathcal{A})$.

For fixed $Y \in \mathrm{obj}(\mathcal{A})$, define the class

$$\mathcal{C} = \mathcal{C}_Y = \{X \in \mathrm{obj}(\mathcal{A}) : \mathrm{Hom}_{\mathcal{A}}(X, Y) \to \mathrm{Hom}_R(FX, FY) \text{ is surjective}\}.$$

[11] A small projective generator is often called a ***progenerator***.

We will prove three properties of the class $\mathcal{C}$:

(i) $P \in \mathcal{C}$.

(ii) If $(X_i)_{i\in I}$ is a family of objects in $\mathcal{C}$, then $\bigoplus_{i\in I} X_i \in \mathcal{C}$.

(iii) If $X, Z \in \mathcal{C}$ and $f\colon X \to Z$ is any morphism, then $\operatorname{coker} f \in \mathcal{C}$.

These properties imply $\mathcal{C} = \operatorname{obj}(\mathcal{A})$: by (i) and (ii), every coproduct of copies of $P$ lies in $\mathcal{C}$; since $P$ is a generator of $\mathcal{A}$, every $Z \in \operatorname{obj}(\mathcal{A})$ is a cokernel of $\bigoplus_{i\in I} P_i \to \bigoplus_{j\in J} P_j$, where all $P_i, P_j$ are isomorphic to $P$. Thus, $Z \in \mathcal{C}$, which says that $\operatorname{Hom}_{\mathcal{A}}(Z, Y) \to \operatorname{Hom}_{FP}(FZ, FY)$ is surjective; that is, $F$ is full.

We now verify these three properties.

To see that $P \in \mathcal{C}$, we must show that $\operatorname{Hom}_{\mathcal{A}}(P, Y) \to \operatorname{Hom}_R(FP, FY)$ is surjective. Since $P$ is a generator of $\mathcal{A}$, there is an exact sequence

$$\bigoplus_{i\in I} P_i \to \bigoplus_{j\in J} P_j \to Y \to 0 \tag{1}$$

which gives the commutative diagram (details below)

$$\begin{array}{ccccccc}
F(\bigoplus_{i\in I} P_i) & \longrightarrow & F(\bigoplus_{j\in J} P_j) & \longrightarrow & FY & \longrightarrow & 0 \\
\downarrow{\scriptstyle =} & & \downarrow{\scriptstyle =} & & \downarrow{\scriptstyle =} & & \\
\operatorname{Hom}_{\mathcal{A}}(P, \bigoplus_{i\in I} P_i) & \longrightarrow & \operatorname{Hom}_{\mathcal{A}}(P, \bigoplus_{j\in J} P_j) & \longrightarrow & \operatorname{Hom}_{\mathcal{A}}(P, Y) & \longrightarrow & 0 \\
\downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\
\operatorname{Hom}_R(FP, F(\bigoplus_{i\in I} P_i)) & \to & \operatorname{Hom}_R(FP, F(\bigoplus_{j\in J} P_j)) & \to & \operatorname{Hom}_R(FP, FY) & \to & 0 \\
\downarrow & & \downarrow & & \downarrow & & \\
\bigoplus_{i\in I} FP_i & \longrightarrow & \bigoplus_{j\in J} FP_j & \longrightarrow & FY & \longrightarrow & 0.
\end{array}$$

Now $F = \operatorname{Hom}_{\mathcal{A}}(P, \quad)$ is an exact functor (because $P$ is projective), and so the top two rows arise by applying $F$ to Eq. (1); the vertical maps between these rows are identities. The third row arises from applying $\operatorname{Hom}_R(FP, \quad)$ to the top row; the vertical maps are the maps $\operatorname{Hom}(X, Y) \to \operatorname{Hom}(FX, FY)$ given by $\varphi \mapsto F\varphi$. The bottom row arises from the third row, for $F$ preserves direct sums (because $P$ is small), and $\operatorname{Hom}_R(FP, FY) = \operatorname{Hom}_R(R, FY) = FY$. Finally, since $\alpha$ and $\beta$ are surjections, so is $\gamma$ (use the Five Lemma, by adding $\to 0$ at the ends of the middle two rows).

For (ii), let $(X_i)_{i\in I}$ be a family for which all $\operatorname{Hom}(X_i, Y) \to \operatorname{Hom}(FX_i, FY)$ are surjections. To see that $\operatorname{Hom}(\bigoplus X_i, Y) \to \operatorname{Hom}(F(\bigoplus X_i), FY)$ is surjective, use the facts that $\operatorname{Hom}(\bigoplus X_i, Y) \cong \prod \operatorname{Hom}(X_i, Y)$ and $\operatorname{Hom}(F(\bigoplus X_i), FY) \cong \operatorname{Hom}(\bigoplus FX_i, FY) \cong \prod \operatorname{Hom}(FX_i, FY)$ (because $F$ preserves direct sums).

For (iii), use the Five Lemma on a variant of the commutative diagram above. We conclude that $F$ is full.

Lastly, for every right $R$-module $M$, we show that there is $A \in \operatorname{obj}(\mathcal{A})$ with $M \cong FA$. There is an exact sequence $\bigoplus_{i\in I} R_i \xrightarrow{f} \bigoplus_{j\in J} R_j \to M \to 0$ in $\mathbf{Mod}_R$,

where $R_i, R_j$ are isomorphic to $R$. If we view each $R_i, R_j$ as $FP_i$, $FP_j$, where all $P_i, P_j$ are isomorphic to $P$, then

$$f \in \operatorname{Hom}\Big(\bigoplus_{i\in I} R_i, \bigoplus_{j\in J} R_j\Big) = \operatorname{Hom}\Big(\bigoplus_{i\in I} FP_i, \bigoplus_{j\in J} FP_j\Big)$$
$$= \operatorname{Hom}\Big(F\Big(\bigoplus_{i\in I} P_i\Big), F\Big(\bigoplus_{j\in J} P_j\Big)\Big).$$

Since $F$ is full, there is $\varphi \in \operatorname{Hom}(\bigoplus_{i\in I} P_i, \bigoplus_{j\in J} P_j)$ with $F\varphi = f$. Using the Five Lemma again, the reader may show that $M \cong F(\operatorname{coker}\varphi)$. •

**Corollary C-4.56.** *If $R$ is a ring and $n \geq 1$, there is an equivalence of categories*

$$\mathbf{Mod}_R \cong \mathbf{Mod}_{\operatorname{Mat}_n(R)}.$$

**Proof.** For any integer $n \geq 1$, the free module $P = \bigoplus_{i=1}^n R_i$, where $R_i \cong R$, is a small projective generator of $\mathbf{Mod}_R$. Theorem C-4.55 gives an equivalence $\mathbf{Mod}_R \cong \mathbf{Mod}_S$, where $S = \operatorname{End}_R(P) \cong \operatorname{Mat}_n(R)$. •

We can now understand why matrix rings arise in the study of semisimple rings. Proposition C-2.27 says that $\operatorname{Mat}_n(k)$ is semisimple when $k$ is a division ring. By Proposition C-2.23, a ring $R$ is semisimple if and only if every $R$-module is projective; that is, every object in $\mathbf{Mod}_R$ is projective. But every $k$-module is projective (even free), so that equivalence of the categories shows that every object in $\mathbf{Mod}_{\operatorname{Mat}_n(k)}$ is also projective. Therefore, $\operatorname{Mat}_n(k)$ is semisimple.

Given rings $R$ and $S$, Corollary C-4.56 raises the question of when $\mathbf{Mod}_R$ and $\mathbf{Mod}_S$ are equivalent. The answer is given in Morita [**159**], which arose by analyzing the proof of Theorem C-4.55. We merely report the main results; for details, see Jacobson [**111**], pp. 177–184, Lam [**135**], Chapters 18 and 19, McConnell–Robson [**153**], Chapter 3, §5, Reiner [**180**], Chapter 4, or Rowen [**194**], Chapter 4.

**Definition.** Call rings $R$ and $S$ ***Morita equivalent*** if their module categories $\mathbf{Mod}_R$ and $\mathbf{Mod}_S$ are equivalent.

For example, Corollary C-4.56 says that every ring $R$ is Morita equivalent to the matrix ring $\operatorname{Mat}_n(R)$, where $n \geq 1$.

Given a ring $R$, every right $R$-module $P$ determines a ***Morita context***

$$(P, R, Q, S, \alpha, \beta).$$

Here, $Q = \operatorname{Hom}_R(P, R)$ and $S = \operatorname{End}_R(P)$. Both $P$ and $Q$ turn out to be bimodules: $P = {}_SP_R$ and $Q = {}_RQ_S$, and there is an $(R,R)$-map $\alpha\colon Q \otimes_S P \to R$, given by $q \otimes p \mapsto qp$, and an $(R,R)$-map $\beta\colon P \otimes_R Q \to S$, given by $p \otimes q \mapsto pq$. When $P_R$ is a small projective generator, both $\alpha$ and $\beta$ are isomorphisms; in this case, $Q_S$ is also a small projective generator.

A Morita context enables us to construct an equivalence $F$ and its inverse $G$.

**Theorem C-4.57 (Morita I).** *Let $P_R$ be a small projective generator with Morita context $(P, R, Q, S, \alpha, \beta)$.*

(i) $\mathrm{Hom}_R(P, \quad)\colon \mathbf{Mod}_R \to \mathbf{Mod}_S$ *is an equivalence, and its inverse is* $\mathrm{Hom}_S(Q, \quad)\colon \mathbf{Mod}_S \to \mathbf{Mod}_R$.

(ii) $\mathrm{Hom}_R(Q, \quad)\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ *is an equivalence and its inverse is* $\mathrm{Hom}_S(P, \quad)\colon {}_S\mathbf{Mod} \to {}_R\mathbf{Mod}$.

**Proof.** Lam [**135**] states and proves this using $\mathrm{Hom}_R(P, \quad) \cong - \otimes_R Q$ and $\mathrm{Hom}_S(Q, \quad) \cong -\otimes_S P$ in part (i) and $\mathrm{Hom}_S(P, \quad) \cong Q \otimes_S -$ and $\mathrm{Hom}_R(Q, \quad) \cong P \otimes_R -$ in part (ii). •

Equivalences $F$ and $G$ essentially arise as in Morita I.

**Theorem C-4.58 (Morita II).** *Let $R$ and $S$ be rings, and let $F\colon \mathbf{Mod}_R \to \mathbf{Mod}_S$ be an equivalence with inverse $G\colon \mathbf{Mod}_S \to \mathbf{Mod}_R$. Then $F \cong \mathrm{Hom}_R(P, \quad)$ and $G \cong \mathrm{Hom}_S(Q, \quad)$, where $P = G(S)$ and $Q = F(R)$.*

Let's apply Morita II to the categories of left and right $R$-modules.

**Corollary C-4.59.** $\mathbf{Mod}_R$ *and* $\mathbf{Mod}_S$ *are equivalent if and only if* ${}_R\mathbf{Mod}$ *and* ${}_S\mathbf{Mod}$ *are equivalent.*

Exercise C-4.43 on page 391 shows that $\mathbf{Mod}_R$ and ${}_{R^{op}}\mathbf{Mod}$ are equivalent, but Exercise C-4.45 shows that $\mathbf{Mod}_R$ and ${}_R\mathbf{Mod}$ may not be equivalent.

**Corollary C-4.60.** *Two rings $R$ and $S$ are Morita equivalent if and only if $S \cong \mathrm{End}_R(P)$ for some small projective generator $P$ of* $\mathbf{Mod}_R$.

If $\mathcal{A}$ is a category, then an ***endomorphism*** of the identity functor $1_\mathcal{A}$ is a natural transformation $\tau\colon 1_\mathcal{A} \to 1_\mathcal{A}$; that is, for every pair of objects $A$ and $B$ and every morphism $f\colon A \to B$, there is a commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{\tau_A} & A \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f} \\ B & \xrightarrow[\tau_B]{} & B. \end{array}$$

It is easy to see, if $\mathcal{A}$ is an abelian category, that the pointwise sum of two endomorphisms defined as the family $(\tau_A + \sigma_A)_{A \in \mathcal{A}}$ is also an endomorphism.

**Definition.** If $\mathcal{A}$ is an abelian category, define

$$\mathrm{End}(\mathcal{A}) = \{\text{endomorphisms } \tau\colon 1_\mathcal{A} \to 1_\mathcal{A}\}.$$

In particular, if $R$ is a ring, define

$$\mathrm{End}(\mathbf{Mod}_R) = \{\text{endomorphisms } \tau\colon 1_{\mathbf{Mod}_R} \to 1_{\mathbf{Mod}_R}\}.$$

It should be easy to see that $\mathrm{End}(\mathbf{Mod}_R)$ is a ring with composition as multiplication, but it is not obvious whether $\mathrm{End}(\mathbf{Mod}_R)$ is a set.

**Proposition C-4.61.** *For any ring $R$, there is a ring isomorphism*

$$Z(R) \cong \mathrm{End}(\mathbf{Mod}_R).$$

**Proof.** If $c \in Z(R)$ and $A$ is a right $R$-module, define $\tau_A^c\colon A \to A$ to be multiplication by $c$:

$$\tau_A^c\colon a \mapsto ac.$$

Since $c \in Z(R)$, the function $\tau_A^c$ is an $R$-map. It is easily checked that $\tau^c = (\tau_A^c)_{A\in\mathbf{Mod}_R}$ is an endomorphism of $1_{\mathbf{Mod}_R}$. Define $\varphi\colon Z(R) \to \mathrm{End}(\mathbf{Mod}_R)$ by

$$\varphi\colon c \mapsto \tau^c = (\tau_A^c)_{A\in\mathbf{Mod}_R}.$$

We claim that $\varphi$ is a bijection (so that $\mathrm{End}(\mathbf{Mod}_R)$ is a set) and a ring isomorphism. The only point which is not obvious is whether $\varphi$ is surjective. Let $\sigma$ be an endomorphism of $1_{\mathcal{A}}$ and let $A$ be a right $R$-module. If $a \in A$, define $f\colon R \to A$ by $f(1) = a$. There is a commutative diagram

$$\begin{array}{ccc} R & \xrightarrow{\sigma_R} & R \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f} \\ A & \xrightarrow[\sigma_A]{} & A. \end{array}$$

Define $c = \sigma_R(1)$. Now $f\sigma_R(1) = f(c) = f(1 \cdot c) = f(1)c = ac$. On the other hand, $\sigma_A f(1) = \sigma_A(a)$. Commutativity gives $\sigma_A(a) = ac$; that is, $\sigma_A = \tau_A^c$. •

**Corollary C-4.62.** *Let $R$ and $S$ be rings.*

(i) *If $R$ and $S$ are Morita equivalent, then $Z(R) \cong Z(S)$.*

(ii) *If $R$ and $S$ are commutative, then $\mathbf{Mod}_R$ and $\mathbf{Mod}_S$ are equivalent if and only if $R \cong S$.*

**Proof.**

(i) If $\mathcal{A}$ and $\mathcal{B}$ are equivalent abelian categories, then $\mathrm{End}(1_{\mathcal{A}}) \cong \mathrm{End}(1_{\mathcal{B}})$. If $\mathcal{A} = \mathbf{Mod}_R$ and $\mathcal{B} = \mathbf{Mod}_S$, then $Z(R) \cong Z(S)$, by Proposition C-4.61.

(ii) Sufficiency is obvious. For necessity, part (i) gives $Z(R) \cong Z(S)$. Since $R$ and $S$ are commutative, $R = Z(R) \cong Z(S) = S$. •

This last result indicates why the Brauer group $\mathrm{Br}(k)$ is involved in investigating central simple algebras instead of only division rings: as we survey the categories ${}_{\Delta}\mathbf{Mod}$, where $\Delta$ is a division ring, then the center $k$ of each $\Delta$ is a categorical invariant.

## Exercises

**C-4.40.** Let $F\colon \mathcal{C} \to \mathcal{D}$ be an *isomorphism* of categories with inverse $G\colon \mathcal{D} \to \mathcal{C}$; that is, $GF = 1_{\mathcal{C}}$ and $FG = 1_{\mathcal{D}}$. Prove that both $(F, G)$ and $(G, F)$ are adjoint pairs.

**C-4.41.** If $F\colon \mathcal{A} \to \mathcal{B}$ is an equivalence of abelian categories, prove the following statements:

(i) If $f$ is monic in $\mathcal{A}$, then $Ff$ is monic in $\mathcal{B}$.

(ii) If $f$ is epic in $\mathcal{A}$, then $Ff$ is epic in $\mathcal{B}$.

(iii) If $A \in \operatorname{obj}(\mathcal{A})$, then $f \mapsto Ff$ is a ring isomorphism $\operatorname{End}(A) \to \operatorname{End}(FA)$.

(iv) If $0 \to A' \to A \to A'' \to 0$ is exact in $\mathcal{A}$, then $0 \to FA' \to FA \to FA'' \to 0$ is exact in $\mathcal{B}$. Moreover, the first sequence is split if and only if the second sequence is split.

(v) If $(A_i)_{i\in I}$ is a family of objects in $\mathcal{A}$, then

$$F\Big(\bigoplus_{i\in I} A_i\Big) \cong \bigoplus_{i\in I} FA_i \quad \text{and} \quad F\Big(\prod_{i\in I} A_i\Big) \cong \prod_{i\in I} FA_i.$$

(vi) If $P$ is projective in $\mathcal{A}$, then $FP$ is projective in $\mathcal{B}$.

(vii) If $P$ is injective in $\mathcal{A}$, then $FP$ is injective in $\mathcal{B}$.

**C-4.42.** Let $F\colon \mathbf{Mod}_R \to \mathbf{Mod}_S$ be an equivalence. Prove that a right $R$-module $A$ has any of the following properties if and only if $FA$ does: simple, semisimple, ACC, DCC, indecomposable. Moreover, $A$ has a composition series if and only if $FA$ does; both have the same length, and $S_1, \dots, S_n$ are composition factors of $A$ if and only if $FS_1, \dots, FS_n$ are composition factors of $FA$.

∗ **C-4.43.** If $R$ is a ring with opposite ring $R^{\mathrm{op}}$, prove that $\mathbf{Mod}_R$ is equivalent to ${}_{R^{\mathrm{op}}}\mathbf{Mod}$.

**Hint.** For each right $R$-module $M$, show that there is an isomorphism $\tau$ with $\tau(M) = M'$, where $M'$ is $M$ made into a left $R$-module as in Proposition B-1.23 in Part 1.

∗ **C-4.44.** Prove that every small projective generator $P$ of $\mathbf{Mod}_R$ is finitely generated.

∗ **C-4.45.** (i) Let $R$ and $S$ be rings, and let $F\colon \mathbf{Mod}_R \to \mathbf{Mod}_S$ (or $F\colon \mathbf{Mod}_R \to {}_S\mathbf{Mod}$) be an equivalence. Prove that a right $R$-module $M$ is finitely generated if and only if $FM$ is a finitely generated right (or left) $S$-module.

**Hint.** Use Exercise B-1.44 on page 300 in Part 1.

(ii) Call a category $\mathbf{Mod}_R$ (or ${}_R\mathbf{Mod}$) ***noetherian*** if every submodule of a finitely generated right (or left) $R$-module $M$ is finitely generated. Let $\mathcal{A}$ and $\mathcal{B}$ be equivalent categories of modules; that is, $\mathcal{A}$ is equivalent to $\mathbf{Mod}_R$ or ${}_R\mathbf{Mod}$ for some ring $R$, and $\mathcal{B}$ is equivalent to $\mathbf{Mod}_S$ or ${}_S\mathbf{Mod}$ for some ring $S$. Prove that $\mathcal{A}$ is noetherian if and only if $\mathcal{B}$ is noetherian.

(iii) Prove that $\mathbf{Mod}_R$ is a noetherian category if and only if $R$ is a right noetherian ring and that ${}_R\mathbf{Mod}$ is a noetherian category if and only if $R$ is a left noetherian ring.

(iv) Give an example of a ring $R$ such that ${}_R\mathbf{Mod}$ and $\mathbf{Mod}_R$ are not equivalent.
**Hint.** Let $R$ be the ring in Exercise B-1.28 on page 288 in Part 1 which is left noetherian but not right noetherian.

## C-4.7. Adjoint Functor Theorem for Modules

Recall the Adjoint Isomorphism, Theorem B-4.98 in Part 1: given modules $A_R$, ${}_RB_S$, and $C_S$, there is a natural isomorphism

$$\tau_{A,B,C}\colon \operatorname{Hom}_S(A \otimes_R B, C) \to \operatorname{Hom}_R(A, \operatorname{Hom}_S(B, C)).$$

Write $F = -\otimes_R B$ and $G = \operatorname{Hom}_S(B, \quad)$, so that the isomorphism reads

$$\operatorname{Hom}_S(FA, C) \cong \operatorname{Hom}_R(A, GC).$$

If we pretend that $\operatorname{Hom}_S(\quad,\quad)$ and $\operatorname{Hom}_R(\quad,\quad)$ are inner products on vector spaces $V, W$, then this reminds us of the *adjoint* of a linear transformation $T\colon V \to W$, the linear transformation $T^*\colon W \to V$ such that

$$(Tv, w) = (v, T^*w)$$

for all $v \in V$ and $w \in W$. This analogy explains why the isomorphism $\tau$ is called the *adjoint isomorphism*.

**Definition.** Let $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$ be covariant functors. The ordered pair $(F, G)$ is an ***adjoint pair*** if, for each $C \in \operatorname{obj}(\mathcal{C})$ and $D \in \operatorname{obj}(\mathcal{D})$, there are bijections

$$\tau_{C,D}\colon \operatorname{Hom}_{\mathcal{D}}(FC, D) \to \operatorname{Hom}_{\mathcal{C}}(C, GD)$$

that are natural transformations in $C$ and in $D$.

In more detail, naturality says that the following two diagrams commute for all $f\colon C' \to C$ in $\mathcal{C}$ and $g\colon D \to D'$ in $\mathcal{D}$:

$$\begin{array}{ccc} \operatorname{Hom}_{\mathcal{D}}(FC, D) & \xrightarrow{(Ff)^*} & \operatorname{Hom}_{\mathcal{D}}(FC', D) \\ \downarrow{\scriptstyle \tau_{C,D}} & & \downarrow{\scriptstyle \tau_{C',D}} \\ \operatorname{Hom}_{\mathcal{C}}(C, GD) & \xrightarrow[f^*]{} & \operatorname{Hom}_{\mathcal{C}}(C', GD), \end{array}$$

$$\begin{array}{ccc} \operatorname{Hom}_{\mathcal{D}}(FC, D) & \xrightarrow{g_*} & \operatorname{Hom}_{\mathcal{D}}(FC, D') \\ \downarrow{\scriptstyle \tau_{C,D}} & & \downarrow{\scriptstyle \tau_{C,D'}} \\ \operatorname{Hom}_{\mathcal{C}}(C, GD) & \xrightarrow[(Gg)_*]{} & \operatorname{Hom}_{\mathcal{C}}(C, GD'). \end{array}$$

**Example C-4.63.**

(i) If $B = {}_RB_S$ is a bimodule, then $\big(-\otimes_R B, \operatorname{Hom}_S(B, \quad)\big)$ is an adjoint pair, by Theorem B-4.98 in Part 1. Similarly, if $B = {}_SB_R$ is a bimodule, then $\big(B \otimes_R -, \operatorname{Hom}_S(B, \quad)\big)$ is an adjoint pair, by Theorem B-4.99 in Part 1.

(ii) Let $U\colon$ **Groups** $\to$ **Sets** be the ***forgetful functor*** which assigns to each group $G$ its underlying set and views each homomorphism as a mere function. Let $F\colon$ **Sets** $\to$ **Groups** be the ***free functor*** defined in Exercise B-4.23 on page 474 in Part 1, which assigns to each set $X$ the free group $FX$ having basis $X$. The function

$$\tau_{X,H}\colon \operatorname{Hom}_{\mathbf{Groups}}(FX, H) \to \operatorname{Hom}_{\mathbf{Sets}}(X, UH),$$

given by $f \mapsto f|X$, is a bijection (its inverse is $\varphi \mapsto \widetilde{\varphi}$, where $X$, being a basis of $FX$, says that every function $\varphi\colon X \to H$ corresponds to a unique homomorphism $\widetilde{\varphi}\colon FX \to H$). Indeed, $\tau_{X,H}$ is a natural bijection, showing that $(F, U)$ is an adjoint pair of functors. This example can be generalized by replacing **Groups** by other categories having free objects, e.g., ${}_R\mathbf{Mod}$ or $\mathbf{Mod}_R$.

(iii) If $U\colon \mathbf{ComRings} \to \mathbf{Sets}$ is the forgetful functor, then $(F, U)$ is an adjoint pair where, for any set $X$, we have $F(X) = \mathbb{Z}[X]$, the ring of all polynomials in commuting variables $X$. More generally, if $k$ is a commutative ring and $\mathbf{ComAlg}_k$ is the category of all commutative $k$-algebras, then $F(X) = k[X]$, the polynomial ring over $k$. This is essentially the same example as in (ii), for $k[X]$ is the free $k$-algebra on $X$. ◀

For many examples of adjoint pairs of functors, see Herrlich–Strecker [**96**], p. 197, and Mac Lane [**144**], Chapter 4, especially pp. 85–86.

**Example C-4.64.** Adjointness is a property of an *ordered pair* of functors; if $(F, G)$ is an adjoint pair of functors, it does not follow that $(G, F)$ is also an adjoint pair. For example, if $F = - \otimes B$ and $G = \mathrm{Hom}(B, \quad)$, then the adjoint isomorphism says that $\mathrm{Hom}(A, B \otimes C) \cong \mathrm{Hom}(A, \mathrm{Hom}(B, C))$ for all $A$ and $C$; that is, $\mathrm{Hom}(FA, C) \cong \mathrm{Hom}(A, GC)$. It does not say that there is an isomorphism (natural or not) $\mathrm{Hom}(\mathrm{Hom}(B, A), C) \cong \mathrm{Hom}(A, B \otimes C)$. Indeed, if $A = \mathbb{Q}, B = \mathbb{Q}/\mathbb{Z}$, and $C = \mathbb{Z}$, then $\mathrm{Hom}(G\mathbb{Q}, \mathbb{Z}) \ncong \mathrm{Hom}(\mathbb{Q}, F\mathbb{Z})$; that is,

$$\mathrm{Hom}\big(\mathrm{Hom}(\mathbb{Q}/\mathbb{Z}, \mathbb{Q}), \mathbb{Z}\big) \ncong \mathrm{Hom}\big(\mathbb{Q}, (\mathbb{Q}/\mathbb{Z}) \otimes \mathbb{Z}\big),$$

for the left side is $\{0\}$, while the right side is isomorphic to $\mathrm{Hom}(\mathbb{Q}, \mathbb{Q}/\mathbb{Z})$, which is not zero because it contains the natural map $\mathbb{Q} \to \mathbb{Q}/\mathbb{Z}$. ◀

**Definition.** Let $\mathcal{C} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{D}$ be functors. If $(F, G)$ is an adjoint pair, then we say that $F$ has a ***right adjoint*** $G$ and that $G$ has a ***left adjoint*** $F$.

Let $(F, G)$ be an adjoint pair, where $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$. If $C \in \mathrm{obj}(\mathcal{C})$, then setting $D = FC$ gives a bijection $\tau\colon \mathrm{Hom}_{\mathcal{D}}(FC, FC) \to \mathrm{Hom}_{\mathcal{C}}(C, GFC)$, so that $\eta_C$, defined by

$$\eta_C = \tau(1_{FC}),$$

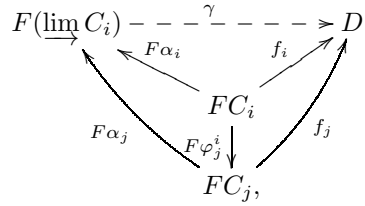
is a morphism $C \to GFC$. Exercise C-4.47 on page 402 says that $\eta\colon 1_{\mathcal{C}} \to GF$ is a natural transformation; it is called the ***unit*** of the adjoint pair.

**Theorem C-4.65.** *Let $(F, G)$ be an adjoint pair of functors, where $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$. Then $F$ preserves direct limits and $G$ preserves inverse limits.*

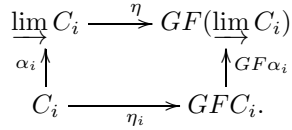
**Remark.** There is no restriction on the index sets of the limits; in particular, they need not be directed. ◀

**Proof.** Let $I$ be a partially ordered set, and let $\{C_i, \varphi_j^i\}$ be a direct system in $\mathcal{C}$ over $I$. By Exercise B-7.2 on page 670 in Part 1, $\{FC_i, F\varphi_j^i\}$ is a direct system in

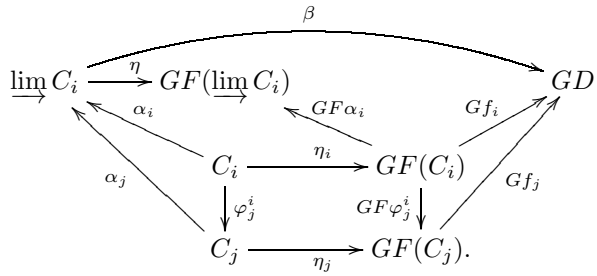
$\mathcal{D}$ over $I$. Consider the following diagram in $\mathcal{D}$:




where $\alpha_i\colon C_i \to \varinjlim C_i$ are the maps in the definition of direct limit. We must show that there exists a unique morphism $\gamma\colon F(\varinjlim C_i) \to D$ making the diagram commute. The idea is to apply $G$ to this diagram and use the unit $\eta\colon 1_{\mathcal{C}} \to GF$ to replace $GF(\varinjlim C_i)$ and $GFC_i$ by $\varinjlim C_i$ and $C_i$, respectively. In more detail, Exercise C-4.47 on page 402 gives morphisms $\eta$ and $\eta_i$ making the following diagram commute:




Apply $G$ to the original diagram and adjoin this diagram to its left:




This diagram commutes: we know that $(GF\varphi^i_j)\eta_i = \eta_j\varphi^i_j$, since $\eta$ is natural, and $Gf_i = Gf_j(GF\varphi^i_j)$, since $G$ is a functor; therefore, $Gf_i\eta_i = Gf_j(GF\varphi^i_j)\eta_i = Gf_j\eta_j\varphi^i_j$. By the definition of direct limit, there exists a unique $\beta\colon \varinjlim C_i \to GD$ (that is, $\beta \in \mathrm{Hom}_{\mathcal{C}}(\varinjlim C_i, GD)$) making the diagram commute. Since $(F, G)$ is an adjoint pair, there exists a natural bijection

$$\tau\colon \mathrm{Hom}_{\mathcal{D}}(F(\varinjlim C_i), D) \to \mathrm{Hom}_{\mathcal{C}}(\varinjlim C_i, GD).$$

Define

$$\gamma = \tau^{-1}(\beta) \in \mathrm{Hom}_{\mathcal{D}}(F(\varinjlim C_i), D).$$

We claim that $\gamma\colon F(\varinjlim C_i) \to D$ makes the first diagram commute. The first commutative square in the definition of adjointness, which involves the morphism

$\alpha_i\colon C_i \to \varinjlim C_i$, gives commutativity of

$$\begin{array}{ccc} \mathrm{Hom}_{\mathcal{D}}(F(\varinjlim C_i), D) & \xrightarrow{(F\alpha_i)^*} & \mathrm{Hom}_{\mathcal{D}}(FC_i, D) \\ \Big\downarrow{\scriptstyle \tau} & & \Big\downarrow{\scriptstyle \tau} \\ \mathrm{Hom}_{\mathcal{C}}(\varinjlim C_i, GD) & \xrightarrow[\alpha_i^*]{} & \mathrm{Hom}_{\mathcal{C}}(C_i, GD). \end{array}$$

Thus, $\tau(F\alpha_i)^* = \alpha_i^*\tau$, and so $\tau^{-1}\alpha_i^* = (F\alpha_i)^*\tau^{-1}$. Evaluating on $\beta$, we have

$$(F\alpha_i)^*\tau^{-1}(\beta) = (F\alpha_i)^*\gamma = \gamma F\alpha_i.$$

On the other hand, since $\beta\alpha_i = (Gf_i)\eta_i$, we have

$$\tau^{-1}\alpha_i^*(\beta) = \tau^{-1}(\beta\alpha_i) = \tau^{-1}((Gf_i)\eta_i).$$

Therefore,

$$\gamma F\alpha_i = \tau^{-1}((Gf_i)\eta_i).$$

The second commutative square in the definition of adjointness commutes, for the morphism $f_i\colon FC_i \to D$ gives commutativity of

$$\begin{array}{ccc} \mathrm{Hom}_{\mathcal{D}}(FC_i, FC_i) & \xrightarrow{(f_i)_*} & \mathrm{Hom}_{\mathcal{D}}(FC_i, D) \\ \Big\downarrow{\scriptstyle \tau} & & \Big\downarrow{\scriptstyle \tau} \\ \mathrm{Hom}_{\mathcal{C}}(C_i, GFC_i) & \xrightarrow[(Gf_i)_*]{} & \mathrm{Hom}_{\mathcal{C}}(C_i, GD); \end{array}$$

that is, $\tau(f_i)_* = (Gf_i)_*\tau$. Evaluating at $1_{FC_i}$, the definition of $\eta_i$ gives $\tau(f_i)_*(1) = (Gf_i)_*\tau(1)$, and so $\tau f_i = (Gf_i)_*\eta_i$. Therefore,

$$\gamma F\alpha_i = \tau^{-1}((Gf_i)\eta_i) = \tau^{-1}\tau f_i = f_i,$$

so that $\gamma$ makes the original diagram commute. We leave the proof of the uniqueness of $\gamma$ as an exercise for the reader, with the hint to use the uniqueness of $\beta$.

The dual proof shows that $G$ preserves inverse limits. •

We are now going to characterize the Hom and tensor functors on module categories, yielding a necessary and sufficient condition for a functor on such categories to be half of an adjoint pair (Theorems C-4.73 and C-4.74).

**Lemma C-4.66.**

(i) *If $M$ is a right $R$-module and $m \in M$, then $\varphi_m\colon R \to M$, defined by $r \mapsto mr$, is a map of right $R$-modules. In particular, if $M = R$ and $u \in R$, then $\varphi_u\colon R \to R$ is a map of right $R$-modules.*

(ii) *If $M$ is a right $R$-module, $m \in M$, and $u \in R$, then*

$$\varphi_{mu} = \varphi_m\varphi_u.$$

(iii) *Let $f\colon M \to N$ be an $R$-map between right $R$-modules. If $m \in M$, then*

$$\varphi_{fm} = f\varphi_m.$$

**Proof.**

(i) $\varphi_m$ is additive because $m(r+s) = mr+ms$; $\varphi_m$ preserves scalar multiplication on the right because $\varphi_m(rs) = m(rs) = (mr)s = \varphi_m(r)s$.

(ii) Now $\varphi_{mr}\colon u \mapsto (mr)u$, while $\varphi_m\varphi_r\colon u \mapsto \varphi_m(ru) = m(ru)$. These values agree because $M$ is a right $R$-module.

(iii) Now $\varphi_{fm}\colon u \mapsto (fm)u$, while $f\varphi_m\colon u \mapsto f(mu)$. These values agree because $f$ is an $R$-map. •

**Theorem C-4.67 (Watts).** *If $F\colon \mathbf{Mod}_R \to \mathbf{Ab}$ is a right exact additive functor that preserves direct sums, then $F$ is naturally isomorphic to $-\otimes_R B$, where $B$ is $F(R)$ made into a left $R$-module.*

**Proof.** We begin by making the abelian group $FR$ (our abbreviation for $F(R)$) into a left $R$-module. If $M$ is a right $R$-module and $m \in M$, then $\varphi_m\colon R \to M$, defined by $r \mapsto mr$, is an $R$-map, by Lemma C-4.66(i), and so the $\mathbb{Z}$-map $F\varphi_m\colon FR \to FM$ is defined. In particular, if $M = R$ and $u \in R$, then $\varphi_u\colon R \to R$ and, for all $x \in FR$, we define $ux$ by

$$ux = (F\varphi_u)x.$$

Let us show that this scalar multiplication makes $FR$ into a left $R$-module. If $M = R$ and $u, v \in R$, then $F\varphi_u, F\varphi_v\colon FR \to FR$, and Lemma C-4.66(ii) gives $\varphi_{uv} = \varphi_u\varphi_v$. Hence,

$$(uv)x = (F\varphi_{uv})x = F(\varphi_u\varphi_v)x = (F\varphi_u)(F\varphi_v)x = u(vx).$$

Denote the left $R$-module $FR$ by $B$, so that $-\otimes_R B\colon \mathbf{Mod}_R \to \mathbf{Ab}$. We claim that $\tau_M\colon M \times FR \to FM$, defined by $(m, x) \mapsto (F\varphi_m)x$, is $R$-biadditive; that is, $\tau_M(mu, x) = \tau_M(m, ux)$ for all $u \in R$. Now

$$\tau_M(mu, x) = (F\varphi_{mu})x = F(\varphi_m\varphi_u)x,$$

by Lemma C-4.66(ii). On the other hand,

$$\tau_M(m, ux) = (F\varphi_m)ux = (F\varphi_m)(F\varphi_u)x = (F\varphi_{mu})x.$$

Thus, $\tau_M$ induces a homomorphism $\sigma_M\colon M \otimes_R B \to FM$. We claim that the family $\sigma = (\sigma_M)_{M\in\mathbf{Mod}_R}$ is a natural transformation $\sigma\colon -\otimes_R B \to F$; that is, the following diagram commutes for $R$-maps $f\colon M \to N$:

$$\begin{array}{ccc} M\otimes_R B & \xrightarrow{\sigma_M} & FM \\ {\scriptstyle f\otimes 1}\downarrow & & \downarrow{\scriptstyle Ff} \\ N\otimes_R B & \xrightarrow[\sigma_N]{} & FN. \end{array}$$

Going clockwise, $m \otimes x \mapsto (F\varphi_m)x \mapsto (Ff)(F\varphi_m)x$; going counterclockwise,

$$m \otimes x \mapsto f(m) \otimes x \mapsto (F\varphi_{fm})x = F(f\varphi_m)x = (Ff)(F\varphi_m)x,$$

by Lemma C-4.66(iii).

Now $\sigma_R\colon R\otimes_R B \to FR$ is an isomorphism (because $B = FR$); moreover, since both $-\otimes_R B$ and $F$ preserve direct sums, $\sigma_A\colon A \otimes_R B \to FA$ is an isomorphism

for every free right $R$-module $A$. Let $M$ be any right $R$-module. There are a free right $R$-module $A$ and a short exact sequence

$$0 \to K \xrightarrow{i} A \to M \to 0;$$

there is also a surjection $f\colon C \to K$ for some free right $R$-module $C$. Splicing these together, there is an exact sequence

$$C \xrightarrow{if} A \to M \to 0.$$

Now the following commutative diagram has exact rows, for both $- \otimes_R B$ and $F$ are right exact:

$$\begin{array}{ccccccc} C \otimes_R B & \longrightarrow & A \otimes_R B & \longrightarrow & M \otimes_R B & \longrightarrow & 0 \\ \downarrow{\scriptstyle \sigma_C} & & \downarrow{\scriptstyle \sigma_A} & & \downarrow{\scriptstyle \sigma_M} & & \\ FC & \longrightarrow & FA & \longrightarrow & FM & \longrightarrow & 0. \end{array}$$

Since $\sigma_C$ and $\sigma_A$ are isomorphisms, the Five Lemma shows that $\sigma_M$ is an isomorphism. Therefore, $\sigma$ is a natural isomorphism. •

**Remark.** If, in Theorem C-4.67, $F$ takes values in $\mathbf{Mod}_S$ instead of in **Ab**, then the first paragraph of the proof can be modified to prove that the right $S$-module $FR$ may be construed as an $(R, S)$-bimodule; thus, the theorem remains true if **Ab** is replaced by $\mathbf{Mod}_S$. ◄

**Example C-4.68.** If $R$ is a commutative ring and $r \in R$, then there is a functor $F\colon {}_R\mathbf{Mod} \to {}_R\mathbf{Mod}$ that takes an $R$-module $M$ to $M/rM$ (if $\varphi\colon M \to N$ is an $R$-map, define $F\varphi\colon M/rM \to N/rN$ by $m + rM \mapsto \varphi(m) + rN$). The reader may check that $F$ is a right exact functor preserving direct sums, and so it follows from Watts' Theorem that $F$ is naturally isomorphic to $- \otimes_R (R/rR)$, for $FR = R/rR$. This generalizes Proposition B-4.91 in Part 1. ◄

**Corollary C-4.69.** *Let $R$ be a right noetherian ring, and let $\mathcal{F}_R$ be the category of all finitely generated right $R$-modules. If $F\colon \mathcal{F}_R \to \mathbf{Mod}_S$ is a right exact additive functor, then $F$ is naturally isomorphic to $- \otimes_R B$, where $B$ is $F(R)$ made into a left $R$-module.*

**Proof.** The proof is almost the same as that of Theorem C-4.67 coupled with the remark after it. Given a finitely generated right $R$-module $M$, we can choose a finitely generated free right $R$-module $A$ mapping onto $M$. Moreover, since $R$ is right noetherian, Proposition B-1.34 in Part 1 shows that the kernel $K$ of the surjection $A \to M$ is also finitely generated (if $K$ were not finitely generated, then there would be no free right $R$-module in the category $\mathcal{F}_R$ mapping onto $K$). Finally, we need not assume that $F$ preserves finite direct sums, for Lemma C-4.1 shows that this follows from the additivity of $F$. •

We now characterize contravariant Hom functors.

**Theorem C-4.70 (Watts).** *If $H\colon {}_R\mathbf{Mod} \to \mathbf{Ab}$ is a contravariant left exact additive functor that converts direct sums to direct products, then $H$ is naturally isomorphic to* $\mathrm{Hom}_R(\ \ , B)$*, where $B$ is $H(R)$ made into a right $R$-module.*

**Proof.** We begin by making the abelian group $HR$ into a right $R$-module. As in the beginning of the proof of Theorem C-4.67, if $M$ is a right $R$-module and $m \in M$, then the function $\varphi_m\colon R \to M$, defined by $r \mapsto mr$, is an $R$-map. In particular, if $M = R$ and $u \in R$, then $H\varphi_u\colon HR \to HR$, and Lemma C-4.66(ii) gives $\varphi_{uv} = \varphi_u\varphi_v$ for all $u, v \in R$. If $x \in HR$, define

$$ux = (H\varphi_u)x.$$

Here, $HR$ is a *right* $R$-module, for the contravariance of $H$ gives

$$(uv)x = (H\varphi_{uv})x = H(\varphi_u\varphi_v)x = (H\varphi_v)(H\varphi_u)x = v(ux).$$

Define $\sigma_M\colon HM \to \mathrm{Hom}_R(M, B)$ by $\sigma_M(x)\colon m \mapsto (H\varphi_m)x$, where $x \in HM$. It is easy to check that $\sigma\colon H \to \mathrm{Hom}_R(\ \ , B)$ is a natural transformation and that $\sigma_R$ is an isomorphism. The remainder of the proof proceeds, *mutatis mutandis*, as that of Theorem C-4.67. •

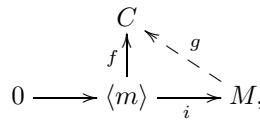
We can characterize covariant Hom functors, but the proof is a bit more complicated.

**Definition.** A left $R$-module $C$ is called a ***cogenerator*** of ${}_R\mathbf{Mod}$ if, for every left $R$-module $M$ and every nonzero $m \in M$, there exists an $R$-map $g\colon M \to C$ with $g(m) \neq 0$.

Exercise B-4.57 on page 501 in Part 1 can be restated to say that $\mathbb{Q}/\mathbb{Z}$ is an injective cogenerator of **Ab**.

**Lemma C-4.71.** *There exists an injective cogenerator of* ${}_R\mathbf{Mod}$.

**Proof.** Define $C$ to be an injective left $R$-module containing $\bigoplus_I R/I$, where $I$ varies over all the left ideals in $R$ (the module $C$ exists, by Theorem B-4.64 in Part 1). If $M$ is a left $R$-module and $m \in M$ is nonzero, then $\langle m\rangle \cong R/J$ for some left ideal $J$. Consider the diagram




where $i$ is the inclusion and $f$ is an isomorphism of $\langle m\rangle$ to some submodule of $C$ isomorphic to $R/J$. Since $C$ is injective, there is an $R$-map $g\colon M \to C$ extending $f$, and so $g(m) \neq 0$. •

An analysis of the proof of Proposition B-7.4 in Part 1 shows that it can be generalized by replacing $\mathrm{Hom}(A, \ \ )$ by any left exact functor that preserves direct products. However, this added generality is only illusory, in light of the following theorem of Watts characterizing representable functors on module categories.

**Theorem C-4.72 (Watts).** *If* $G\colon {}_R\mathbf{Mod} \to \mathbf{Ab}$ *is a covariant additive functor preserving inverse limits, then* $G$ *is naturally isomorphic to* $\mathrm{Hom}_R(B, \ \ )$ *for some left* $R$*-module* $B$.

**Proof.** For a module $M$ and a set $X$, let $M^X$ denote the direct product of copies of $M$ indexed by $X$; more precisely, $M^X$ is the set of all functions $X \to M$. In particular, $1_M \in M^M$, and we write $e = 1_M \in M^M$. If $m \in M$ and $\pi_m \colon M^M \to M$ is the $m$th projection, then the $m$th coordinate of $e$ is $\pi_m(e) = m$.

Choose an injective cogenerator $C$ of ${}_R\mathbf{Mod}$. Let $\Pi = C^{GC}$, and let its projection maps be $p_x \colon \Pi \to C$ for all $x \in GC$. Since $G$ preserves inverse limits, it preserves direct products, and so $G\Pi$ is a direct product with projection maps $Gp_x$. More precisely, if $\pi_x \colon (GC)^{GC} \to GC$ are the projection maps, then there is a unique isomorphism $\theta$ making the following diagrams commute for all $x \in GC$:

$$\begin{array}{ccc} G\Pi & \xleftarrow{\ \theta\ } & (GC)^{GC} \\ & {\scriptstyle Gp_x}\searrow \quad \swarrow{\scriptstyle \pi_x} & \\ & GC. & \end{array}$$

Thus, $(Gp_x)\theta = \pi_x$ for all $x \in GC$. Write

$$e = 1_{GC} \in (GC)^{GC}.$$

Define $\tau \colon \operatorname{Hom}_R(\Pi, C) \to GC$ by

$$\tau \colon f \mapsto (Gf)(\theta e).$$

If $f \colon \Pi \to C$, then $Gf \colon G\Pi \to GC$; since $\theta e \in G\Pi$, $\tau(f) = (Gf)(\theta e)$ makes sense.

The map $\tau$ is surjective, for if $x \in GC$, then $\tau(p_x) = (Gp_x)(\theta e) = \pi_x(e) = x$. We now describe $\ker \tau$. If $S \subseteq \Pi$, denote the inclusion $S \to \Pi$ by $i_S$. Define

$$B = \bigcap_{S \in \mathcal{S}} S, \quad \text{where } \mathcal{S} = \{\text{submodules } S \subseteq \Pi : \theta e \in \operatorname{im} G(i_S)\}.$$

We show that $\mathcal{S}$ is closed under finite intersections. All the maps in the first diagram below are inclusions, so that $i_S\lambda = i_{S\cap T}$. Since $G$ preserves inverse limits, it preserves pullbacks; since the first diagram is a pullback, the second diagram is also a pullback:

$$\begin{array}{ccc} S \cap T & \xrightarrow{\lambda} & S \\ {\scriptstyle \mu}\downarrow & & \downarrow{\scriptstyle i_S} \\ T & \xrightarrow[i_T]{} & \Pi \end{array} \qquad \text{and} \qquad \begin{array}{ccc} G(S \cap T) & \xrightarrow{G\lambda} & GS \\ {\scriptstyle G\mu}\downarrow & & \downarrow{\scriptstyle G(i_S)} \\ GT & \xrightarrow[G(i_T)]{} & G\Pi. \end{array}$$

By the definition of $\mathcal{S}$, there are $u \in GS$ with $(Gi_S)u = \theta e$ and $v \in GT$ with $(Gi_T)v = \theta e$. By Exercise C-3.42 on page 292, there is $d \in G(S \cap T)$ with $(Gi_S)(G\lambda)d = \theta e$. But $(Gi_S)(G\lambda) = Gi_{S\cap T}$, so that $\theta e \in \operatorname{im} G(i_{S\cap T})$ and $S \cap T \in \mathcal{S}$. It now follows from Example B-7.3(iii) on page 654 in Part 1 that $B = \bigcap S \cong \varprojlim S$, so that $B \in \mathcal{S}$.

Now $G$ is left exact, so that exactness of $0 \to \ker f \xrightarrow{\nu} \Pi \xrightarrow{f} C$ gives exactness of $0 \to G(\ker f) \xrightarrow{G\nu} G\Pi \xrightarrow{Gf} GC$. Thus, $\operatorname{im} G\nu = \ker(Gf)$. If

$$j \colon B \to \Pi$$

is the inclusion, then $\ker \tau = \ker j^*$, where $j^* \colon f \mapsto fj$ is the induced map $j^* \colon \operatorname{Hom}_R(\Pi, C) \to \operatorname{Hom}_R(B, C)$: if $f \in \ker \tau$, then $(Gf)\theta e = 0$, and $\theta e \in \ker Gf = \operatorname{im} G\nu$; thus, $\ker f \in \mathcal{S}$. Hence, $B \subseteq \ker f$, $fj = 0$, $f \in \ker j^*$, and $\ker \tau \subseteq \ker j^*$.

For the reverse inclusion, assume that $f \in \ker j^*$, so that $B \subseteq \ker f$. Then $\operatorname{im} Gj \subseteq \operatorname{im} G\nu = \ker Gf$. But $\theta e \in \ker Gf$; that is, $(Gf)\theta e = 0$, and $f \in \ker \tau$. Therefore, $\ker j^* = \ker \tau$.

In the diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \operatorname{Hom}_R(\Pi/B, C) & \longrightarrow & \operatorname{Hom}_R(\Pi, C) & \xrightarrow{j^*} & \operatorname{Hom}_R(B, C) & \longrightarrow & 0 \\
 & & \Big\| & & \Big\| & & \Big\downarrow{\scriptstyle \sigma_C} & & \\
0 & \longrightarrow & \operatorname{Hom}_R(\Pi/B, C) & \longrightarrow & \operatorname{Hom}_R(\Pi, C) & \xrightarrow[\tau]{} & GC & \longrightarrow & 0,
\end{array}$$

the first two vertical arrows are identities, so that the diagram commutes. Exactness of $0 \to B \xrightarrow{j} \Pi \to \Pi/B \to 0$ and injectivity of $C$ give exactness of the top row, while the bottom row is exact because $\tau$ is surjective and $\ker \tau = \ker j^*$. It follows that the two cokernels are isomorphic: there is an isomorphism

$$\sigma_C \colon \operatorname{Hom}_R(B, C) \to GC,$$

given by $\sigma_C \colon f \mapsto (Gf)\theta e$ (for the fussy reader, this is Proposition B-1.46 in Part 1).

For any module $M$, there is a map $M \to C^{\operatorname{Hom}_R(M,C)}$ given by $m \mapsto (fm)$, that "vector" whose $f$th coordinate is $fm$; this map is an injection because $C$ is a cogenerator. Similarly, if $N = \operatorname{coker}(M \to C^{\operatorname{Hom}_R(M,C)})$, there is an injection $N \to C^Y$ for some set $Y$; splicing these together gives an exact sequence

$$\begin{array}{ccccccc}
0 & \longrightarrow & M & \longrightarrow & C^{\operatorname{Hom}_R(M,C)} & \twoheadrightarrow & C^Y \\
 & & & & & \searrow & \uparrow \\
 & & & & & & N.
\end{array}$$

Since both $G$ and $\operatorname{Hom}_R(B, \quad)$ are left exact, there is a commutative diagram with exact rows

$$\begin{array}{ccccccc}
0 & \longrightarrow & \operatorname{Hom}_R(B, M) & \longrightarrow & \operatorname{Hom}_R(B, C^{\operatorname{Hom}_R(M,C)}) & \longrightarrow & \operatorname{Hom}_R(B, C^Y) \\
 & & \Big\downarrow{\scriptstyle \sigma_M} & & \Big\downarrow{\scriptstyle \sigma_{C^{\operatorname{Hom}_R(M,C)}}} & & \Big\downarrow{\scriptstyle \sigma_{C^Y}} \\
0 & \longrightarrow & GM & \longrightarrow & GC^{\operatorname{Hom}_R(M,C)} & \longrightarrow & GC^Y.
\end{array}$$

The vertical maps $\sigma_{C^{\operatorname{Hom}_R(M,C)}}$ and $\sigma_{C^Y}$ are isomorphisms, so that Proposition B-1.47 in Part 1 gives a unique isomorphism $\sigma_M \colon \operatorname{Hom}_R(B, M) \to GM$. It remains to prove that the isomorphisms $\sigma_M$ constitute a natural transformation. Recall, for any set $X$, that $\operatorname{Hom}_R(B, C^X) \cong \operatorname{Hom}_R(B, C)^X$ via $f \mapsto (p_x f)$, where $p_x$ is the $x$th projection. The map $\sigma_{C^X} \colon \operatorname{Hom}_R(B, C^X) \to GC^X$ is given by $f \mapsto \big((Gp_x f)\theta e\big) = \big((Gp_x f)\big)\theta e = (Gf)\theta e$. Therefore, $\sigma_M \colon \operatorname{Hom}_R(B, M) \to GM$ is given by $f \mapsto (Gf)\theta e$, and Theorem C-4.8, Yoneda's Lemma, shows that $\sigma$ is a natural isomorphism. •

**Remark.** No easy description of the module $B$ is known. However, we know that $B$ is not $G(R)$. For example, if $G = \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Q}, \quad)$, then Watts's Theorem applies to give $\operatorname{Hom}_{\mathbb{Z}}(B, \quad) \cong \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Q}, \quad)$. Now Theorem C-4.10 says that $B \cong \mathbb{Q}$, but $B \ncong G(\mathbb{Z}) = \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Q}, \mathbb{Z}) = \{0\}$. ◄

**Theorem C-4.73 (Adjoint Functor Theorem for Modules).** *If* $F\colon \mathbf{Mod}_R \to \mathbf{Ab}$ *is an additive functor, then the following statements are equivalent:*

(i) $F$ *preserves direct limits.*

(ii) $F$ *is right exact and preserves direct sums.*

(iii) $F \cong - \otimes_R B$ *for some left* $R$*-module* $B$*.*

(iv) $F$ *has a right adjoint; there is a functor* $G\colon \mathbf{Ab} \to \mathbf{Mod}_R$ *so that* $(F, G)$ *is an adjoint pair.*

**Proof.**

(i) $\Rightarrow$ (ii). Cokernels and direct sums are direct limits.

(ii) $\Rightarrow$ (iii). Theorem C-4.67.

(iii) $\Rightarrow$ (iv). Take $G = \mathrm{Hom}_R(B, \quad)$ in the Adjoint Isomorphism Theorem.

(iv) $\Rightarrow$ (i). Theorem C-4.65. •

**Theorem C-4.74.** *If* $G\colon {}_R\mathbf{Mod} \to \mathbf{Ab}$ *is an additive functor, then the following statements are equivalent:*

(i) $G$ *preserves inverse limits.*

(ii) $G$ *is left exact and preserves direct products.*

(iii) $G$ *is representable; i.e.,* $G \cong \mathrm{Hom}_R(B, \quad)$ *for some left* $R$*-module* $B$*.*

(iv) $G$ *has a left adjoint; there is a functor* $F\colon \mathbf{Ab} \to {}_R\mathbf{Mod}$ *so that* $(F, G)$ *is an adjoint pair.*

**Proof.**

(i) $\Rightarrow$ (ii). Kernels and direct products are inverse limits.

(ii) $\Rightarrow$ (iii). Theorem C-4.72.

(iii) $\Rightarrow$ (iv). Take $F = - \otimes_R B$ in the Adjoint Isomorphism Theorem.

(iv) $\Rightarrow$ (i). A left exact additive functor that preserves products must preserve inverse limits (Exercise B-7.10 on page 671 in Part 1). •

The general ***Adjoint Functor Theorem*** says that a functor $G$ on an arbitrary category has a left adjoint (that is, there exists a functor $F$ so that $(F, G)$ is an adjoint pair) if and only if $G$ preserves inverse limits and $G$ satisfies a "solution set condition" (Mac Lane [**144**], pp. 116–127 and 230). One consequence is a proof of the existence of free objects when a forgetful functor has a left adjoint; see Barr [**16**]. The Adjoint Functor Theorem also says that $F$ has a right adjoint if and only if $F$ preserves all direct limits and satisfies a solution set condition. Theorems C-4.73 and C-4.74 are special cases of the Adjoint Functor Theorem.

It can be proved that adjoints are unique if they exist: if $(F, G)$ and $(F, G')$ are adjoint pairs, where $F\colon \mathcal{A} \to \mathcal{B}$ and $G, G'\colon \mathcal{B} \to \mathcal{A}$, then $G \cong G'$; similarly, if $(F, G)$ and $(F', G)$ are adjoint pairs, then $F \cong F'$ (Mac Lane [**144**], p. 83). Here is the special case for module categories.

**Proposition C-4.75.** *Let* $F\colon {}_R\mathbf{Mod} \to \mathbf{Ab}$ *and* $G, G'\colon \mathbf{Ab} \to {}_R\mathbf{Mod}$ *be functors. If* $(F, G)$ *and* $(F, G')$ *are adjoint pairs, then* $G \cong G'$.

**Proof.** For every left $R$-module $C$, there are natural isomorphisms

$$\mathrm{Hom}_R(C, G\quad) \cong \mathrm{Hom}_{\mathbb{Z}}(FC,\quad) \cong \mathrm{Hom}_R(C, G'\quad).$$

Thus, $\mathrm{Hom}_R(C,\quad)\circ G \cong \mathrm{Hom}_R(C,\quad)\circ G'$ for every left $R$-module $C$. In particular, if $C = R$, then $\mathrm{Hom}_R(R,\quad) \cong 1$, the identity functor on ${}_R\mathbf{Mod}$, and so $G \cong G'$. •

**Remark.** In functional analysis, one works with topological vector spaces; moreover, there are many different topologies imposed on vector spaces, depending on the sort of problem being considered. We know that if $A, B, C$ are modules, then the Adjoint Isomorphism Theorem, Theorem B-4.98 in Part 1, gives a natural isomorphism

$$\mathrm{Hom}(A \otimes B, C) \cong \mathrm{Hom}(A, \mathrm{Hom}(B, C)).$$

Thus, $- \otimes B$ is the left adjoint of $\mathrm{Hom}(B,\quad)$. In the category of topological vector spaces, Grothendieck defined ***topological tensor products*** as left adjoints of $\mathrm{Hom}(B,\quad)$. Since the Hom sets consist of *continuous* linear transformations, they depend on the topology, and so topological tensor products also depend on the topology. ◀

## Exercises

**C-4.46.** Give an example of an additive functor $H\colon \mathbf{Ab} \to \mathbf{Ab}$ that has neither a left nor a right adjoint.

∗ **C-4.47.** Let $(F, G)$ be an adjoint pair, where $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$, and let the natural bijection be $\tau_{C,D}\colon \mathrm{Hom}(FC, D) \to \mathrm{Hom}(C, GC)$.

(i) If $D = FC$, there is a natural bijection

$$\tau_{C,FC}\colon \mathrm{Hom}(FC, FC) \to \mathrm{Hom}(C, GFC)$$

with $\tau(1_{FC}) = \eta_C\colon C \to GFC$. Prove that $\eta\colon 1_{\mathcal{C}} \to GF$ is a natural transformation, where $\eta$ is the unit of the adjoint pair.

(ii) If $C = GD$, there is a natural bijection

$$\tau^{-1}_{GD,D}\colon \mathrm{Hom}(GD, GD) \to \mathrm{Hom}(FGD, D)$$

with $\tau^{-1}(1_D) = \varepsilon_D\colon FGD \to D$. Prove that $\varepsilon\colon FG \to 1_{\mathcal{D}}$ is a natural transformation. (We call $\varepsilon$ the ***counit*** of the adjoint pair.)

**C-4.48.** Let $(F, G)$ be an adjoint pair of functors between module categories. Prove that if $G$ is exact, then $F$ preserves projectives; that is, if $P$ is a projective module, then $FP$ is projective. Dually, prove that if $F$ is exact, then $G$ preserves injectives.

**C-4.49.** (i) Let $F\colon \mathbf{Groups} \to \mathbf{Ab}$ be the functor with $F(G) = G/G'$, where $G'$ is the commutator subgroup of a group $G$, and let $U\colon \mathbf{Ab} \to \mathbf{Groups}$ be the functor taking every abelian group $A$ into itself (that is, $UA$ regards $A$ as a not necessarily abelian group). Prove that $(F, U)$ is an adjoint pair of functors.

(ii) Prove that the unit of the adjoint pair $(F, U)$ is the natural map $G \to G/G'$.

∗ **C-4.50.** (i) If $I$ is a partially ordered set, let $\mathbf{Dir}({}_R\mathbf{Mod}/I)$ denote the class of all direct systems of left $R$-modules over $I$, together with their morphisms.

(ii) Prove that $\mathbf{Dir}({}_R\mathbf{Mod}/I)$ is a category and that $\varinjlim\colon \mathbf{Dir}({}_R\mathbf{Mod}/I) \to {}_R\mathbf{Mod}$ is a functor.

(iii) Let $\mathcal{C}, \mathcal{D}$ be categories. Prove that the constant functors $\mathcal{C} \to \mathcal{D}$ (see Example B-4.15 in Part 1) define a functor $|\ \ |\colon \mathcal{C} \to \mathcal{C}^{\mathcal{D}}$: to each object $C$ in $\mathcal{C}$ assign the constant functor $|C|$, and to each morphism $\varphi\colon C \to C'$ in $\mathcal{C}$, assign the natural transformation $|\varphi|\colon |C| \to |C'|$ defined by $|\varphi|_D = \varphi$.

(iv) If $\mathcal{C}$ is cocomplete, prove that $(\varinjlim, |\ \ |)$ is an adjoint pair, and conclude that $\varinjlim$ preserves direct limits.

(v) Let $I$ be a partially ordered set and let $\mathbf{Inv}({}_R\mathbf{Mod}/I)$ denote the class of all inverse systems, together with their morphisms, of left $R$-modules over $I$. Prove that $\mathbf{Inv}({}_R\mathbf{Mod}/I)$ is a category and that $\varprojlim\colon \mathbf{Inv}({}_R\mathbf{Mod}/I) \to {}_R\mathbf{Mod}$ is a functor.

(vi) Prove that if $\mathcal{C}$ is complete, then $(|\ \ |, \varprojlim)$ is an adjoint pair and $\varprojlim$ preserves inverse limits.

**C-4.51.** (i) If $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ is an ascending sequence of submodules of a module $A$, prove that $A/\bigcup A_i \cong \bigcup A/A_i$; that is, $\operatorname{coker}(\varinjlim A_i \subseteq A) \cong \varinjlim \operatorname{coker}(A_i \to A)$.

(ii) Generalize (i): prove that any two direct limits (perhaps with distinct index sets) commute.

(iii) Prove that any two inverse limits (perhaps with distinct index sets) commute.

(iv) Give an example in which direct limit and inverse limit do not commute.

**C-4.52.** (i) Define ACC in ${}_R\mathbf{Mod}$, and prove that if ${}_S\mathbf{Mod} \cong {}_R\mathbf{Mod}$, then ${}_S\mathbf{Mod}$ has ACC. Conclude that if $R$ is left noetherian, then $S$ is left noetherian.

(ii) Give an example showing that ${}_R\mathbf{Mod}$ and $\mathbf{Mod}_R$ are not isomorphic.

**C-4.53.** (i) A ***generator*** of a category $\mathcal{C}$ is an object $G$ such that $\operatorname{Hom}(G, \ \ )\colon \mathcal{C} \to \mathbf{Sets}$ is a faithful functor; that is, if $f, g\colon A \to B$ are distinct morphisms in $\mathcal{C}$, then there exists a morphism $h\colon G \to A$ with $fh \neq gh$. Prove, when $\mathcal{C} = {}_R\mathbf{Mod}$, that this definition coincides with the definition of generator on page 385.

(ii) Recall that a ***cogenerator*** of a category $\mathcal{C}$ is an object $C$ such that $\operatorname{Hom}(\ , C)\colon \mathcal{C} \to \mathbf{Sets}$ is a faithful functor; that is, if $f, g\colon A \to B$ are distinct morphisms in $\mathcal{C}$, then there exists a morphism $h\colon B \to C$ with $hf \neq hg$. Prove, when $\mathcal{C} = {}_R\mathbf{Mod}$, that this definition coincides with the definition of cogenerator on page 398.

## C-4.8. Algebraic $K$-Theory

This section may be regarded as an introduction to algebraic $K$-theory. In contrast to homological algebra, this topic is often called *homotopical algebra*, for there are topological interpretations involving homotopy groups as well as homology groups. There are two different constructions of *Grothendieck groups*: $K_0$ and $G_0$. Although we will define $K_0(C)$ and $G_0(C)$ for various types of $C$ (semigroups, categories, rings), the most important versions have $C$ a ring. We shall see that $K_0(R) \cong G_0(R)$ for every ring $R$.

Exercise C-4.54 on page 417 shows that $K_0(R) \cong K_0(R^{\mathrm{op}})$ for every ring $R$, so that "$R$-module" in this section will always mean "left $R$-module".

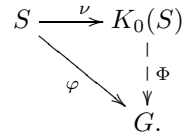
## ▪ The Functor $K_0$

We begin by generalizing the construction of the additive group of the integers $\mathbb{Z}$ from the additive semigroup of the natural numbers $\mathbb{N}$.

**Definition.** If $S$ is a commutative semigroup, then $K_0(S)$ is an abelian group solving the following universal problem in the category

**CS**

of commutative semigroups: there is a semigroup map $\nu\colon S \to K_0(S)$ such that, for every abelian group $G$ and every semigroup map $\varphi\colon S \to G$, there exists a unique group homomorphism $\Phi\colon K_0(S) \to G$ with $\Phi\nu = \varphi$,




We call $K_0(S)$ the ***Grothendieck group*** of the semigroup $S$.

In generalizing the Riemann–Roch Theorem, Grothendieck recognized (see Proposition C-4.90 below) that if $C$ is the category of all finitely generated $R$-modules and a filtration of a module

$$M = M_0 \supseteq M_1 \supseteq M_2 \supseteq \cdots \supseteq M_n = \{0\}$$

has factor modules $Q_i = M_{i-1}/M_i$, then $M$ is a telescoping sum

$$[M] = [Q_1] + \cdots + [Q_n] \quad \textit{in } G_0(\mathcal{C})$$

in the Grothendieck group.

**Proposition C-4.76.** *For every commutative semigroup $S$, the group $K_0(S)$ exists.*

**Proof.** Let $F(S)$ be the free abelian group with basis $S$, and let

$$H = \big\langle x + y - z\colon x + y = z \text{ in } S \big\rangle.$$

Define $K_0(S) = F(S)/H$, and define $\nu\colon S \to K_0(S)$ by $\nu\colon x \mapsto x + H$. A semigroup map $\varphi\colon S \to G$, being a function defined on a basis, has a unique extension to a group homomorphism $\widetilde{\varphi}\colon F(S) \to G$; note that $H \subseteq \ker \widetilde{\varphi}$, because $\varphi(x + y) = \varphi(x) + \varphi(y)$. Define $\Phi\colon K_0(S) \to G$ by $x + H \mapsto \widetilde{\varphi}(x)$. •

If $x \in S$, we will denote

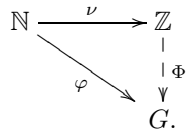
$$\nu(x) = x + H \text{ by } [x].$$

As always, solutions to a universal mapping problem are unique to isomorphism.

**Corollary C-4.77.** *If $\mathbb{N}$ is viewed as an additive commutative semigroup, then*

$$K_0(\mathbb{N}) = \mathbb{Z}.$$

**Proof.** In the diagram below, let $\nu\colon \mathbb{N} \to \mathbb{Z}$ be the inclusion. Define $\Phi(n) = \varphi(n)$ if $n \geq 0$, and define $\Phi(n) = -\varphi(-n)$ if $n < 0$:




•

**Example C-4.78.** The last corollary is not really a construction of $\mathbb{Z}$ from $\mathbb{N}$, because the construction of $K_0(\mathbb{N})$ in Proposition C-4.76 uses free abelian groups. We now sketch a second construction of $K_0(S)$, where $S$ is a commutative semigroup, which remedies this.

Define a relation on $S \times S$ by

$$(x, y) \sim (a, b) \text{ if there exists } s \in S \text{ with } x + b + s = a + y + s.$$

It is easily verified that this is an equivalence relation (if $R$ is a domain, then $R - \{0\}$ is a multiplicative semigroup, and this relation is essentially the cross product relation arising in the usual construction of $\mathrm{Frac}(R)$). Denote the equivalence class of $(x, y)$ by $[x, y]$, and note that $[x, x] = [y, y]$ for all $x, y \in S$. Define

$$[x, y] + [x', y'] = [x + x', y + y'].$$

A routine check shows that the orbit space $(S \times S)/\!\sim$ is an abelian group: addition is well-defined and associative; the identity element is $[x, x]$; the inverse of $[x, y]$ is $[y, x]$. Finally, define $\nu\colon S \to (S \times S)/\!\sim$ by $x \mapsto [x + x, x]$. The group $K_0(S) = (S \times S)/\!\sim$ is a solution of the universal problem.

For example, if $S = \mathbb{N}$, then $(2, 5) \sim (5, 8)$ because $2 + 8 + 0 = 5 + 5 + 0$. We usually write $-3 = [2, 5]$; that is, we view $[x, y]$ as "$x - y$" (this explains the definition of $\nu(x) = [x + x, x]$). The definition of $\sim$ for $\mathbb{N}$ is simpler than that for an arbitrary commutative semigroup because $\mathbb{N}$ is a monoid which satisfies the cancellation law ($u + s = v + s$ implies $u = v$). ◀

**Proposition C-4.79.** *The Grothendieck group defines a functor $K_0\colon \mathbf{CS} \to \mathbf{Ab}$, where* **CS** *is the category of all commutative semigroups.*

**Proof.** If $f\colon S \to T$ is a semigroup map, define $K_0(f)\colon K_0(S) \to K_0(T)$ by $[s] \mapsto [f(s)]$. The proof that $K_0$ is a functor is straightforward. •

Semigroups arising from categories provide interesting examples of Grothendieck groups. Since most categories are not small, we first deal with a set-theoretic problem this might cause. If $\mathcal{C}$ is a category, let

$$\mathrm{Iso}(\mathcal{C})$$

denote the family of all isomorphism classes of objects in $\mathcal{C}$. We will abuse notation by allowing an object $A$ to denote its own isomorphism class.

**Definition.** A category $\mathcal{C}$ is ***virtually small*** if its family $\mathrm{Iso}(\mathcal{C})$ of isomorphism classes is a *set*; that is, $\mathrm{Iso}(\mathcal{C})$ has a cardinal number.

**Example C-4.80.**

(i) The category $\mathcal{C}$ of all countable abelian groups is not a small category (indeed, its (full) subcategory of all groups of order 1 is not small). However, $\mathcal{C}$ is virtually small.

(ii) Any *full* subcategory $\mathcal{U}$ of a virtually small category $\mathcal{C}$ is virtually small. (We must assume that $\mathcal{U}$ is full; otherwise, objects in $\mathcal{U}$ which are isomorphic in $\mathcal{C}$ might not be isomorphic in $\mathcal{U}$. For example, the discrete subcategory (having the same objects but whose only morphisms are identities) of the category of all countable abelian groups is not virtually small.)

(iii) For any ring $R$, the full subcategory of ${}_R\mathbf{Mod}$ consisting of all finitely generated left $R$-modules is virtually small. Using part (ii), we see that the full subcategory

$$R\text{-}\mathbf{Proj}$$

generated by all finitely generated projective left $R$-modules is virtually small. ◀

**Lemma C-4.81.** *Let $\mathcal{C}$ be a full additive subcategory of ${}_R\mathbf{Mod}$ for some ring $R$. If $\mathcal{C}$ is virtually small, then the family* $\mathrm{Iso}(\mathcal{C})$ *of all its isomorphism classes is a semigroup in which*

$$A + A' = A \oplus A', \text{ where } A, A' \in \mathrm{Iso}(\mathcal{C}).$$

**Proof.** First, $\mathrm{Iso}(\mathcal{C})$ is a set, for $\mathcal{C}$ is virtually small. Since $\mathcal{C}$ is additive, $A, A' \in \mathrm{Iso}(\mathcal{C})$ implies $A \oplus A' \in \mathrm{Iso}(\mathcal{C})$ (we are using our notational convention of allowing an object to denote its own isomorphism class). Thus, direct sum gives a well-defined binary operation on $\mathrm{Iso}(\mathcal{C})$. This operation is associative, for $A \oplus (A' \oplus A'') \cong (A \oplus A') \oplus A''$ gives $A \oplus (A' \oplus A'') = (A \oplus A') \oplus A''$ in $\mathrm{Iso}(\mathcal{C})$. •

**Definition.** If $R$ is a ring, the ***Grothendieck group*** $K_0(R)$ is defined by

$$K_0(R) = K_0(\mathrm{Iso}(R\text{-}\mathbf{Proj})),$$

where $R$-**Proj** is the full subcategory of all finitely generated projective left $R$-modules.

We have defined $K_0(\mathbb{Z})$ in two different ways. On the one hand, it was defined by viewing $\mathbb{Z}$ as a semigroup (and $K_0(\mathbb{Z}) \cong \mathbb{Z}$). On the other hand, we have just defined $K_0(\mathbb{Z})$ as $K_0(\mathrm{Iso}(\mathbb{Z}\text{-}\mathbf{Proj}))$. But $\mathbb{Z}$-**Proj** is just the semigroup of isomorphism classes of finitely generated projective $\mathbb{Z}$-modules (that is, of finitely generated free abelian groups). We shall see, in Proposition C-4.83 below, that this second $K_0(\mathbb{Z})$ is also isomorphic to $\mathbb{Z}$.

**Remark.**

(i) It is not interesting to remove the finitely generated hypothesis in the definition of $K_0(R)$. If $\mathcal{C}$ is closed under countable direct sums, then $A \in \mathrm{obj}(\mathcal{C})$ implies $A' \in \mathrm{obj}(\mathcal{C})$, where $A'$ is the direct sum of countably many copies of $A$. But the isomorphism $A \oplus A' \cong A'$ gives the equation $[A] + [A'] = [A']$ in the group $K_0$, which forces $[A] = 0$. Thus, this Grothendieck group is trivial.

(ii) If $R$ is commutative, then tensor product makes $K_0(R)$ into a commutative ring. If $A, B \in R\text{-}\mathbf{Proj}$, then $A \otimes_R B \in R\text{-}\mathbf{Proj}$ (Exercise B-4.81 on page 520 in Part 1). Moreover, if $A \cong A'$ and $B \cong B'$, then $A \otimes_R B \cong A' \otimes_R B'$, so that $(A, B) \mapsto A \otimes_R B$ gives a well-defined multiplication on $K_0(R)$. The reader may check that $K_0(R)$ is a commutative ring with unit $[R]$.

(iii) Forming a Grothendieck group on $R\text{-}\mathbf{Proj}$ when $R$ is commutative, with binary operation tensor product instead of direct sum, leads to the notion of *Picard group* in algebraic geometry. ◀

We have seen that the Grothendieck group defines a functor on the category of commutative semigroups; we now show that it also defines a functor on **Rings**, the category of rings.

**Proposition C-4.82.** *There is a functor* $K_0\colon \mathbf{Rings} \to \mathbf{Ab}$ *with* $R \mapsto K_0(R)$. *Moreover, if* $\varphi\colon R \to S$ *is a ring homomorphism, then* $K_0(\varphi)$ *takes free left* $R$*-modules to free left* $S$*-modules*; *that is,* $K_0(\varphi)\colon [R^n] \mapsto [S^n]$ *for all* $n \geq 0$.

**Proof.** If $\varphi\colon R \to S$ is a ring homomorphism, then we may view $S$ as an $(S, R)$-bimodule by defining $sr = s\varphi(r)$ for all $r \in R$ and $s \in S$. For any left $R$-module $P$, the tensor product $S \otimes_R P$ now makes sense, and it is a left $S$-module, by Corollary B-4.83 in Part 1. The restriction of the functor $S \otimes_R -\colon {}_R\mathbf{Mod} \to {}_S\mathbf{Mod}$ takes $R\text{-}\mathbf{Proj} \to S\text{-}\mathbf{Proj}$, for if $P$ is a finitely generated projective $R$-module, then $S \otimes_R P$ is a finitely generated projective $S$-module (Exercise B-4.81 on page 520 in Part 1). Now define $K_0(\varphi)\colon K_0(R) \to K_0(S)$ by $[P] \mapsto [S \otimes_R P]$. In particular, if $P = R^n$ is a finitely generated free left $R$-module, then $K_0(\varphi)\colon [R^n] \mapsto [S \otimes_R R^n] = [S^n]$. •

For every ring $R$, there is a ring homomorphism $\iota\colon \mathbb{Z} \to R$ which takes the integer 1 to the unit element of $R$. Since $K_0$ is a functor on the category of rings, there is a homomorphism $K_0(\iota)\colon K_0(\mathbb{Z}) \to K_0(R)$. By Proposition C-4.82, $\operatorname{im} K_0(\iota)$ is the subgroup of $K_0(R)$ generated by all finitely generated free $R$-modules. If $V$ is an infinite-dimensional vector space over a field $k$, then $R = \operatorname{End}_k(V)$ has the property that $R \cong R \oplus R$ as $R$-modules (Rotman [**187**], p. 58). It follows that $\operatorname{im} K_0(\iota) = \{0\}$ for this ring. In fact, $K_0(R) = \{0\}$ (Rosenberg [**183**], p. 10).

**Definition.** For any ring $R$, the ***reduced Grothendieck group*** is defined by

$$\widetilde{K}_0(R) = \operatorname{coker} K_0(\iota) = K_0(R)/\operatorname{im} K_0(\iota).$$

In the next chapter, we will show that $\widetilde{K}_0(R)$ is the class group of the ring of integers $R$ in an algebraic number field.

Since $\operatorname{im} K_0(\iota)$ is the subgroup of $K_0(R)$ generated by all finitely generated free $R$-modules, the reduced Grothendieck group is the important part of $K_0(R)$, for it describes nonfree projective $R$-modules. For example, if $R$ is a ring for which every finitely generated projective $R$-module is free, then $K_0(R) = \operatorname{im} K_0(\iota)$ and $\widetilde{K}_0(R) = \{0\}$. We shall see, on page 412, that the converse of this is false.

Let us compute $K_0(R)$ for some nice rings $R$.

**Proposition C-4.83.** *Let $R$ be a ring all of whose finitely generated projective left $R$-modules are free and which has a rank function; that is, $R^m \cong R^n$ implies $m = n$. Then $K_0(R) \cong \mathbb{Z}$.*[12] *In particular, $K_0(R) \cong \mathbb{Z}$ if $R$ is a division ring, a* PID*, or $k[x_1, \ldots, x_n]$ when $k$ is a field.*

**Proof.** The function $\rho\colon \mathrm{Iso}(R\text{-}\mathbf{Proj}) \to \mathbb{N}$ induced by rank, namely, $\rho(R^n) = n$, is well-defined; it is a semigroup map, for $\rho(V \oplus W) = \rho(V) + \rho(W)$, and it is an isomorphism because we are assuming that two free modules $P$ and $Q$ are isomorphic if and only if $\rho(P) = \rho(Q)$. Since $K_0\colon \mathbf{CS} \to \mathbf{Ab}$ is a functor, $K_0(\rho)$ is an isomorphism $K_0(R) \to K_0(\mathbb{N}) = \mathbb{Z}$.

Each of the stated rings satisfies the hypotheses: see Proposition B-1.40 in Part 1 for division rings, Theorem B-2.28 in Part 1 for PIDs, and the Quillen–Suslin Theorem (Rotman [**187**], p. 209) for polynomial rings in several variables over a field. •

## ▪ The Functor $G_0$

As we said earlier, we are now going to define a new Grothendieck group, $G_0(\mathcal{C})$, which coincides with $K_0(R)$ when $\mathcal{C} = R\text{-}\mathbf{Proj}$.

**Definition.** A ***$G$-category*** $\mathcal{C}$ is a virtually small full subcategory of ${}_R\mathbf{Mod}$ for some ring $R$.

**Example C-4.84.**

(i) $R$-**Proj** is a $G$-category.

(ii) All finitely generated $R$-modules form a $G$-category.

(iii) The category of all countable torsion-free abelian groups is a $G$-category. ◀

We continue to denote the family of all isomorphism classes of objects in $\mathcal{C}$ by $\mathrm{Iso}(\mathcal{C})$, and we continue letting an object $A$ denote its own isomorphism class.

**Definition.** If $\mathcal{C}$ is a $G$-category, let $\mathcal{F}(\mathcal{C})$ be the free abelian group with basis $\mathrm{Iso}(\mathcal{C})$,[13] and let $\mathcal{E}$ be the subgroup of $\mathcal{F}(\mathcal{C})$ generated by all elements of the form

$$A + C - B \quad \text{if} \quad 0 \to A \to B \to C \to 0 \text{ is exact.}$$

The ***Grothendieck group*** $G_0(\mathcal{C})$ is the abelian group

$$G_0(\mathcal{C}) = \mathcal{F}(\mathcal{C})/\mathcal{E}.$$

For any object $A$ in $\mathcal{C}$, we denote the coset $A + \mathcal{E}$ in $G_0(\mathcal{C})$ by

$$[A].$$

In particular, if $R$ is a ring, define

$$G_0(R) = G_0(R\text{-}\mathbf{Proj}).$$

[12] The ring $R = \mathrm{End}_k(V)$ mentioned on page 407 does not have a rank function.

[13] Since a basis of a free abelian group must be a set and not a proper class, we are using the hypothesis that $\mathcal{C}$ is virtually small.

**Proposition C-4.85.** *For every ring $R$, we have $G_0(R) \cong K_0(R)$.*

**Proof.** Recall Proposition C-4.76; if $S$ is a commutative semigroup, then $K_0(S) = F(S)/H$, where $H = \langle x + y - z \colon x + y = z \text{ in } S \rangle$. By definition, $K_0(R) = K_0(S)$, where $S$ is the semigroup $\mathrm{Iso}(R\text{-}\mathbf{Proj})$ with operation $A + B = A \oplus B$; hence, $K_0(R) = F(\mathrm{Iso}(R\text{-}\mathbf{Proj}))/H$, where $H$ is generated by all $A + B - A \oplus B$. Now $G_0(R) = F(R\text{-}\mathbf{Proj})/\mathcal{E}$, where $\mathcal{E}$ is generated by all $A + B - X$ with $0 \to A \to X \to B \to 0$ exact. But every exact sequence $0 \to A \to X \to B \to 0$ splits, because every object in $R$-**Proj** is projective, and so $X \cong A \oplus B$. •

When are two elements in $G_0(\mathcal{C})$ equal?

**Proposition C-4.86.** *Let $\mathcal{C}$ be a $G$-category.*

(i) *If $x \in G_0(\mathcal{C})$, then $x = [A] - [B]$ for $A, B \in \mathrm{obj}(\mathcal{C})$.*

(ii) *Let $A, B \in \mathrm{obj}(\mathcal{C})$. Then $[A] = [B]$ in $G_0(\mathcal{C})$ if and only if there are $C, U, V \in \mathrm{obj}(\mathcal{C})$ and exact sequences*

$$0 \to U \to A \oplus C \to V \to 0 \quad \textit{and} \quad 0 \to U \to B \oplus C \to V \to 0.$$

**Proof.**

(i) Since $G_0(\mathcal{C})$ is generated by $\mathrm{Iso}(\mathcal{C})$, each element $x \in G_0(\mathcal{C})$ is a $\mathbb{Z}$-linear combination of objects: $x = \sum_i m_i C_i$ (we allow $C_i$ to be repeated; that is, we assume each $m_i = \pm 1$). If we denote those $C$ with positive coefficient $m_i$ by $A_1, \dots, A_r$ and those with negative $m_i$ by $B_1, \dots, B_t$, then

$$x = \sum_{i=1}^{r} [A_i] - \sum_{j=1}^{s} [B_j].$$

Define $A = A_1 \oplus \cdots \oplus A_r$. That $[A] = [A_1 \oplus \cdots \oplus A_r]$ is proved by induction on $r \geq 2$. If $r = 2$, then exactness of $0 \to A_1 \to A_1 \oplus A_2 \to A_2 \to 0$ gives $[A_1] + [A_2] = [A_1 \oplus A_2]$; the inductive step is also easy. Similarly, if $B = B_1 \oplus \cdots \oplus B_s$, then $[B] = [B_1 \oplus \cdots \oplus B_s]$. Therefore, $x = [A] - [B]$.

(ii) If there exist modules $C, U$, and $V$ as in the statement, then

$$[A \oplus C] = [U] + [V] = [B \oplus C].$$

But exactness of $0 \to A \to A \oplus C \to C \to 0$ gives $[A \oplus C] = [A] + [C]$. Similarly, $[B \oplus C] = [B] + [C]$, so that $[A] + [C] = [B] + [C]$ and $[A] = [B]$.

Conversely, if $[A] = [B]$ in $G_0(\mathcal{C})$, then $A - B \in \mathcal{E}$, and there is an equation in $F(\mathcal{C})$ (as in part (i), we allow repetitions, so that all coefficients are $\pm 1$),

$$A - B = \sum_i (X_i' + X_i'' - X_i) - \sum_j (Y_j' + Y_j'' - Y_j),$$

where there are exact sequences $0 \to X_i' \to X_i \to X_i'' \to 0$ and $0 \to Y_j' \to Y_j \to Y_j'' \to 0$. Transposing to eliminate negative coefficients,

$$A + \sum_i (X_i' + X_i'') + \sum_j Y_j = B + \sum_i X_i + \sum_j (Y_j' + Y_j'').$$

This is an equation in a free abelian group, so that expressions as linear combinations of a basis are unique. Therefore, the set $\{A, X_i', X_i'', Y_j\}$ of objects on the left-hand side, with multiplicities, coincides with the set of objects $\{B, X_i, Y_i', Y_j''\}$ on the right-hand side, with multiplicities. Therefore, the direct sum of the objects on the left is isomorphic to the direct sum of the objects on the right:

$$A \oplus X' \oplus X'' \oplus Y \cong B \oplus X \oplus Y' \oplus Y'',$$

where $X' = \bigoplus_i X_i'$, $X = \bigoplus_i X_i$, and so forth. Let $C$ denote their common value:

$$A \oplus X' \oplus X'' \oplus Y = C = B \oplus X \oplus Y' \oplus Y''.$$

By Exercise B-1.48 on page 309 in Part 1, there are exact sequences

$$0 \to X' \oplus Y'' \to X \oplus Y'' \to X'' \to 0,$$

$$0 \to X' \oplus Y'' \to (X \oplus Y'') \oplus Y' \to X'' \oplus Y' \to 0,$$

and

$$0 \to X' \oplus Y'' \to A \oplus (X \oplus Y' \oplus Y'') \to A \oplus (X'' \oplus Y') \to 0.$$

The middle object is $C$. Applying Exercise B-1.48 in Part 1 once again, there is an exact sequence

$$0 \to X' \oplus Y' \to B \oplus C \to B \oplus (A \oplus X'' \oplus Y'') \to 0.$$

Define $U = X' \oplus Y'$ and $V = B \oplus A \oplus X'' \oplus Y''$; with this notation, the last exact sequence is

$$0 \to U \to B \oplus C \to V \to 0.$$

Similar manipulation yields an exact sequence $0 \to U \to A \oplus C \to V \to 0$. •

**Corollary C-4.87.** *Let $R$ be a ring.*

(i) *If $x \in G_0(R)$, then there are projective $R$-modules $P$ and $Q$ with*

$$x = [P] - [Q].$$

(ii) *If $A$ and $B$ are projective $R$-modules, then $[A] = [B]$ in $G_0(R)$ if and only if there is a projective $R$-module $C$ with*

$$A \oplus C \cong B \oplus C.$$

**Proof.** Recall that $G_0(R) = G_0(R\text{-}\mathbf{Proj})$, so that objects are projective $R$-modules. Thus, (i) follows at once from Proposition C-4.86. The second statement in the proposition says that there are projectives $C, U, V$ (for every object in $R$-**Proj** is projective) and exact sequences $0 \to U \to A \oplus C \to V \to 0$ and $0 \to U \to B \oplus C \to V \to 0$. These sequences split, because $V$ is projective; that is, $A \oplus C \cong U \oplus V$ and $B \oplus C \cong U \oplus V$. Hence, $A \oplus C \cong B \oplus C$. •

**Corollary C-4.88.** *Let $A$ and $B$ be projective $R$-modules. Then $[A] = [B]$ in $K_0(R)$ if and only if there is a projective $R$-module $P$ with*

$$A \oplus P \cong B \oplus P.$$

**Proof.** This follows from Corollary C-4.87 because $K_0(R) \cong G_0(R)$. •

**Definition.** Let $R$ be a commutative ring, and let $\mathcal{C}$ be a subcategory of ${}_R\mathbf{Mod}$. Two $R$-modules $A$ and $B$ are called ***stably isomorphic in*** $\mathcal{C}$ if there exists a module $C \in \mathrm{obj}(\mathcal{C})$ with $A \oplus C \cong B \oplus C$.

With this terminology, Corollary C-4.88 in Part 1 says that two modules determine the same element of $K_0(R)$ if and only if they are stably isomorphic. It is clear that isomorphic modules are stably isomorphic; the next example shows that the converse need not hold.

**Example C-4.89.**

(i) If $\mathcal{F}$ is the category of all finite abelian groups, then Exercise B-3.18 on page 377 in Part 1 shows that two finite abelian groups are stably isomorphic in $\mathcal{F}$ if and only if they are isomorphic.

(ii) Stably isomorphic finitely generated projective modules need not be isomorphic; here is an example of Kaplansky as described in [**137**]:

Let $A$ be the coordinate ring of the real 2-sphere; i.e., $A = \mathbb{R}[x, y, z]$ with the relation $x^2 + y^2 + z^2 = 1$. Let $\varepsilon\colon A^3 \to A$ be the $A$-linear functional given by $\varepsilon(\alpha, \beta, \gamma) = \alpha x + \beta y + \gamma z$. Since $\varepsilon(x, y, z) = 1$, we have a splitting $A^3 \cong \ker\varepsilon \oplus A$, so $P = \ker\varepsilon$ is stably free. Assume, for the moment, that $P$ is actually free, so (necessarily) $P \cong A^2$. Then, $A^3$ will have a new basis consisting of $(x, y, z)$ and two other triples $(f, g, h), (f', g', h')$. The matrix

$$\begin{bmatrix} x & f & f' \\ y & g & g' \\ z & h & h' \end{bmatrix}$$

is therefore invertible over $A$, and has determinant equal to a unit $e \in A$. We think of $f, g, h, e, e^{-1}, \cdots$ as functions on $S^2$ (they are polynomial expressions in the 'coordinate functions' $x, y, z$). Consider the continuous vector field on $S^2$ given by $v \in S^2 \mapsto (f(v), g(v), h(v)) \in \mathbb{R}^3$. Since $e = f' \cdot (yh - zg) - g' \cdot (xh - zf) + h' \cdot (xg - yf)$ is clearly nowhere zero on $S^2$, the vector $(f(v), g(v), h(v))$ is nowhere collinear with the vector $v$. Taking the orthogonal projections onto the tangent planes of $S^2$, we obtain a continuous vector field of nowhere vanishing tangents. This is well-known to be impossible by elementary topology. It follows that $P$ cannot be free over $A$.

Here is the heart of this construction.

**Theorem (Swan).** *Let $X$ be a compact Hausdorff space and let $R$ be the ring of continuous real-valued functions on $X$. If $E \xrightarrow{p} X$ is a vector bundle with global sections*

$$\Gamma(X) = \{\text{continuous } s\colon X \to E \mid ps = 1_X\},$$

*then $\Gamma(X)$ is a finitely generated projective $R$-module, and every such $R$-module arises in this way.*

See Swan [**218**], Bass [**17**], p. 735, or Rosenberg [**183**], pp. 32–36. ◀

Grothendieck proved that if $R$ is left noetherian, then $K_0(R) \cong K_0(R[x])$ (Rosenberg [**183**], p. 141). It follows, for $k$ a field, that $K_0(k[x_1, \dots, x_n]) \cong K_0(k) = \mathbb{Z}$, and so the reduced Grothendieck group $\widetilde{K}_0(R) = \{0\}$ (recall that $\widetilde{K}_0(R) = K_0(R)/\operatorname{im} K_0(\iota)$, where $\operatorname{im} K_0(\iota)$ is the subgroup generated by free $R$-modules). Serre wondered whether the converse holds for $R = k[x_1, \dots, x_n]$: are finitely generated projective $k[x_1, \dots, x_n]$-modules free? The Quillen–Suslin Theorem (Rotman [**187**], p. 209) proves that this is so.

A finitely generated projective $R$-module $P$ is ***stably free*** if it is stably isomorphic to a free module. Free modules are, obviously, stably free, but the coordinate ring of the 2-sphere in Example C-4.89(ii) shows there are stably free modules which are not free. A ring $R$ is called a ***Hermite ring*** if every finitely generated projective $R$-module is stably free. In light of Corollary C-4.88, a ring $R$ is Hermite if and only if $\widetilde{K}_0(R) = \{0\}$. If $R$ is a ring for which every finitely generated projective $R$-module is free, then $R$ is a Hermite ring. The converse is not true. Ojanguren–Sridharan [**170**] showed that if $\Delta$ is a noncommutative division ring, then $R = \Delta[x, y]$ (where coefficients commute with indeterminates) is a Hermite ring having projectives which are not free.

We now present computations of $G_0(\mathcal{C})$ for some familiar $G$-categories $\mathcal{C}$. The next result shows that filtrations of modules give rise to alternating sums in $G_0$.

**Proposition C-4.90.** *Let $R$ be a commutative ring and let $\mathcal{C}$ be the category of all finitely generated $R$-modules. If $M \in \operatorname{obj}(\mathcal{C})$ and*

$$M = M_0 \supseteq M_1 \supseteq M_2 \supseteq \cdots \supseteq M_n = \{0\}$$

*has factor modules $Q_i = M_{i-1}/M_i$, then*

$$[M] = [Q_1] + \cdots + [Q_n] \quad \text{in } G_0(\mathcal{C}).$$

**Proof.** Since $Q_i = M_{i-1}/M_i$, there is a short exact sequence

$$0 \to M_i \to M_{i-1} \to Q_i \to 0,$$

so that $[Q_i] = [M_{i-1}] - [M_i]$ in $G_0(\mathcal{C})$. We now have a telescoping sum:

$$\sum_{i=1}^{n} [Q_i] = \sum_{i=1}^{n} \big([M_{i-1}] - [M_i]\big) = [M_0] - [M_n] = [M]. \quad \bullet$$

**Definition.** Let $\mathcal{C}$ be a $G$-category. Call an object $S$ ***simple in*** $\mathcal{C}$ if, for every object $X$ and every monic $u\colon X \to S$ in $\mathcal{C}$, either $u = 0$ or $u$ is an isomorphism. A filtration

$$M = M_0 \supseteq M_1 \supseteq M_2 \supseteq \cdots \supseteq M_n = \{0\}$$

in a category $\mathcal{C}$ of modules is called a ***composition series*** of $M$ if each of its factors $Q_i = M_{i-1}/M_i$ is simple in $\mathcal{C}$.

A ***Jordan–Hölder category*** is a $G$-category $\mathcal{C}$ such that

(i) each object $M$ has a composition series;

(ii) for every two composition series

$$M = M_0 \supseteq M_1 \supseteq M_2 \supseteq \cdots \supseteq M_n = \{0\}$$

and

$$M = M'_0 \supseteq M'_1 \supseteq M'_2 \supseteq \cdots \supseteq M'_m = \{0\},$$

we have $m = n$ and a permutation $\sigma \in S_n$ such that $Q'_j \cong Q_{\sigma j}$ for all $j$, where $Q_i = M_{i-1}/M_i$ and $Q'_j = M'_{j-1}/M'_j$.

Define the ***length*** $\ell(M)$ of a module $M$ in a Jordan–Hölder category to be the number $n$ of terms in a composition series. If the simple factors of a composition series are $Q_1, \ldots, Q_n$, we define

$$\mathrm{jh}(M) = Q_1 \oplus \cdots \oplus Q_n.$$

A composition series may have several isomorphic factors, and $\mathrm{jh}(M)$ records their multiplicity. Since $Q_1 \oplus Q_2 \cong Q_2 \oplus Q_1$, $\mathrm{jh}(M)$ does not depend on the ordering $Q_1, \ldots, Q_n$.

**Lemma C-4.91.** *Let $\mathcal{C}$ be a Jordan–Hölder category, and let $Q_1, \ldots, Q_n$, $Q'_1, \ldots, Q'_m$ be simple objects in $\mathcal{C}$.*

(i) *If $Q_1 \oplus \cdots \oplus Q_n \cong Q'_1 \oplus \cdots \oplus Q'_m$, then $m = n$ and there is a permutation $\sigma \in S_n$ such that $Q'_j \cong Q_{\sigma j}$ for all $j$, where $Q_i = M_{i-1}/M_i$ and $Q'_j = M'_{j-1}/M'_j$.*

(ii) *If $M$ and $M'$ are modules in $\mathrm{obj}(\mathcal{C})$ and there is a simple object $S$ in $\mathcal{C}$ with*

$$S \oplus \mathrm{jh}(M) \cong S \oplus \mathrm{jh}(M'),$$

*then $\mathrm{jh}(M) \cong \mathrm{jh}(M')$.*

**Proof.**

(i) Since $Q_1, \ldots, Q_n$ are simple, the filtration

$$Q_1 \oplus \cdots \oplus Q_n \supseteq Q_2 \oplus \cdots \oplus Q_n \supseteq Q_3 \oplus \cdots \oplus Q_n \supseteq \cdots$$

is a composition series of $Q_1 \oplus \cdots \oplus Q_n$ with factors $Q_1, \ldots, Q_n$; similarly, $Q'_1 \oplus \cdots \oplus Q'_m$ has a composition series with factors $Q'_1, \ldots, Q'_m$. The result follows from $\mathcal{C}$ being a Jordan–Hölder category.

(ii) This result follows from part (i) because $S$ is simple. •

**Lemma C-4.92.** *If $0 \to A \to B \to C \to 0$ is an exact sequence in a Jordan–Hölder category, then*

$$\mathrm{jh}(B) \cong \mathrm{jh}(A) \oplus \mathrm{jh}(C).$$

**Proof.** The proof is by induction on the length $\ell(C)$. Let $A = A_0 \supseteq A_1 \supseteq \cdots \supseteq A_n = \{0\}$ be a composition series for $A$ with factors $Q_1, \ldots, Q_n$. If $\ell(C) = 1$, then $C$ is simple, and so

$$B \supseteq A \supseteq A_1 \supseteq \cdots \supseteq A_n = \{0\}$$

is a composition series for $B$ with factors $C, Q_1, \ldots, Q_n$. Therefore,

$$\mathrm{jh}(B) = C \oplus Q_1 \oplus \cdots \oplus Q_n = \mathrm{jh}(C) \oplus \mathrm{jh}(A).$$

For the inductive step, let $\ell(C) > 1$. Choose a maximal submodule $C_1$ of $C$ (which exists because $C$ has a composition series). If $\nu\colon B \to C$ is the given

surjection, define $B_1 = \nu^{-1}(C_1)$. There is a commutative diagram (with vertical arrows inclusions)

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & B & \xrightarrow{\nu} & C & \longrightarrow & 0 \\ & & \uparrow & & \uparrow & & \uparrow & & \\ 0 & \longrightarrow & A & \longrightarrow & B_1 & \longrightarrow & C_1 & \longrightarrow & 0. \end{array}$$

Since $C_1$ is a maximal submodule of $C$, the quotient module

$$C'' = C/C_1$$

is simple. Note that $B/B_1 \cong (B/A)/(B_1/A) \cong C/C_1 = C''$. By the base step, we have

$$\operatorname{jh}(C) = C'' \oplus \operatorname{jh}(C_1) \quad \text{and} \quad \operatorname{jh}(B) = C'' \oplus \operatorname{jh}(B_1).$$

By the inductive hypothesis,

$$\operatorname{jh}(B_1) = \operatorname{jh}(A) \oplus \operatorname{jh}(C_1).$$

Therefore,

$$\begin{aligned} \operatorname{jh}(B) &= C'' \oplus \operatorname{jh}(B_1) \\ &\cong C'' \oplus \operatorname{jh}(A) \oplus \operatorname{jh}(C_1) \\ &\cong \operatorname{jh}(A) \oplus C'' \oplus \operatorname{jh}(C_1) \\ &\cong \operatorname{jh}(A) \oplus \operatorname{jh}(C). \quad \bullet \end{aligned}$$

**Theorem C-4.93.** *Let $\mathcal{C}$ be a G-category in which every $M$ in $\mathcal{C}$ has a composition series. Then $\mathcal{C}$ is a Jordan–Hölder category if and only if $G_0(\mathcal{C})$ is a free abelian group with basis the set $\mathcal{B}$ of all nonisomorphic simple $S$ in $\mathcal{C}$.*

**Proof.** Assume that $G_0(\mathcal{C})$ is free abelian with basis $\mathcal{B}$. Since $[0]$ is not a member of a basis, we have $[S] \neq [0]$ for every simple object $S$; moreover, if $S \ncong S'$, then $[S] \neq [S']$, for a basis repeats no elements. Let $M \in \operatorname{obj}(\mathcal{C})$, and let $Q_1, \ldots, Q_n$ and $Q'_1, \ldots, Q'_m$ be simple quotients arising, respectively, as factors of two composition series of $M$. By Proposition C-4.90, we have

$$[Q_1] + \cdots + [Q_n] = [M] = [Q'_1] + \cdots + [Q'_m].$$

Uniqueness of expression in terms of the basis $\mathcal{B}$ says, for each $Q'_j$, that there exists $Q_i$ with $[Q_i] = [Q'_j]$; in fact, the number of any $[Q_i]$ on the left-hand side is equal to the number of copies of $[Q'_j]$ on the right-hand side. Therefore, $\mathcal{C}$ is a Jordan–Hölder category.

Conversely, assume that the Jordan–Hölder Theorem holds for $\mathcal{C}$. Since every $M \in \operatorname{obj}(\mathcal{C})$ has a composition series, Proposition C-4.90 shows that $\mathcal{B}$ generates $G_0(\mathcal{C})$. Let $S$ in $\operatorname{obj}(\mathcal{C})$ be simple. If $[S] = [T]$, then Proposition C-4.86 says there are $C, U, V \in \operatorname{obj}(\mathcal{C})$ and exact sequences $0 \to U \to S \oplus C \to V \to 0$ and $0 \to U \to T \oplus C \to V \to 0$. Lemma C-4.92 gives

$$\operatorname{jh}(S) \oplus \operatorname{jh}(C) \cong \operatorname{jh}(U) \oplus \operatorname{jh}(V) \cong \operatorname{jh}(T) \oplus \operatorname{jh}(C).$$

By Lemma C-4.91, we may cancel the simple summands one by one until we are left with $S \cong T$. A similar argument shows that if $S$ is simple, then $[S] \neq [0]$.

Finally, let us show that every element in $G_0(\mathcal{C})$ has a unique expression as a linear combination of elements in $\mathcal{B}$. Suppose there are positive integers $m_i$ and $n_j$ so that

$$\sum_i m_i[S_i] - \sum_j n_j[T_j] = [0], \tag{1}$$

where the $S_i$ and $T_j$ are simple and $S_i \ncong T_j$ for all $i, j$. If we denote the direct sum of $m_i$ copies of $S_i$ by $m_i S_i$, then Eq. (1) gives

$$\Big[\bigoplus_i m_i S_i\Big] = \Big[\bigoplus_j n_j T_j\Big].$$

By Proposition C-4.86, there are modules $C, U, V$ and exact sequences

$$0 \to U \to C \oplus \bigoplus_i m_i S_i \to V \to 0 \quad \text{and} \quad 0 \to U \to C \oplus \bigoplus_j n_j T_j \to V \to 0,$$

and Lemma C-4.92 gives

$$\mathrm{jh}\Big(\bigoplus_i m_i S_i\Big) \cong \mathrm{jh}\Big(\bigoplus_j n_j T_j\Big).$$

By Lemma C-4.91, some $S_i$ occurs on the right-hand side, contradicting $S_i \ncong T_j$ for all $i, j$. Therefore, Eq. (1) cannot occur. •

The reader is probably curious why the Grothendieck group $K_0(R)$ has a subscript (he or she may also be curious about why the letter K is used). Grothendieck constructed $K_0(R)$ to generalize the Riemann–Roch Theorem in algebraic geometry; he used the letter K to abbreviate the German *Klasse*. Atiyah–Hirzebruch [**13**] applied this idea to the category of vector bundles over compact Hausdorff spaces $X$ obtaining *topological K-theory*: a sequence of functors $K^{-n}(X)$ for $n \geq 0$. This sequence satisfies all the Eilenberg–Steenrod Axioms characterizing the cohomology groups of topological spaces (actually, of pairs $(X, Y)$, where $X$ is a compact Hausdorff space and $Y \subseteq X$ is a closed subspace) except for the *Dimension Axiom.*[14] In particular, there is a long exact sequence

$$\to K^{-1}(X, Y) \to K^{-1}(X) \to K^{-1}(Y) \to K^0(X, Y) \to K^0(X) \to K^0(Y).$$

At that time, the Grothendieck group $K_0(R)$ had no need for a subscript, for there was no analog of the higher groups $K^{-i}$. This was remedied by Bass–Schanuel who defined $K_1$ as well as relative groups $K_0(R, I)$ and $K_1(R, I)$, where $I$ is a two-sided ideal in $R$. Moreover, they exhibited a six-term exact sequence involving three $K_0$ (now subscripted) and three $K_1$. In his review (MR0249491(40#2736)) of Bass's book [**17**], Alex Heller wrote,

> Algebraic $K$-theory flows from two sources. The (J. H. C.) Whitehead torsion, introduced in order to study the topological notion of simple homotopy type, leads to the groups $K_1$. The Grothendieck group of projective modules over a ring leads to $K_0$. The latter

[14] The Dimension Axiom says that if $X$ is a one-point space and $G$ is an abelian group, then $H^0(X, G) \cong G$ and $H^i(X, G) \cong \{0\}$ for all $i > 0$.

> notion was applied by Atiyah and Hirzebruch in order to construct a new cohomology theory which has been enormously fruitful in topology.
>
> The observation that these two ideas could be unified in a beautiful and powerful theory with widespread applications in algebra is due to Bass, who is also responsible for a major portion of those applications.

We have seen that $K_0$ can be regarded as a functor on the category of rings, so that a ring homomorphism $\varphi\colon R \to S$ gives a homomorphism $K_0(R) \to K_0(S)$; studying its kernel leads to relative $K_0$. Milnor [**156**] defined a functor $K_2$ extending the exact sequence of lower $K$'s, which has applications to field theory and group theory. Quillen then constructed an infinite sequence $(K_n(\mathcal{C}))_{n\geq 0}$ of abelian groups by associating a topological space $X(\mathcal{C})$ to certain categories $\mathcal{C}$. He then defined $K_n(\mathcal{C}) = \pi_n(X(\mathcal{C}))$ for all $n \geq 0$, the homotopy groups of this space, and proved that his $K_n$ coincide with those for $n = 0, 1, 2$ (Rosenberg [**183**]).

Here are two interesting applications. The Merkurjev–Suslin Theorem relates $K$-theory to Brauer groups. Let $n$ be a positive integer and let $F$ be a field containing a primitive $n$th root of unity. If the characteristic of $F$ is either 0 or does not divide $n$, then $K_2(F) \otimes_{\mathbb{Z}} \mathbb{Z}_n \cong \mathrm{Br}(F)[n]$, the subgroup of the Brauer group $\mathrm{Br}(F)$ consisting of all elements whose order divides $n$. The second application involves what are called the ***Lichtenbaum Conjectures***, which relate the orders of certain $K$-groups with values of zeta functions. A generalization of this conjecture was solved by Voevodsky in 2011; here is a portion of its review in Math Reviews (MR2811603 (2012j:14030)):

> This landmark paper completes the publication of Voevodsky's celebrated proof of the Bloch-Kato conjecture—it is now a theorem that the norm residue homomorphism
>
> $$K_n^M(k)/l \to H^n_{\text{ét}}(k, \mu_l^{\otimes n})$$
>
> is an isomorphism for all fields $k$, all primes $l$ with $(l, \operatorname{char} k) = 1$ and all $n$. The norm residue homomorphism is a special case of a comparison morphism between (Beilinson) motivic cohomology and (Lichtenbaum) étale motivic cohomology, and indeed the more general Beilinson-Lichtenbaum conjecture is also a consequence of the results of the paper. The proof of these conjectures is one of the fundamental results in algebraic K-theory and motivic cohomology, and has involved a lot of time as well as a lot of work by a lot of people. The result itself is a great piece of mathematics that allows a much better understanding of the relation between motivic cohomology or algebraic K-theory with their étale counterparts, but even more importantly, the methods developed to prove it (derived categories of motives, motivic cohomology and homotopy) have already had a large impact on mathematics and will continue to do so in the years to come. Needless to say, Voevodsky

was awarded the Fields Medal in 2002 for developing these methods leading to a proof of the case $l = 2$, which had been known as Milnor's conjecture. . . .

## Exercises

∗ **C-4.54.** For any ring $R$, prove that $K_0(R) \cong K_0(R^{\mathrm{op}})$.

**Hint.** Prove that $\mathrm{Hom}_R(\quad, R)\colon R\text{-}\mathbf{Proj} \to R^{\mathrm{op}}\text{-}\mathbf{Proj}$ is an equivalence of categories.

**C-4.55.** Prove that $K_0(R) \cong K_0(\mathrm{Mat}_n(R))$ for every ring $R$.

**Hint.** The functor $\mathrm{Hom}_R(R^n, \quad)$ preserves finitely generated projectives.

**C-4.56.** If a ring $R$ is a direct product, $R = R_1 \times R_2$, prove that

$$K_0(R) \cong K_0(R_1) \oplus K_0(R_2).$$

**C-4.57.** If a commutative semigroup $S$ is an abelian group, prove that $K_0(S) \cong S$.

**C-4.58.** Let $R$ be a domain, and let $a \in R$ be neither 0 nor a unit. If $\mathcal{C}$ is the category of all finitely generated $R$-modules, prove that $[R/Ra] = 0$ in $G_0(\mathcal{C})$.

**Hint.** Use the exact sequence $0 \to R \xrightarrow{\mu_a} R \to R/Ra \to 0$, where $\mu_a\colon r \mapsto ar$.

**C-4.59.** If $\mathcal{C}$ and $\mathcal{A}$ are $G$-categories, prove that every exact functor $F\colon \mathcal{C} \to \mathcal{A}$ defines a homomorphism $K_0(\mathcal{C}) \to K_0(\mathcal{A})$ with $[C] \mapsto [FC]$ for all $C \in \mathrm{obj}(\mathcal{C})$.

**C-4.60.** Let $\mathcal{C}$ be the category of all finitely generated abelian groups.

(i) Prove that $K_0(\mathcal{C})$ is free abelian of countably infinite rank.

(ii) Prove that $G_0(\mathcal{C}) \cong \mathbb{Z}$.

(iii) When are two finitely generated abelian groups stably isomorphic?

**C-4.61.** If $\mathcal{C}$ is the category of all finitely generated $R$-modules over a PID $R$, compute $K_0(\mathcal{C})$ and $G_0(\mathcal{C})$.

**C-4.62.** If $\mathcal{C}$ is the category of all finitely generated $\mathbb{Z}_6$-modules, prove that $K_0(\mathcal{C}) \cong \mathbb{Z} \oplus \mathbb{Z}$.

**C-4.63.** If $R = k[x_1, \ldots, x_n]$, where $k$ is a field, use Hilbert's Syzygy Theorem (Corollary C-5.150) to prove that $G_0(R) \cong \mathbb{Z}$. (Of course, this also follows from the Quillen–Suslin Theorem that finitely generated projective $R$-modules are free.)

∗ **C-4.64. (Eilenberg)** Prove that if $P$ is a projective $R$-module (over some commutative ring $R$), then there exists a free $R$-module $Q$ with $P \oplus Q$ a free $R$-module. Conclude that $K_0(\mathcal{C}) = \{0\}$ for $\mathcal{C}$ the category of countably generated projective $R$-modules.

**Hint.** $Q$ need not be finitely generated.

# Commutative Rings III

## C-5.1. Local and Global

It is often easier to examine algebraic structures "one prime at a time". For example, let $G$ and $H$ be finite groups. If $G \cong H$, then their Sylow $p$-subgroups are isomorphic for all primes $p$; studying $G$ and $H$ *locally* means studying their $p$-subgroups. This local information is not enough to determine whether $G \cong H$; for example, the symmetric group $S_3$ and the cyclic group $\mathbb{Z}_6$ are nonisomorphic groups having isomorphic Sylow 2-subgroups and isomorphic Sylow 3-subgroups. The *global problem* assumes that the Sylow $p$-subgroups of groups $G$ and $H$ are isomorphic, for all primes $p$, and asks what else is necessary to conclude that $G \cong H$. For general groups, this global problem is intractible (although there are partial results). However, this strategy does succeed for finite *abelian* groups. The local problem involves primary components (Sylow subgroups), which are direct sums of cyclic groups, and the global problem is solved by Primary Decomposition: every finite abelian group is the direct sum of its primary components. In this case, the local information is sufficient to solve the global problem. In general, local problems are simpler than global ones, and their solutions can give very useful information.

### ▪ Subgroups of $\mathbb{Q}$

We begin with another group-theoretic illustration of local and global investigation, after which we will consider localization of commutative rings.

**Definition.** Let $R$ be a domain with $Q = \text{Frac}(R)$. If $M$ is an $R$-module, define

$$\text{rank}(M) = \dim_Q(Q \otimes_R M).$$

For example, the rank of an abelian group $G$ is defined as $\dim_{\mathbb{Q}}(\mathbb{Q} \otimes_{\mathbb{Z}} G)$. Recall, when $R$ is a domain, that an $R$-module $M$ is *torsion-free* if it has no nonzero elements of finite order; that is, if $r \in R$ and $m \in M$ are nonzero, then $rm$ is nonzero.

**Lemma C-5.1.** *Let $R$ be a domain with $Q = \mathrm{Frac}(R)$, and let $M$ be an $R$-module.*

(i) *If $0 \to M' \to M \to M'' \to 0$ is a short exact sequence of $R$-modules, then*
$$\mathrm{rank}(M) = \mathrm{rank}(M') + \mathrm{rank}(M'').$$

(ii) *An $R$-module $M$ is torsion if and only if* $\mathrm{rank}(M) = 0$.

(iii) *Let $M$ be torsion-free. Then $M$ has rank* 1 *if and only if it is isomorphic to a nonzero $R$-submodule of $Q$.*

**Proof.**

(i) By Corollary B-4.106 in Part 1, the fraction field $Q$ is a flat $R$-module. Therefore, $0 \to Q \otimes_R M' \to Q \otimes_R M \to Q \otimes_R M'' \to 0$ is a short exact sequence of vector spaces over $Q$, and the result is a standard result of linear algebra, Exercise A-7.9 on page 259 in Part 1.

(ii) If $M$ is torsion, then $Q \otimes_R M = \{0\}$, by a routine generalization of Proposition B-4.92 in Part 1 (divisible $\otimes$ torsion $= \{0\}$). Hence, $\mathrm{rank}(M) = 0$.

If $M$ is not torsion, then there is an exact sequence $0 \to R \to M$ (if $m \in M$ has infinite order, then $r \mapsto rm$ is an injection $R \to M$). Since $Q$ is flat, the sequence $0 \to Q \otimes_R R \to Q \otimes_R M$ is exact. By Proposition B-4.84 in Part 1, $Q \otimes_R R \cong Q \neq \{0\}$, so that $Q \otimes_R M \neq \{0\}$. Therefore, $\mathrm{rank}(M) > 0$.

(iii) If $\mathrm{rank}(M) = 1$, then $M \neq \{0\}$, and exactness of $0 \to R \to Q$ gives exactness of $\mathrm{Tor}_1^R(Q/R, M) \to R \otimes_R M \to Q \otimes_R M$. Since $M$ is torsion-free, Lemma C-3.101 gives $\mathrm{Tor}_1^R(Q/R, M) = \{0\}$. As always, $R \otimes_R M \cong M$, while $Q \otimes_R M \cong Q$ because $M$ has rank 1. Therefore, $M$ is isomorphic to an $R$-submodule of $Q$.

Conversely, if $M$ is isomorphic to an $R$-submodule of $Q$, there is an exact sequence $0 \to M \to Q$. Since $Q$ is flat, by Corollary B-4.106 in Part 1, we have exactness of $0 \to Q \otimes_R M \to Q \otimes_R Q$. Now this last sequence is an exact sequence of vector spaces over $Q$. Since $Q \otimes_R Q \cong Q$ is one-dimensional, its nonzero subspace $Q \otimes_R M$ is also one-dimensional; that is, $\mathrm{rank}(M) = 1$. •

**Example C-5.2.** The following abelian groups are torsion-free of rank 1.

(i) The group $\mathbb{Z}$ of integers.

(ii) The additive group $\mathbb{Q}$.

(iii) The ***2-adic fractions*** $\mathbb{Z}_{(2)} = \{a/b \in \mathbb{Q} : b \text{ is odd}\}$.

(iv) The set of all rationals having a finite decimal expansion.

(v) The set of all rationals having squarefree denominator. ◀

**Proposition C-5.3.** *Let $R$ be a domain with $Q = \mathrm{Frac}(R)$. Two submodules $A$ and $B$ of $Q$ are isomorphic if and only if there is $c \in Q$ with $B = cA$.*

**Proof.** If $B = cA$, then $A \cong B$ via $a \mapsto ca$. Conversely, suppose that $f\colon A \to B$ is an isomorphism. Regard $f$ as an $R$-injection $A \to Q$ by composing it with the inclusion $B \to Q$. Since $Q$ is an injective $R$-module, by Proposition B-4.58 in Part 1, there is an $R$-map $g\colon Q \to Q$ extending $f$. But Proposition B-4.58 in Part 1 says that there is $c \in Q$ with $g(x) = cx$ for all $x \in Q$. Thus, $B = f(A) = g(A) = cA$. •

Recall, for each prime $p$, that the ring of *$p$-adic fractions* is the subring of $\mathbb{Q}$:

$$\mathbb{Z}_{(p)} = \{a/b \in \mathbb{Q} : (b, p) = 1\}.$$

For example, if $p = 2$, then $\mathbb{Z}_{(2)}$ is the group of 2-adic fractions.

**Proposition C-5.4.**

(i) *For each prime $p$, the ring $\mathbb{Z}_{(p)}$ has a unique maximal ideal; that is, it is a local*[1] PID.

(ii) *If $G$ is a torsion-free abelian group of rank* 1, *then $\mathbb{Z}_{(p)} \otimes_{\mathbb{Z}} G$ is a torsion-free $\mathbb{Z}_{(p)}$-module of rank* 1.

(iii) *If $M$ is a torsion-free $\mathbb{Z}_{(p)}$-module of rank* 1, *then $M \cong \mathbb{Z}_{(p)}$ or $M \cong \mathbb{Q}$.*

**Proof.**

(i) We show that the only nonzero ideals $I$ in $\mathbb{Z}_{(p)}$ are $(p^n)$, for $n \geq 0$; it will then follow that $\mathbb{Z}_{(p)}$ is a PID and that $(p)$ is its unique maximal ideal. Each nonzero $x \in \mathbb{Z}_{(p)}$ has the form $a/b$ for integers $a$ and $b$, where $(b, p) = 1$. But $a = p^n a'$, where $n \geq 0$ and $(a', p) = 1$; that is, there is a unit $u \in \mathbb{Z}_{(p)}$, namely, $u = a'/b$, with $x = up^n$. Let $I \neq (0)$ be an ideal. Of all the nonzero elements in $I$, choose $x = up^n \in I$, where $u$ is a unit, with $n$ minimal. Then $I = (x) = (p^n)$, for if $y \in I$, then $y = vp^m$, where $v$ is a unit and $n \leq m$. Hence, $p^n \mid y$ and $y \in (p^n)$.

(ii) Since $\mathbb{Z}_{(p)} \subseteq \mathbb{Q}$, it is an additive torsion-free abelian group of rank 1, and so it is flat (Corollary B-4.105 in Part 1). Hence, exactness of $0 \to G \to \mathbb{Q}$ gives exactness of

$$0 \to \mathbb{Z}_{(p)} \otimes_{\mathbb{Z}} G \to \mathbb{Z}_{(p)} \otimes_{\mathbb{Z}} \mathbb{Q}.$$

By Exercise C-5.4 on page 426, $\mathbb{Z}_{(p)} \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathbb{Q} = \mathrm{Frac}(\mathbb{Z}_{(p)})$, so that $\mathbb{Z}_{(p)} \otimes_{\mathbb{Z}} G$ is a torsion-free $\mathbb{Z}_{(p)}$-module of rank 1.

(iii) There is no loss in generality in assuming that $M \subseteq \mathbb{Q}$ and that $1 \in M$. Consider the equations $p^n y_n = 1$ for $n \geq 0$. We claim that if all these equations are solvable for $y_n \in M$, then $M = \mathbb{Q}$. If $a/b \in \mathbb{Q}$, then $a/b = a/p^n b'$, where $(b', p) = 1$, and so $a/b = (a/b')y_n$; as $a/b' \in \mathbb{Z}_{(p)}$, we have $a/b \in M$. We may now assume that there is a largest $n \geq 0$ for which the equation $p^n y_n = 1$ is solvable for $y_n \in M$. We claim that $M = \langle y_n \rangle$, the cyclic submodule generated by $y_n$, which will show that $M \cong \mathbb{Z}_{(p)}$. If $m \in M$, then $m = c/d = p^r c'/p^s d' = (c'/d')(1/p^{s-r})$, where $(c', p) = 1 = (d', p)$. Since $c'/d'$ is a unit in $\mathbb{Z}_{(p)}$, we have $1/p^{s-r} \in M$, and so $s - r \leq n$; that is, $s - r = n - \ell$ for some $\ell \geq 0$. Hence, $1/p^{s-r} = 1/p^{n-\ell} = p^\ell/p^n = p^\ell y_n$, and so $m = (c'p^\ell/d')y_n \in \langle y_n \rangle$. •

**Definition.** A ***discrete valuation ring***, abbreviated DVR, is a local PID that is not a field.

[1] In this section, *local rings* are commutative rings having a unique maximal ideal. Often, local rings are also assumed to be noetherian. On the other hand, the term sometimes means a noncommutative ring having a unique maximal left or right ideal.

Some examples of DVRs are $\mathbb{Z}_{(p)}$, the $p$-adic integers $\mathbb{Z}_p$, and formal power series $k[[x]]$, where $k$ is a field.

**Definition.** Two abelian groups $G$ and $H$ are ***locally isomorphic*** if $\mathbb{Z}_{(p)} \otimes_{\mathbb{Z}} G \cong \mathbb{Z}_{(p)} \otimes_{\mathbb{Z}} H$ for all primes $p$.

We have solved the local problem for torsion-free abelian groups $G$ of rank 1; associate to $G$ the family $\mathbb{Z}_{(p)} \otimes_{\mathbb{Z}} G$ of $\mathbb{Z}_{(p)}$-modules, one for each prime $p$.

**Example C-5.5.** Let $G$ be the subgroup of $\mathbb{Q}$ consisting of those rationals having squarefree denominator. Then $G$ and $\mathbb{Z}$ are locally isomorphic, but they are not isomorphic, because $\mathbb{Z}$ is finitely generated and $G$ is not. ◀

We are now going to solve the global problem for torsion-free abelian groups of rank 1.

**Definition.** Let $G$ be an abelian group, let $x \in G$, and let $p$ be a prime. Then $x$ is ***divisible by*** $p^n$ ***in*** $G$ if there exists $y_n \in G$ with $p^n y_n = x$. Define the $p$-***height*** of $x$, denoted by $h_p(x)$, by

$$h_p(x) = \begin{cases} \infty & \text{if } x \text{ is divisible by } p^n \text{ in } G \text{ for all } n \geq 0, \\ k & \text{if } x \text{ is divisible by } p^k \text{ in } G \text{ but not by } p^{k+1}. \end{cases}$$

The ***height sequence*** (or ***characteristic***) of $x$ in $G$, where $x$ is nonzero, is the sequence

$$\chi(x) = \chi_G(x) = (h_2(x), h_3(x), h_5(x), \dots, h_p(x), \dots).$$

Thus, $\chi(x)$ is a sequence $(h_p)$, where $h_p = \infty$ or $h_p \in \mathbb{N}$. Let $G \subseteq \mathbb{Q}$ and let $x \in G$ be nonzero. If $\chi(x) = (h_p)$ and $a = p_1^{f_1} \cdots p_n^{f_n}$, then $\frac{1}{a}x \in G$ if and only if $f_{p_i} \leq h_{p_i}$ for $i = 1, \dots, n$.

**Example C-5.6.** Each of the groups in Example C-5.2 contains $x = 1$.

(i) In $\mathbb{Z}$,
$$\chi_{\mathbb{Z}}(1) = (0, 0, 0, \dots).$$

(ii) In $\mathbb{Q}$,
$$\chi_{\mathbb{Q}}(1) = (\infty, \infty, \infty, \dots).$$

(iii) In $\mathbb{Z}_{(2)}$,
$$\chi_{\mathbb{Z}_{(2)}}(1) = (0, \infty, \infty, \dots).$$

(iv) If $G$ is the group of all rationals having a finite decimal expansion, then
$$\chi_G(1) = (\infty, 0, \infty, 0, 0, \dots).$$

(v) If $H$ is the group of all rationals having squarefree denominator, then
$$\chi_H(1) = (1, 1, 1, \dots). \quad ◀$$

**Lemma C-5.7.** *Let $(k_2, k_3, \dots, k_p, \dots)$ be a sequence, where $k_p \in \mathbb{N} \cup \{\infty\}$ for all $p$. There exists a unique subgroup $G \subseteq \mathbb{Q}$ containing 1 such that*

$$\chi_G(1) = (k_2, k_3, \dots, k_p, \dots).$$

**Proof.** Define

$$G = \{a/c \in \mathbb{Q} : c = p_1^{e_1} \cdots p_n^{e_n} \text{ and } e_p \le k_p \text{ whenever } k_p < \infty\}.$$

That $G$ is a subgroup follows from the elementary fact that if $c = p_1^{e_1} \cdots p_n^{e_n}$ and $d = p_1^{f_1} \cdots p_n^{f_n}$, then $1/c - 1/d = m/p_1^{\ell_1} \cdots p_n^{\ell_n}$, where $m \in \mathbb{Z}$ and $\ell_p \le \max\{e_p, f_p\} \le k_p$ for all $p$. It is clear that $\chi_G(1) = (k_2, k_3, \dots, k_p, \dots)$. To prove uniqueness of $G$, let $H$ be another such subgroup. Suppose $k_p < \infty$, and let $u/v \in \mathbb{Q}$, where $u/v$ is in lowest terms. If $v = p_1^{r_1} \cdots p_n^{r_n}$ and $r_p > k_p$, then $u/v \notin G$. But if $u/v \in H$, then $h_p(1) \ge r_p > k_p$, contradicting $\chi_H(1) = (k_2, k_3, \dots, k_p, \dots)$. Hence, $u/v \notin H$ if and only if $u/v \notin G$; that is, $G = H$. •

Different elements in a torsion-free abelian group of rank 1 may have different height sequences. For example, if $G$ is the group of rationals having finite decimal expansions, then 1 and $\frac{63}{8}$ lie in $G$, and

$$\chi(1) = (\infty, 0, \infty, 0, 0, 0, \dots) \quad \text{and} \quad \chi(\tfrac{63}{8}) = (\infty, 2, \infty, 1, 0, 0, \dots).$$

Thus, these height sequences agree when $h_p = \infty$, but they disagree for some finite $p$-heights: $h_3(1) \ne h_3(\frac{63}{8})$ and $h_7(1) \ne h_7(\frac{63}{8})$.

**Definition.** Two height sequences $(h_2, h_3, \dots, h_p, \dots)$ and $(k_2, k_3, \dots, k_p, \dots)$ are ***equivalent***, denoted by

$$(h_2, h_3, \dots, h_p, \dots) \sim (k_2, k_3, \dots, k_p, \dots),$$

if there are only finitely many $p$ for which $h_p \ne k_p$ and, for such primes $p$, neither $h_p$ nor $k_p$ is $\infty$.

It is easy to see that equivalence just defined is, in fact, an equivalence relation.

**Lemma C-5.8.** *If $G$ is a torsion-free abelian group of rank 1 and $x, y \in G$ are nonzero, then their height sequences $\chi(x)$ and $\chi(y)$ are equivalent.*

**Proof.** By Lemma C-5.1(iii), we may assume that $G \subseteq \mathbb{Q}$. If $b = p_1^{e_1} \cdots p_n^{e_n}$, then it is easy to see that $h_p(bx) = h_p(x)$ for all $p \notin \{p_1, \dots, p_n\}$, while

$$h_{p_i}(bx) = e_i + h_{p_i}(x)$$

for $i = 1, \dots, n$ (we agree that $e_i + \infty = \infty$). Hence, $\chi(x) \sim \chi(bx)$. Since $x, y \in G \subseteq \mathbb{Q}$, we have $x/y = a/b$ for integers $a, b$, so that $bx = ay$. Therefore, $\chi(x) \sim \chi(bx) = \chi(ay) \sim \chi(y)$. •

**Definition.** The equivalence class of a height sequence is called a ***type***. If $G$ is a torsion-free abelian group of rank 1, then its ***type***, denoted by $\tau(G)$, is the type of a height sequence $\chi(x)$, where $x$ is a nonzero element of $G$.

Lemma C-5.8 shows that $\tau(G)$ depends only on $G$ and not on the choice of nonzero element $x \in G$. We now solve the global problem for subgroups of $\mathbb{Q}$.

**Theorem C-5.9.** *If $G$ and $H$ are torsion-free abelian groups of rank 1, then $G \cong H$ if and only if $\tau(G) = \tau(H)$.*

**Proof.** Let $\varphi\colon G \to H$ be an isomorphism. If $x \in G$ is nonzero, it is easy to see that $\chi(x) = \chi(\varphi(x))$, and so $\tau(G) = \tau(H)$.

For the converse, choose nonzero $x \in G$ and $y \in H$. By the definition of equivalence, there are primes $p_1, \ldots, p_n, q_1, \ldots, q_m$ with $h_{p_i}(x) < h_{p_i}(y) < \infty$, with $\infty > h_{q_j}(x) > h_{q_j}(y)$, and with $h_p(x) = h_p(y)$ for all other primes $p$. If we define $b = \prod p_i^{h_{p_i}(y) - h_{p_i}(x)}$, then $bx \in G$ and $h_{p_i}(bx) = \big(h_{p_i}(y) - h_{p_i}(x)\big) + h_{p_i}(x) = h_{p_i}(y)$. A similar construction, using $a = \prod q_j^{h_{q_j}(x) - h_{q_j}(y)}$, gives $\chi_G(bx) = \chi_H(ay)$. We have found elements $x' = bx \in G$ and $y' = ay \in H$ having the same height sequence.

Since $G$ has rank 1, there is an injection $f\colon G \to \mathbb{Q}$; write $A = \operatorname{im} f$, so that $A \cong G$. If $u = f(x')$, then $\chi_A(u) = \chi_G(x')$. If $g\colon \mathbb{Q} \to \mathbb{Q}$ is defined by $q \mapsto q/u$, then $g(A)$ is a subgroup of $\mathbb{Q}$ containing 1 with $\chi_{g(A)}(1) = \chi_A(u)$. Of course, $G \cong g(A)$. Similarly, there is a subgroup $C \subseteq \mathbb{Q}$ containing 1 with $\chi_C(1) = \chi_H(y') = \chi_G(x')$. By Lemma C-5.7, $G \cong g(A) = C \cong H$. •

We can also solve the global problem for subrings of $\mathbb{Q}$.

**Corollary C-5.10.** *The following are equivalent for subrings $R$ and $S$ of $\mathbb{Q}$:*

(i) $R = S$.

(ii) $R \cong S$.

(iii) *$R$ and $S$ are locally isomorphic*: $R_{(p)} \cong S_{(p)}$ *for all primes $p$.*

**Proof.**

(i) $\Rightarrow$ (ii). This is obvious.

(ii) $\Rightarrow$ (iii). This, too, is obvious.

(iii) $\Rightarrow$ (i). If $1/b \in R$, where $b \in \mathbb{Z}$, then its prime factorization is $b = \pm p_1^{e_1} \cdots p_t^{e_t}$, where $e_i > 0$. Thus, $h_{p_i}(1) > 0$, and so $h_{p_i}(1) = \infty$. Every ring contains 1; here, its height is a sequence of 0's and $\infty$'s. Since $R_{(p)} \cong S_{(p)}$, the height of 1 in $R$ is the same as its height in $S$: $\chi_R(1) = \chi_S(1)$. Therefore $R = S$, by the uniqueness in Lemma C-5.7. •

The uniqueness theorems for groups and rings just proved are complemented by existence theorems.

**Corollary C-5.11.** *Given any type $\tau$, there exists a torsion-free abelian group $G$ of rank* 1*, unique up to isomorphism, with $\tau(G) = \tau$. Hence, there are uncountably many nonisomorphic subgroups of $\mathbb{Q}$.*

**Proof.** The existence of $G$ follows at once from Lemma C-5.7, while uniqueness up to isomorphism follows from Theorem C-5.9. But there are uncountably many types; for example, two height sequences of 0's and $\infty$'s are equivalent if and only if they are equal. •

**Corollary C-5.12.**

(i) *If $R$ is a subring of $\mathbb{Q}$, then the height sequence of* 1 *consists of* 0*'s and* $\infty$*'s.*

(ii) *There are uncountably many nonisomorphic subrings of $\mathbb{Q}$. In fact, the additive groups of distinct subrings of $\mathbb{Q}$ are not isomorphic.*

**Proof.**

(i) If $h_p(1) > 0$, then $\frac{1}{p} \in R$. Since $R$ is a ring, $\left(\frac{1}{p}\right)^n = \frac{1}{p^n} \in R$ for all $n \geq 1$, and so $h_p(1) = \infty$.

(ii) If $R$ and $S$ are distinct subrings of $\mathbb{Q}$, then the height sequences of 1 are distinct, by part (i). Both statements follow from the observation that two height sequences whose only terms are 0 and $\infty$ are equivalent if and only if they are equal. •

Kurosh classified torsion-free $\mathbb{Z}_{(p)}$-modules $G$ of finite rank with invariants $\operatorname{rank}(G) = n$, $p$-***rank***$(G) = \dim_{\mathbb{F}_p}(\mathbb{F}_p \otimes G)$, and an equivalence class of $n \times n$ nonsingular matrices $M_p$ over the $p$-adic numbers $\mathbb{Q}_p$ (Fuchs [**73**], pp. 154–158). This theorem was globalized by Derry; the invariants are rank, $p$-ranks for all primes $p$, and an equivalence class of matrix sequences $(M_p)$, where $M_p$ is an $n \times n$ nonsingular matrix over $\mathbb{Q}_p$. Alas, there is no normal form for the matrix sequences (nor for the matrices in the local case), and so this theorem does not have many applications. However, Arnold [**6**] used this result to set up an interesting duality.

An abelian group is ***decomposable*** if it is a direct sum of two proper subgroups; otherwise, it is ***indecomposable***. For example, every torsion-free abelian group of rank 1 is indecomposable. A torsion-free abelian group $G$ is called ***completely decomposable*** if it is a direct sum of (possibly infinitely many) groups of rank 1. Baer proved (Fuchs [**73**], p. 114) that every direct summand of a completely decomposable group is itself completely decomposable. Moreover, if $G = \bigoplus_{i\in I} A_i = \bigoplus_{i\in I} B_i$, where all $A_i$ and $B_i$ have rank 1, then, for each type $\tau$, the number of $A_i$ of type $\tau$ is equal to the number of $B_i$ of type $\tau$ (Fuchs [**73**], p. 112). This nice behavior of decompositions does not hold more generally. Every torsion-free abelian group $G$ of finite rank is a direct sum of indecomposable groups (this is not true for infinite rank), but there is virtually no uniqueness for such a decomposition. For example, there exists a group $G$ of rank 6 with

$$G = A_1 \oplus A_2 = B_1 \oplus B_2 \oplus B_3,$$

with all the direct summands indecomposable, with $\operatorname{rank}(A_1) = 1$, $\operatorname{rank}(A_2) = 5$, and $\operatorname{rank}(B_j) = 2$ for $j = 1, 2, 3$ (Fuchs [**73**], p. 135). Thus, the number of indecomposable direct summands in a decomposition is not uniquely determined by $G$, nor is the isomorphism class of any of its indecomposable direct summands. Here is an interesting theorem of Corner that can be used to produce bad examples of torsion-free groups such as the group $G$ of rank 6 just mentioned. Let $R$ be a ring whose additive group is countable, torsion-free, and reduced (it has no nonzero divisible subgroups). Then there exists an abelian group $G$, also countable, torsion-free, and reduced, with $\operatorname{End}(G) \cong R$. Moreover, if the additive group of $R$ has finite rank $n$, then $G$ can be chosen to have rank $2n$ (Fuchs [**73**], p. 231). Jónsson

introduced the notion of *quasi-isomorphism*: two torsion-free abelian groups $G$ and $H$ of finite rank are ***quasi-isomorphic***, denoted by

$$G \doteq H,$$

if each is isomorphic to a subgroup of finite index in the other, and he defined a group to be ***strongly indecomposable*** if it is not quasi-isomorphic to a decomposable group. He then proved that every torsion-free abelian group $G$ of finite rank is quasi-isomorphic to a direct sum of strongly indecomposable groups, and this decomposition is unique in the following sense: if $G \doteq A_1 \oplus \cdots \oplus A_n \doteq B_1 \oplus \cdots \oplus B_m$, then $n = m$ and, after possible reindexing, $A_i \doteq B_i$ for all $i$ (Fuchs [**73**], p. 150). Torsion-free abelian groups still hide many secrets.

## Exercises

**C-5.1.** Prove that $\mathbb{Z}_{(p)} \ncong \mathbb{Q}$ as $\mathbb{Z}_{(p)}$-modules.

**C-5.2.** Prove that the following statements are equivalent for a torsion-free abelian group $G$ of rank 1:

(i) $G$ is finitely generated.

(ii) $G$ is cyclic.

(iii) If $x \in G$ is nonzero, then $h_p(x) \neq \infty$ for all primes $p$ and $h_p(x) = 0$ for almost all $p$.

(iv) $\tau(G) = \tau(\mathbb{Z})$.

**C-5.3.** (i) If $G$ is a torsion-free abelian group of rank 1, prove that the additive group of $\mathrm{End}(G)$ is torsion-free of rank 1.

**Hint.** Use Proposition B-4.58 in Part 1.

(ii) Let $x \in G$ be nonzero with $\chi(x) = (h_2(x), h_3(x), \dots, h_p(x), \dots)$, and let $R$ be the subring of $\mathbb{Q}$ in which $\chi(1) = (k_2, k_3, \dots, k_p, \dots)$, where

$$k_p = \begin{cases} \infty & \text{if } h_p(x) = \infty, \\ 0 & \text{if } h_p(x) \text{ is finite.} \end{cases}$$

Prove that $\mathrm{End}(G) \cong R$. Prove that there are infinitely many $G$ with $\mathrm{Aut}(G) \cong \mathbb{Z}_2$.

(iii) Prove that $G$ and $\mathrm{End}(G)$ are locally isomorphic abelian groups.

$*$ **C-5.4.** (i) Prove that if $G$ and $H$ are torsion-free abelian groups of finite rank, then

$$\mathrm{rank}(G \otimes_{\mathbb{Z}} H) = \mathrm{rank}(G)\,\mathrm{rank}(H).$$

(ii) If $G$ and $H$ are torsion-free abelian groups of rank 1, then $G \otimes_{\mathbb{Z}} H$ is torsion-free of rank 1, by part (i). If $(h_p)$ is the height sequence of a nonzero element $x \in G$ and $(k_p)$ is the height sequence of a nonzero element $y \in H$, prove that the height sequence of $x \otimes y$ is $(m_p)$, where $m_p = h_p + k_p$ (we agree that $\infty + k_p = \infty$).

**C-5.5.** Let $G$ and $G'$ be nonzero subgroups of $\mathbb{Q}$, and let $\chi_G(g) = (k_p)$ and $\chi_{G'}(g') = (k'_p)$ for $g \in G$ and $g' \in G'$.

(i) Let $T$ be the set of all types. Define $\tau \leq \tau'$, where $\tau, \tau' \in T$, if there are height sequences $(k_p) \in \tau$ and $(k'_p) \in \tau'$ with $k_p \leq k'_p$ for all primes $p$. Prove that $\leq$ is a well-defined relation which makes $T$ a partially ordered set.

(ii) Prove that $G$ is isomorphic to a subgroup of $G'$ if and only if $\tau(G) \le \tau(G')$.

(iii) Note that $G \cap G'$ and $G + G'$ are subgroups of $\mathbb{Q}$. Prove that there are elements $x \in G \cap G'$ and $y \in G + G'$ with $\chi_{G\cap G'}(x) = (\min\{k_p, k'_p\})$ and $\chi_{G+G'}(y) = (\max\{k_p, k'_p\})$. Define $\tau(G) \wedge \tau(G')$ to be the type of $(\min\{k_p, k'_p\})$ and $\tau(G) \vee \tau(G')$ to be the type of $(\max\{k_p, k'_p\})$. Prove that these are well-defined operations which make $T$ a lattice.

(iv) Prove that there is an exact sequence $0 \to A \to G \oplus G' \to B \to 0$ with $A \cong G \cap G'$ and $B \cong G + G'$.

(v) Prove that $H = \mathrm{Hom}(G, G')$ is a torsion-free abelian group with $\mathrm{rank}(H) \le 1$ and that $H \neq \{0\}$ if and only if $\tau(G) \le \tau(G')$. Prove that $\mathrm{End}(G) = \mathrm{Hom}(G, G)$ has height sequence $(h_p)$, where $h_p = \infty$ if $k_p = \infty$ and $h_p = 0$ otherwise.

(vi) Prove that $G \otimes_{\mathbb{Z}} G'$ has rank 1 and that $\chi_{G\otimes_{\mathbb{Z}} G'}(g \otimes g') = (k_p + k'_p)$.

**C-5.6.** Let $A$ be a torsion-free abelian group. If $a \in A$ is nonzero, define $\tau(a) = \tau(\chi_A(a))$, and define

$$A(\tau) = \{a \in A : \tau(a) \ge \tau\} \cup \{0\}.$$

Prove that $A(\tau)$ is a fully invariant subgroup of $A$ and that $A/A(\tau)$ is torsion-free.

**C-5.7.** If $G$ is a $p$-primary abelian group, prove that $G$ is a $\mathbb{Z}_{(p)}$-module.

**C-5.8.** Let $\mathcal{C}$ be the category of all torsion-free abelian groups of finite rank.

(i) Prove that the Grothendieck group $G_0(\mathcal{C})$ is generated by all $[A]$ with $\mathrm{rank}(A) = 1$.

(ii) Prove that the Grothendieck group $G_0(\mathcal{C})$ is a commutative ring if we define multiplication by

$$[A][B] = [A \otimes_{\mathbb{Z}} B],$$

and prove that rank induces a ring homomorphism $G_0(\mathcal{C}) \to \mathbb{Z}$.
**Hint.** See the remark on page 406.

**Remark.** Rotman [**189**] showed that $G_0(\mathcal{C})$ is a commutative ring isomorphic to the subring of $\mathbb{Z}^{\mathbb{N}}$ consisting of all functions having finite image.

Consider the Grothendieck group $G'_0(\mathcal{C}) = \mathcal{F}(\mathcal{C})/\mathcal{E}'$, where $\mathcal{F}(\mathcal{C})$ is the free abelian group with basis $\mathrm{Iso}(\mathcal{C})$ and $\mathcal{E}'$ is the subgroup of $\mathcal{F}(\mathcal{C})$ generated by all *split* short exact sequences (thus, the relations are $[A \oplus C] = [A] + [C]$). This Grothendieck group was studied by Lady [**131**]: $G'_0(\mathcal{C})/tG'_0(\mathcal{C})$ is a free abelian group of uncountable rank. He also characterized $tG'_0(\mathcal{C})$ as all $[A] - [B]$ with $A$ and $B$ ***nearly isomorphic***: for each $n > 0$, $B$ contains a subgroup $A_n$ with $A_n \cong A$ and finite index $[B : A_n]$ relatively prime to $n$. ◀

## C-5.2. Localization

**All rings in this section are commutative**.

The ring $\mathbb{Z}$ has infinitely many prime ideals, but $\mathbb{Z}_{(2)}$ has only one nonzero prime ideal, namely, (2) (all odd primes in $\mathbb{Z}$ are invertible in $\mathbb{Z}_{(2)}$). Now $\mathbb{Z}_{(2)}$-modules are much simpler than $\mathbb{Z}$-modules. For example, Proposition C-5.4(iii) says that there are only two $\mathbb{Z}_{(2)}$-submodules of $\mathbb{Q}$ (up to isomorphism): $\mathbb{Z}_{(2)}$ and $\mathbb{Q}$. On the other hand, Corollary C-5.11 says that there are uncountably many nonisomorphic

$\mathbb{Z}$-submodules (= subgroups) of $\mathbb{Q}$. Similar observations lead to a localization-globalization strategy to attack algebraic and number-theoretic problems. The fundamental assumption underlying this strategy is that the local case is simpler than the global. Given a prime ideal $\mathfrak{p}$ in a commutative ring $R$, we will construct local rings $R_{\mathfrak{p}}$; localization looks at problems involving the rings $R_{\mathfrak{p}}$, while globalization uses all such local information to answer questions about $R$. We confess that this section is rather dry and formal.

**Definition.** Every ring $R$ is a monoid under multiplication. A subset $S \subseteq R$ of a commutative ring $R$ is ***multiplicative*** if $S$ is a submonoid not containing 0; that is, $0 \notin S$, $1 \in S$, and $s, s' \in S$ implies $ss' \in S$.

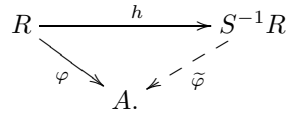
**Example C-5.13.**

(i) If $\mathfrak{p}$ is a prime ideal in $R$, then its set-theoretic complement $S = R - \mathfrak{p}$ is multiplicative (if $a \notin \mathfrak{p}$ and $b \notin \mathfrak{p}$, then $ab \notin \mathfrak{p}$).

(ii) If $R$ is a domain, then the set $S = R^\times$ of all its nonzero elements is multiplicative (this is a special case of part (i), for (0) is a prime ideal in a domain).

(iii) If $a \in R$ is not nilpotent, then the set of its powers $S = \{a^n : n \geq 0\}$ is multiplicative. ◀

**Definition.** If $S \subseteq R$ is multiplicative, a ***localization*** of $R$ is an ordered pair $(S^{-1}R, h)$ which is a solution to the following universal mapping problem: given a commutative $R$-algebra $A$ and an $R$-algebra map $\varphi\colon R \to A$ with $\varphi(s)$ invertible in $A$ for all $s \in S$, there exists a unique $R$-algebra map $\widetilde{\varphi}\colon S^{-1}R \to A$ with $\widetilde{\varphi}h = \varphi$:




The map $h\colon R \to S^{-1}R$ is called the ***localization map***.

As is any solution to a universal mapping problem, a localization $S^{-1}R$ is unique up to isomorphism if it exists, and so we call $S^{-1}R$ *the* localization at $S$. The reason for excluding 0 from a multiplicative set is now apparent, for 0 is invertible only in the zero ring.

Given a multiplicative subset $S \subseteq R$, most authors construct the localization $S^{-1}R$ by generalizing the (tedious) construction of the fraction field of a domain $R$. They define a relation on $R \times S$ by $(r, s) \equiv (r', s')$ if there exists $s'' \in S$ with $s''(rs' - r's) = 0$ (when $R$ is a domain and $S = R^\times$ is the subset of its nonzero elements, this definition reduces to the usual definition of equality of fractions involving cross multiplication). After proving that $\equiv$ is an equivalence relation, $S^{-1}R$ is defined to be the set of all equivalence classes, addition and multiplication are defined and proved to be well-defined, all the $R$-algebra axioms are verified, and the elements of $S$ are shown to be invertible. We prefer to develop the existence and

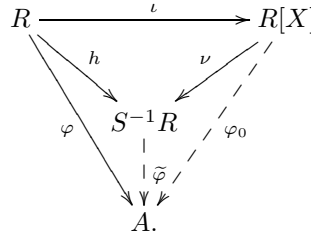
first properties of $S^{-1}R$ in another manner,[2] which is less tedious and which will show that the equivalence relation generalizing cross multiplication arises naturally.

**Theorem C-5.14.** *If $S \subseteq R$ is multiplicative, then the localization $(S^{-1}R, h)$ exists.*

**Proof.** Let $X = (x_s)_{s\in S}$ be an indexed set with $x_s \mapsto s$ a bijection $X \to S$, and let $R[X]$ be the polynomial ring over $R$ with indeterminates $X$. Define

$$S^{-1}R = R[X]/J,$$

where $J$ is the ideal generated by $\{sx_s - 1 : s \in S\}$, and define $h\colon R \to S^{-1}R$ by $h\colon r \mapsto r + J$, where $r$ is a constant polynomial. It is clear that $S^{-1}R$ is an $R$-algebra, that $h$ is an $R$-algebra map, and that each $h(s)$ is invertible. Assume now that $A$ is an $R$-algebra and that $\varphi\colon R \to A$ is an $R$-algebra map with $\varphi(s)$ invertible for all $s \in S$. Consider the diagram in which the top arrow $\iota\colon R \to R[X]$ sends each $r \in R$ to the constant polynomial $r$ and $\nu\colon R[X] \to R[X]/J = S^{-1}R$ is the natural map:

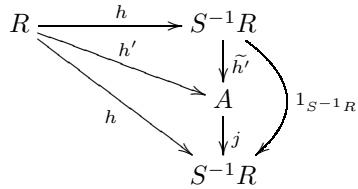



The top triangle commutes because both $h$ and $\nu\iota$ send $r \in R$ to $r+J$. Define an $R$-algebra map $\varphi_0\colon R[X] \to A$ by $\varphi_0(x_s) = \varphi(s)^{-1}$ for all $x_s \in X$. Clearly, $J \subseteq \ker\varphi_0$, for $\varphi_0(sx_s - 1) = 0$, and so there is an $R$-algebra map $\widetilde{\varphi}\colon S^{-1}R = R[X]/J \to A$ making the diagram commute. The map $\widetilde{\varphi}$ is the unique such map because $S^{-1}R$ is generated by $\operatorname{im} h \cup \{h(s)^{-1} : s \in S\}$ as an $R$-algebra. •

We now describe the elements in $S^{-1}R$.

**Proposition C-5.15.** *If $S \subseteq R$ is multiplicative, then each $y \in S^{-1}R$ has a (not necessarily unique) factorization $y = h(r)h(s)^{-1}$, where $h\colon R \to S^{-1}R$ is the localization map, $r \in R$, and $s \in S$.*

**Proof.** Define $A = \{y \in S^{-1}R : y = h(r)h(s)^{-1}, r \in R, s \in S\}$. It is easy to check that $A$ is an $R$-subalgebra of $S^{-1}R$ containing $\operatorname{im} h$. Since $\operatorname{im} h \subseteq A$, there is an $R$-algebra map $h'\colon R \to A$ that is obtained from $h$ by changing its target. Consider the diagram




[2] I first saw this expounded in MIT lecture notes of M. Artin.

where $j\colon A \to S^{-1}R$ is the inclusion and $\widetilde{h'}\colon S^{-1}R \to A$ is given by universality (so the top triangle commutes). The lower triangle commutes, because $h(r) = h'(r)$ for all $r \in R$, and so the large triangle commutes: $(j\widetilde{h'})h = h$. But $1_{S^{-1}R}$ also makes this diagram commute, so that uniqueness gives $j\widetilde{h'} = 1_{S^{-1}R}$. By set theory, $j$ is surjective; that is, $S^{-1}R = A$. •

In light of this proposition, the elements of $S^{-1}R$ can be regarded as "fractions" $h(r)h(s)^{-1}$, where $r \in R$ and $s \in S$.

**Notation.** Let $h\colon R \to S^{-1}R$ be the localization map. If $r \in R$ and $s \in S$, define

$$r/s = h(r)h(s)^{-1}.$$

In particular, $r/1 = h(r)$.

Is the localization map $h\colon r \mapsto r/1$ an injection?

**Proposition C-5.16.** *If $S \subseteq R$ is multiplicative and $h\colon R \to S^{-1}R$ is the localization map, then*

$$\ker h = \{r \in R : sr = 0 \text{ for some } s \in S\}.$$

**Proof.** If $sr = 0$, then $0 = h(s)h(r)$ in $S^{-1}R$, so that $0 = h(s)^{-1}h(s)h(r) = h(r)$ ($h(s)$ is a unit). Hence, $r \in \ker h$, and $\{r \in R : sr = 0 \text{ for some } s \in S\} \subseteq \ker h$.

For the reverse inclusion, suppose that $h(r) = 0$ in $S^{-1}R$. Since $S^{-1}R = R[X]/J$, where $J = (sx_s - 1 : s \in S)$, there is an equation $r = \sum_{i=1}^{n} f_i(X)(s_i x_{s_i} - 1)$ in $R[X]$ which involves only finitely many elements $\{s_1, \dots, s_n\} \subseteq S$; let $S_0$ be the submonoid of $S$ they generate (so $S_0$ is multiplicative). If $h_0\colon R \to S_0^{-1}R$ is the localization map, then $r \in \ker h_0$. In fact, if $s = s_1 \cdots s_n$ and $h'\colon R \to \langle s \rangle^{-1}R$ is the localization map (where $\langle s \rangle$ is the multiplicative set $\{s^n : n \geq 0\}$), then every $h'(s_i)$ is invertible, for $s_i^{-1} = s^{-1}s_1 \cdots \widehat{s_i} \cdots s_n$ (omit the factor $s_i$). Now $\langle s \rangle^{-1}R = R[x]/(sx - 1)$, so that $r \in \ker h'$ says that there is $f(x) = \sum_{i=0}^{m} a_i x^i \in R[x]$ with

$$r = f(x)(sx - 1) = \Big(\sum_{i=0}^{m} a_i x^i\Big)(sx - 1) = \sum_{i=0}^{m} \big(sa_i x^{i+1} - a_i x^i\big) \text{ in } R[x].$$

Expanding and equating coefficients of like powers of $x$ gives

$$r = -a_0, \quad sa_0 = a_1, \quad \dots, \quad sa_{m-1} = a_m, \quad sa_m = 0.$$

Hence, $sr = -sa_0 = -a_1$, and, by induction, $s^i r = -a_i$ for all $i$. In particular, $s^m r = -a_m$, and so $s^{m+1}r = -sa_m = 0$, as desired. •

When are two "fractions" $r/s$ and $r'/s'$ equal? The next corollary shows why the equivalence relation in the usual account of localization arises.

**Corollary C-5.17.** *Let $S \subseteq R$ be multiplicative. If both $r/s, r'/s' \in S^{-1}R$, where $s, s' \in S$, then $r/s = r'/s'$ if and only if there exists $s'' \in S$ with $s''(rs' - r's) = 0$ in $R$.*

**Remark.** If $S$ contains no zero-divisors, then $s''(rs' - r's) = 0$ if and only if $rs' - r's = 0$, because $s''$ is a unit, and so $rs' = r's$. ◀

**Proof.** If $r/s = r'/s'$, then multiplying by $ss'$ gives $(rs' - r's)/1 = 0$ in $S^{-1}R$. Hence, $rs' - r's \in \ker h$, and Proposition C-5.16 gives $s'' \in S$ with $s''(rs' - r's) = 0$ in $R$.

Conversely, if $s''(rs' - r's) = 0$ in $R$ for some $s'' \in S$, then $h(s'')h(rs' - r's) = 0$ in $S^{-1}R$. As $h(s'')$ is a unit, we have $h(r)h(s') = h(r')h(s)$; as $h(s)$ and $h(s')$ are units, $h(r)h(s)^{-1} = h(r')h(s')^{-1}$; that is, $r/s = r'/s'$. •

**Corollary C-5.18.** *Let $S \subseteq R$ be multiplicative.*

(i) *If $S$ contains no zero-divisors, then the localization map $h\colon R \to S^{-1}R$ is an injection.*

(ii) *If $R$ is a domain with $Q = \operatorname{Frac}(R)$, then $S^{-1}R \subseteq Q$. Moreover, if $S = R^\times$, then $S^{-1}R = Q$.*

**Proof.**

(i) This follows easily from Proposition C-5.16.

(ii) The localization map $h\colon R \to S^{-1}R$ is an injection, by Proposition C-5.16. The result now follows from Proposition C-5.15. •

If $R$ is a domain and $S \subseteq R$ is multiplicative, then Corollary C-5.18 says that $S^{-1}R$ consists of all elements $a/s \in \operatorname{Frac}(R)$ with $a \in R$ and $s \in S$.

Let us now investigate the ideals in $S^{-1}R$.

**Definition.** If $S \subseteq R$ is multiplicative and $I$ is an ideal in $R$, then we denote the ideal in $S^{-1}R$ generated by $h(I)$ by $S^{-1}I$.

**Example C-5.19.**

(i) If $S \subseteq R$ is multiplicative and $I$ is an ideal in $R$ containing an element $s \in S$ (that is, $I \cap S \neq \varnothing$), then $S^{-1}I$ contains $s/s = 1$, and so $S^{-1}I = S^{-1}R$.

(ii) Let $S$ consist of all the odd integers (that is, $S$ is the complement of the prime ideal $(2)$), let $I = (3)$, and let $I' = (5)$. Then $S^{-1}I = S^{-1}\mathbb{Z} = S^{-1}I'$. Therefore, the function from the ideals in $\mathbb{Z}$ to the ideals in $S^{-1}\mathbb{Z} = \mathbb{Z}_{(2)} = \{a/b \in \mathbb{Q} : b \text{ is odd}\}$, given by $I \mapsto S^{-1}I$, is not injective. ◀

**Corollary C-5.20.** *Let $S \subseteq R$ be multiplicative.*

(i) *Every ideal $J$ in $S^{-1}R$ is of the form $S^{-1}I$ for some ideal $I$ in $R$. In fact, if $R$ is a domain and $I = J \cap R$, then $J = S^{-1}I$; in the general case, if $I = h^{-1}(h(R) \cap J)$, then $J = S^{-1}I$.*

(ii) *Let $I$ be an ideal in $R$. Then $S^{-1}I = S^{-1}R$ if and only if $I \cap S \neq \varnothing$.*

(iii) *If $\mathfrak{q}$ is a prime ideal in $R$ with $\mathfrak{q} \cap S = \varnothing$, then $S^{-1}\mathfrak{q}$ is a prime ideal in $S^{-1}R$.*

(iv) *The function $f\colon \mathfrak{q} \mapsto S^{-1}\mathfrak{q}$ is a bijection from the family of all prime ideals in $R$ disjoint from $S$ to the family of all prime ideals in $S^{-1}R$.*

(v) *If $R$ is noetherian, then $S^{-1}R$ is also noetherian.*

**Proof.**

(i) Let $J = (j_\lambda : \lambda \in \Lambda)$. By Proposition C-5.15, we have $j_\lambda = h(r_\lambda)h(s_\lambda)^{-1}$, where $r_\lambda \in R$ and $s_\lambda \in S$. Define $I$ to be the ideal in $R$ generated by $\{r_\lambda : \lambda \in \Lambda\}$; that is, $I = h^{-1}(h(R) \cap J)$. It is clear that $S^{-1}I = J$; in fact, since all $s_\lambda$ are units in $S^{-1}R$, we have $J = (h(r_\lambda) : \lambda \in \Lambda)$.

(ii) If $s \in I \cap S$, then $s/1 \in S^{-1}I$. But $s/1$ is a unit in $S^{-1}R$, and so $S^{-1}I = S^{-1}R$. Conversely, if $S^{-1}I = S^{-1}R$, then $h(a)h(s)^{-1} = 1$ for some $a \in I$ and $s \in S$. Therefore, $s - a \in \ker h$, and so there is $s'' \in S$ with $s''(s - a) = 0$. Therefore, $s''s = s''a \in I$. Since $S$ is multiplicative, $s''s \in I \cap S$.

(iii) Suppose that $\mathfrak{q}$ is a prime ideal in $R$. First, $S^{-1}\mathfrak{q}$ is a proper ideal, for $\mathfrak{q} \cap S = \varnothing$. If $(a/s)(b/t) = c/u$, where $a, b \in R$, $c \in \mathfrak{q}$, and $s, t, u \in S$, then there is $s'' \in S$ with $s''(uab - stc) = 0$. Hence, $s''uab \in \mathfrak{q}$. Now $s''u \notin \mathfrak{q}$ (because $s''u \in S$ and $S \cap \mathfrak{q} = \varnothing$); hence, $ab \in \mathfrak{q}$ (because $\mathfrak{q}$ is prime). Thus, either $a$ or $b$ lies in $\mathfrak{q}$, and either $a/s$ or $b/t$ lies in $S^{-1}\mathfrak{q}$. Therefore, $S^{-1}\mathfrak{q}$ is a prime ideal.

(iv) Suppose that $\mathfrak{p}$ and $\mathfrak{q}$ are prime ideals in $R$ with $f(\mathfrak{p}) = S^{-1}\mathfrak{p} = S^{-1}\mathfrak{q} = f(\mathfrak{q})$; we may assume that $\mathfrak{p} \cap S = \varnothing = \mathfrak{q} \cap S$. If $a \in \mathfrak{p}$, then there is $b \in \mathfrak{q}$ and $s \in S$ with $a/1 = b/s$. Hence, $sa - b \in \ker h$, where $h$ is the localization map, and so there is $s' \in S$ with $s'sa = s'b \in \mathfrak{q}$. But $s's \in S$, so that $s's \notin \mathfrak{q}$. Since $\mathfrak{q}$ is prime, we have $a \in \mathfrak{q}$; that is, $\mathfrak{p} \subseteq \mathfrak{q}$. The reverse inclusion is proved similarly. Thus, $f$ is injective.

Let $\mathfrak{P}$ be a prime ideal in $S^{-1}R$. By (i), there is some ideal $I$ in $R$ with $\mathfrak{P} = S^{-1}I$. We must show that $I$ can be chosen to be a prime ideal in $R$. Now $h(R) \cap \mathfrak{P}$ is a prime ideal in $h(R)$, and so $\mathfrak{p} = h^{-1}(h(R) \cap \mathfrak{P})$ is a prime ideal in $R$. By (i), $\mathfrak{P} = S^{-1}\mathfrak{p}$, and so $f$ is surjective.

(v) If $J$ is an ideal in $S^{-1}R$, then (i) shows that $J = S^{-1}I$ for some ideal $I$ in $R$. Since $R$ is noetherian, we have $I = (r_1, \ldots, r_n)$, and so $J = (r_1/1, \ldots, r_n/1)$. Hence, every ideal in $S^{-1}R$ is finitely generated, and so $S^{-1}R$ is noetherian. •

**Notation.** If $\mathfrak{p}$ is a prime ideal in a commutative ring $R$ and $S = R - \mathfrak{p}$, then $S^{-1}R$ is denoted by

$$R_\mathfrak{p}.$$

**Example C-5.21.** If $p$ is a nonzero prime in $\mathbb{Z}$, then $\mathfrak{p} = (p)$ is a prime ideal, and $\mathbb{Z}_\mathfrak{p} = \mathbb{Z}_{(p)}$. ◄

**Proposition C-5.22.** *If $R$ is a domain, then $\bigcap_\mathfrak{m} R_\mathfrak{m} = R$, where the intersection is over all the maximal ideals $\mathfrak{m}$ in $R$.*

**Proof.** Since $R$ is a domain, $R_\mathfrak{m} \subseteq \mathrm{Frac}(R)$ for all $\mathfrak{m}$, and so the intersection in the statement is defined. Moreover, it is plain that $R \subseteq R_\mathfrak{m}$ for all $\mathfrak{m}$, so that $R \subseteq \bigcap_m R_\mathfrak{m}$. For the reverse inclusion, let $a \in \bigcap R_\mathfrak{m}$. Consider the colon ideal

$$I = (R : a) = \{r \in R : ra \in R\}.$$

If $I = R$, then $1 \in I$, and $a = 1a \in R$, as desired. If $I$ is a proper ideal, then there exists a maximal ideal $\mathfrak{m}$ with $I \subseteq \mathfrak{m}$. Now $a/1 \in R_{\mathfrak{m}}$, so there is $r \in R$ and $\sigma \notin \mathfrak{m}$ with $a/1 = r/\sigma$; that is, $\sigma a = r \in R$. Hence, $\sigma \in I \subseteq \mathfrak{m}$, contradicting $\sigma \notin \mathfrak{m}$. Therefore, $R = \bigcap_{\mathfrak{m}} R_{\mathfrak{m}}$. •

The next proposition explains why $S^{-1}R$ is called localization.

**Theorem C-5.23.** *If $\mathfrak{p}$ is a prime ideal in a commutative ring $R$, then $R_{\mathfrak{p}}$ is a local ring with unique maximal ideal $\mathfrak{p}R_{\mathfrak{p}} = \{r/s : r \in \mathfrak{p} \text{ and } s \notin \mathfrak{p}\}$.*

**Proof.** If $x \in R_{\mathfrak{p}}$, then $x = r/s$, where $r \in R$ and $s \notin \mathfrak{p}$. If $r \notin \mathfrak{p}$, then $r/s$ is a unit in $R_{\mathfrak{p}}$; that is, all nonunits lie in $\mathfrak{p}R_{\mathfrak{p}}$. Hence, if $I$ is any ideal in $R_{\mathfrak{p}}$ that contains an element $r/s$ with $r \notin \mathfrak{p}$, then $I = R_{\mathfrak{p}}$. It follows that every proper ideal in $R_{\mathfrak{p}}$ is contained in $\mathfrak{p}R_{\mathfrak{p}}$, and so $R_{\mathfrak{p}}$ is a local ring with unique maximal ideal $\mathfrak{p}R_{\mathfrak{p}}$. •

Here is an application of localization. Recall that a prime ideal $\mathfrak{p}$ in a commutative ring $R$ is a ***minimal prime ideal*** if there is no prime ideal strictly contained in it. In a domain, (0) is the unique minimal prime ideal.

**Proposition C-5.24.** *If $\mathfrak{p}$ is a minimal prime ideal in a commutative ring $R$, then every $x \in \mathfrak{p}$ is nilpotent; that is, $x^n = 0$ for some $n = n(x) \geq 1$.*

**Proof.** Let $x \in \mathfrak{p}$ be nonzero. By Corollary C-5.20(iv), there is only one prime ideal in $R_{\mathfrak{p}}$, namely, $\mathfrak{p}R_{\mathfrak{p}}$, and $x/1$ is a nonzero element in it. Indeed, $x$ is nilpotent if and only if $x/1$ is nilpotent, by Proposition C-5.16. Thus, we have normalized the problem; we may now assume that $x \in \mathfrak{p}$ and that $\mathfrak{p}$ is the only prime ideal in $R$. If $x$ is not nilpotent, then $S = \{1, x, x^2, \dots\}$ is multiplicative. By Zorn's Lemma, there exists an ideal $I$ maximal with $I \cap S = \varnothing$. Now we show that $I$ is a prime ideal; that is, $\mathfrak{p} = I$. Suppose $a, b \in R$, $a \notin I$, $b \notin I$, and $ab \in I$. Then $Ra + I$ and $Rb + I$ are both larger than $I$ so each contains an element $S$. If $x^j = ra + z$ and $x^t = r'b + z'$ with $z, z' \in I$, then $x^{j+t} \in rr'ab + I = I$, since $ab \in I$—a contradiction. So $I$ is prime and $I = \mathfrak{p}$. But $x \in S \cap \mathfrak{p} = S \cap I = \varnothing$, a contradiction. Therefore, $x$ is nilpotent. •

The structure of projective modules over a general ring can be quite complicated, but the next proposition shows that projective modules over local rings are free.

**Lemma C-5.25.** *Let $R$ be a local ring with maximal ideal $\mathfrak{m}$. An element $r \in R$ is a unit if and only if $r \notin \mathfrak{m}$.*

**Proof.** It is clear that if $r$ is a unit, then $r \notin \mathfrak{m}$, for $\mathfrak{m}$ is a proper ideal. Conversely, assume that $r$ is not a unit. By Zorn's Lemma, there is a maximal ideal $\mathfrak{m}'$ containing the principal ideal $(r)$. Since $R$ is local, $\mathfrak{m}$ is the only maximal ideal; hence, $\mathfrak{m}' = \mathfrak{m}$ and $r \in \mathfrak{m}$. •

**Proposition C-5.26.** *If $R$ is a local ring, then every finitely generated*[3] *projective $R$-module $B$ is free.*

**Proof.** Let $R$ be a local ring with maximal ideal $\mathfrak{m}$, and let $\{b_1, \ldots, b_n\}$ be a *smallest set of generators* of $B$ in the sense that $B$ cannot be generated by fewer than $n$ elements. Let $F$ be the free $R$-module with basis $x_1, \ldots, x_n$, and define $\varphi\colon F \to B$ by $\varphi(x_i) = b_i$ for all $i$. Thus, there is an exact sequence

$$0 \to K \to F \xrightarrow{\varphi} B \to 0, \tag{1}$$

where $K = \ker \varphi$.

We claim that $K \subseteq \mathfrak{m}F$. If, on the contrary, $K \subsetneq \mathfrak{m}F$, there is an element $y = \sum_{i=1}^{n} r_i x_i \in K$ which is not in $\mathfrak{m}F$, that is, some coefficient, say, $r_1 \notin \mathfrak{m}$. Lemma C-5.25 says that $r_1$ is a unit. Now $y \in K = \ker \varphi$ gives $\sum r_i b_i = 0$. Hence, $b_1 = -r_1^{-1}\big(\sum_{i=2}^{n} r_i b_i\big)$, which implies that $B = \langle b_2, \ldots, b_n \rangle$, contradicting the original generating set being smallest.

Returning to the exact sequence (1), projectivity of $B$ gives $F = K \oplus B'$, where $B'$ is a submodule of $F$ with $B' \cong B$. Hence, $\mathfrak{m}F = \mathfrak{m}K \oplus \mathfrak{m}B'$. Since $\mathfrak{m}K \subseteq K \subseteq \mathfrak{m}F$, Corollary B-2.16 in Part 1 gives

$$K = \mathfrak{m}K \oplus (K \cap \mathfrak{m}B').$$

But $K \cap \mathfrak{m}B' \subseteq K \cap B' = \{0\}$, so that $K = \mathfrak{m}K$. The submodule $K$ is finitely generated, being a direct summand (and hence a homomorphic image) of the finitely generated module $F$, so that Nakayama's Lemma (Corollary C-2.8) gives $K = \{0\}$. Therefore, $\varphi$ is an isomorphism and $B$ is free. •

Having localized a commutative ring, we now localize its modules. Recall a general construction we introduced in Exercise B-4.25 on page 475 in Part 1. Given a ring homomorphism $\varphi\colon R \to R^*$, every $R^*$-module $A^*$ can be viewed as an $R$-module: if $a^* \in A^*$ and $r \in R$, define

$$ra^* = \varphi(r)a^*$$

and denote $A^*$ viewed as an $R$-module in this way by ${}_{\varphi}A^*$. Indeed, $\varphi$ induces a *change of rings functor*; namely, ${}_{\varphi}\square\colon {}_{R^*}\mathbf{Mod} \to {}_{R}\mathbf{Mod}$, which is additive and exact. In particular, the localization map $h\colon R \to S^{-1}R$ allows us to view an $S^{-1}R$-module $M$ as an *induced $R$-module*

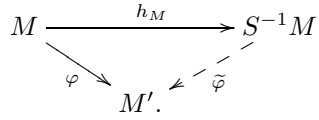
$${}_{h}M,$$

where $rm = h(r)m = (r/1)m$ for all $r \in R$ and $m \in M$.

If $M$ is an $R$-module and $s \in R$, let $\mu_s$ denote the multiplication map $M \to M$ defined by $m \mapsto sm$. Given a subset $S \subseteq R$, note that the map $\mu_s\colon M \to M$ is invertible for every $s \in S$ (that is, every $\mu_s$ is an automorphism) if and only if $M$ is an $S^{-1}R$-module.

[3] It is a theorem of Kaplansky [**123**] that the finiteness hypothesis can be omitted: every projective module over a local ring is free. He even proves freeness when $R$ is a noncommutative local ring.

**Definition.** Let $M$ be an $R$-module and let $S \subseteq R$ be multiplicative. A ***localization*** of $M$ is an ordered pair $(S^{-1}M, h_M)$, where $S^{-1}M$ is an $S^{-1}R$-module and $h_M\colon M \to S^{-1}M$ is an $R$-map (called the ***localization map***), which is a solution to the universal mapping problem: if $M'$ is an $S^{-1}R$-module and $\varphi\colon M \to M'$ is an $R$-map, then there exists a unique $S^{-1}R$-map $\widetilde{\varphi}\colon S^{-1}M \to M'$ with $\widetilde{\varphi}h_M = \varphi$,




The obvious candidate for $(S^{-1}M, h_M)$, namely, $(S^{-1}R \otimes_R M, h \otimes 1_M)$, where $h\colon R \to S^{-1}R$ is the localization map, actually is the localization.

**Proposition C-5.27.** *Let $R$ be a commutative ring, and let $S \subseteq R$ be multiplicative.*

(i) *If $M$ is an $S^{-1}R$-module, then $M$ is naturally isomorphic to $S^{-1}R \otimes_R {}_hM$ via $m \mapsto 1 \otimes m$, where ${}_hM$ is the $R$-module induced from $M$.*

(ii) *If $M$ is an $R$-module, then $(S^{-1}R \otimes_R M, h_M)$ is a localization of $M$, where $h_M\colon M \to S^{-1}R \otimes_R M$ is given by $m \mapsto 1 \otimes m$.*

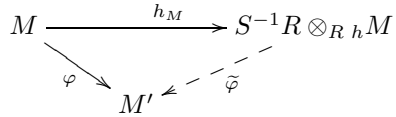
**Proof.**

(i) As above, there is an $R$-module ${}_hM$ with $rm = (r/1)m$ for $r \in R$ and $m \in M$. Define $g\colon M \to S^{-1}R \otimes_R {}_hM$ by $m \mapsto 1 \otimes m$. Now $g$ is an $S^{-1}R$-map: if $s \in S$ and $m \in M$, then

$$g(s^{-1}m) = 1 \otimes s^{-1}m = s^{-1}s \otimes s^{-1}m = s^{-1} \otimes m = s^{-1}g(m).$$

To see that $g$ is an isomorphism, we construct its inverse. Since $M$ is an $S^{-1}R$-module, the function $S^{-1}R \times {}_hM \to M$, defined by $(rs^{-1}, m) \mapsto (rs^{-1})m$, is an $R$-bilinear function, and it induces an $R$-map $S^{-1}R \otimes_R ({}_hM) \to M$, which is obviously inverse to $g$. Proof of naturality of $g$ is left to the reader.

(ii) Consider the diagram




where $M$ is an $R$-module, $h_M\colon m \mapsto 1 \otimes m$, $M'$ is an $S^{-1}R$-module, and $\varphi\colon M \to M'$ is an $R$-map. Since $\varphi$ is merely an $R$-map, we may regard it as a map $\varphi\colon M \to {}_hM'$. By (i), there is an isomorphism $g\colon M' \to S^{-1}R \otimes_R {}_hM'$. Define an $R$-map $\widetilde{\varphi}\colon S^{-1}R \otimes_R M \to M'$ by $g^{-1}(1 \otimes \varphi)$. •

One of the most important properties of $S^{-1}R$ is that it is flat as an $R$-module. To prove this, we first generalize the argument in Proposition C-5.16.

**Proposition C-5.28.** *Let $S \subseteq R$ be multiplicative. If $M$ is an $R$-module and $h_M\colon M \to S^{-1}M$ is the localization map, then*

$$\ker h_M = \{m \in M : sm = 0 \text{ for some } s \in S\}.$$

**Proof.** Denote $\{m \in M : sm = 0 \text{ for some } s \in S\}$ by $K$. If $sm = 0$, for $m \in M$ and $s \in S$, then $h_M(m) = (1/s)h_M(sm) = 0$, and so $K \subseteq \ker h_M$. For the reverse inclusion, proceed as in Proposition C-5.16: if $m \in K$, there is $s \in S$ with $sm = 0$. Reduce to $S = \langle s \rangle$ for some $s \in S$, where $\langle s \rangle = \{s^n : n \geq 0\}$, so that $S^{-1}R = R[x]/(sx - 1)$. Now $R[x] \otimes_R M \cong \bigoplus_i Rx^i \otimes_R M$, because $R[x]$ is the free $R$-module with basis $\{1, x, x^2, \dots\}$. Thus, each element in $R[x] \otimes_R M$ has a unique expression of the form $\sum_i x^i \otimes m_i$, where $m_i \in M$. Hence,

$$\ker h_M = \Big\{ m \in M : 1 \otimes m = (sx - 1) \sum_{i=0}^{n} x^i \otimes m_i \Big\}.$$

The proof now finishes as the proof of Proposition C-5.16. Expanding and equating coefficients gives equations

$$1 \otimes m = -1 \otimes m_0, \quad x \otimes sm_0 = x \otimes m_1, \ \dots,$$
$$x^n \otimes sm_{n-1} = x^n \otimes m_n, \quad x^{n+1} \otimes sm_n = 0.$$

It follows that

$$m = -m_0, \quad sm_0 = m_1, \quad \dots, \quad sm_{n-1} = m_n, \quad sm_n = 0.$$

Hence, $sm = -sm_0 = -m_1$, and, by induction, $s^i m = -m_i$ for all $i$. In particular, $s^n m = -m_n$ and so $s^{n+1}m = -sm_n = 0$ in $M$. Therefore, $\ker h_M \subseteq K$. •

We now generalize Proposition C-5.16 from rings to modules.

**Corollary C-5.29.** *Let $S \subseteq R$ be multiplicative and let $M$ be an $R$-module.*

(i) *Every element $u \in S^{-1}M = S^{-1} \otimes_R M$ has the form $u = s^{-1} \otimes m$ for some $s \in S$ and some $m \in M$.*

(ii) *$s_1^{-1} \otimes m_1 = s_2^{-1} \otimes m_2$ in $S^{-1} \otimes_R M$ if and only if $s(s_2 m_1 - s_1 m_2) = 0$ in $M$ for some $s \in S$.*

**Proof.**

(i) If $u \in S^{-1}R \otimes_R M$, then $u = \sum_i (r_i/s_i) \otimes m_i$, where $r_i \in R$, $s_i \in S$, and $m_i \in M$. If we define $s = \prod s_i$ and $\hat{s}_i = \prod_{j \neq i} s_j$, then

$$u = \sum (1/s_i) r_i \otimes m_i = \sum (\hat{s}_i/s) r_i \otimes m_i$$
$$= (1/s) \sum \hat{s}_i r_i \otimes m_i = (1/s) \otimes \sum \hat{s}_i r_i m_i = (1/s) \otimes m,$$

where $m = \sum \hat{s}_i r_i m_i \in M$.

(ii) If $s \in S$ with $s(s_2 m_1 - s_1 m_2) = 0$ in $M$, then $(s/1)(s_2 \otimes m_1 - s_1 \otimes m_2) = 0$ in $S^{-1}R \otimes_R M$. As $s/1$ is a unit, $s_2 \otimes m_1 - s_1 \otimes m_2 = 0$, and so $s_1^{-1} \otimes m_1 = s_2^{-1} \otimes m_2$.

Conversely, suppose that $s_1^{-1} \otimes m_1 = s_2^{-1} \otimes m_2$ in $S^{-1} \otimes_R M$; then we have $(1/s_1 s_2)(s_2 \otimes m_1 - s_1 \otimes m_2) = 0$. Since $1/s_1 s_2$ is a unit, we have $(s_2 \otimes m_1 - s_1 \otimes m_2) = 0$ and $s_2 m_1 - s_1 m_2 \in \ker h_M$. By Proposition C-5.28, there exists $s \in S$ with $s(s_2 m_1 - s_1 m_2) = 0$ in $M$. •

**Theorem C-5.30.** *If $S \subseteq R$ is multiplicative, then $S^{-1}R$ is a flat $R$-module.*

**Proof.** We must show that if $0 \to A \xrightarrow{f} B$ is exact, then so is

$$0 \to S^{-1}R \otimes_R A \xrightarrow{1\otimes f} S^{-1}R \otimes_R B.$$

By Corollary C-5.29, every $u \in S^{-1}A$ has the form $u = s^{-1} \otimes a$ for some $s \in S$ and $a \in A$. In particular, if $u \in \ker(1\otimes f)$, then $(1\otimes f)(u) = s^{-1}\otimes f(a) = 0$. Multiplying by $s$ gives $1 \otimes f(a) = 0$ in $S^{-1}B$; that is, $f(a) \in \ker h_B$. By Proposition C-5.28, there is $t \in S$ with $0 = tf(a) = f(ta)$. Since $f$ is an injection, $ta \in \ker f = \{0\}$. Hence, $0 = 1 \otimes ta = t(1 \otimes a)$. But $t$ is a unit in $S^{-1}R$, so that $1 \otimes a = 0$ in $S^{-1}A$. Therefore, $1 \otimes f$ is an injection, and $S^{-1}R$ is a flat $R$-module. •

**Corollary C-5.31.** *If $S \subseteq R$ is multiplicative, then localization $M \mapsto S^{-1}M = S^{-1}R \otimes_R M$ defines an exact functor ${}_R\mathbf{Mod} \to {}_{S^{-1}R}\mathbf{Mod}$.*

**Proof.** Localization is the functor $S^{-1}R \otimes_R \square$, and it is exact because $S^{-1}R$ is a flat $R$-module. •

Since tensor product commutes with direct sum, it is clear that if $M$ is a free (or projective) $R$-module, then $S^{-1}M$ is a free (or projective) $S^{-1}R$-module.

**Proposition C-5.32.** *Let $S \subseteq R$ be multiplicative. If $A$ and $B$ are $R$-modules, then there is a natural isomorphism*

$$\varphi\colon S^{-1}(B \otimes_R A) \to S^{-1}B \otimes_{S^{-1}R} S^{-1}A.$$

**Proof.** Every element $u \in S^{-1}(B \otimes_R A)$ has the form $u = s^{-1}m$ for $s \in S$ and $m \in B \otimes_R A$, by Corollary C-5.29; hence, $u = s^{-1}\sum a_i \otimes b_i$:

$$\begin{array}{ccccc} A \times B & \longrightarrow & A \otimes_R B & \longrightarrow & S^{-1}(A \otimes_R B) \\ & \searrow & \downarrow & \swarrow & \\ & & S^{-1}B \otimes_{S^{-1}R} S^{-1}A. & & \end{array}$$

The idea is to define $\varphi(u) = \sum s^{-1}a_i \otimes b_i$ but, as usual with tensor product, the problem is whether obvious maps are well-defined. We suggest that the reader use the universal property of localization to complete the proof. •

**Proposition C-5.33.** *Let $S \subseteq R$ be multiplicative. If $B$ is a flat $R$-module, then $S^{-1}B$ is a flat $S^{-1}R$-module.*

**Proof.** Recall Proposition C-5.27(i): if $A$ is an $S^{-1}R$-module, then $A$ is naturally isomorphic to $S^{-1}R \otimes_R {}_hA$ . The isomorphism of Proposition C-5.32,

$$S^{-1}B \otimes_{S^{-1}R} A = (S^{-1}R \otimes_R B) \otimes_{S^{-1}R} A \to S^{-1}R \otimes_R (B \otimes_R A),$$

can now be used to give a natural isomorphism

$$S^{-1}B \otimes_{S^{-1}R} \square \to (S^{-1}B \otimes_R \square)(B \otimes_R \square)_h\square,$$

where ${}_h\square\colon {}_{S^{-1}R}\mathbf{Mod} \to {}_R\mathbf{Mod}$ is the change of rings functor induced by the localization map $h\colon R \to S^{-1}R$. As each factor is an exact functor, by Exercise B-4.25 on page 475 in Part 1, so is the composite $S^{-1}B \otimes_{S^{-1}R} \square$; that is, $S^{-1}B$ is flat. •

We now investigate localization of injective modules. Preserving injectivity is more subtle than preserving projectives and flats. We will need an analog for Hom of Proposition C-5.32, but the next example shows that the obvious analog is not true without some extra hypothesis.

**Example C-5.34.** If the analog of Proposition C-5.32 for Hom were true, then we would have $S^{-1}\operatorname{Hom}_R(B,A) \cong \operatorname{Hom}_{S^{-1}R}(S^{-1}B, S^{-1}A)$, but such an isomorphism may not exist. If $R = \mathbb{Z}$ and $S^{-1}R = \mathbb{Q}$, then

$$\mathbb{Q} \otimes_{\mathbb{Z}} \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Q},\mathbb{Z}) \ncong \operatorname{Hom}_{\mathbb{Q}}(\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Q}, \mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Z}).$$

The left-hand side is $\{0\}$, because $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Q},\mathbb{Z}) = \{0\}$. On the other hand, the right-hand side is $\operatorname{Hom}_{\mathbb{Q}}(\mathbb{Q},\mathbb{Q}) \cong \mathbb{Q}$. ◀

**Lemma C-5.35.** *Let $S \subseteq R$ be multiplicative, and let $M$ and $A$ be $R$-modules with $A$ finitely presented. Then there is a natural isomorphism*

$$\tau_A \colon S^{-1}\operatorname{Hom}_R(A,M) \to \operatorname{Hom}_{S^{-1}R}(S^{-1}A, S^{-1}M).$$

**Proof.** It suffices to construct natural isomorphisms

$$\varphi_A \colon S^{-1}\operatorname{Hom}_R(A,M) \to \operatorname{Hom}_R(A, S^{-1}M)$$

and

$$\theta_A \colon \operatorname{Hom}_R(A, S^{-1}M) \to \operatorname{Hom}_{S^{-1}R}(S^{-1}A, S^{-1}M),$$

for then we can define $\tau_A = \theta_A\varphi_A$.

If, now, $A$ is a finitely presented $R$-module, then there is an exact sequence

$$R^t \to R^n \to A \to 0. \tag{2}$$

Applying the contravariant functors $\operatorname{Hom}_R(\ \ , M')$ and $\operatorname{Hom}_{S^{-1}R}(\ \ , M')$, where $M' = S^{-1}M$ is first viewed as an $R$-module, gives a commutative diagram with exact rows

$$\begin{array}{ccccccc}
0 & \longrightarrow & \operatorname{Hom}_R(A,M') & \longrightarrow & \operatorname{Hom}_R(R^n,M') & \longrightarrow & \operatorname{Hom}_R(R^t,M') \\
 & & \downarrow{\scriptstyle \theta_A} & & \downarrow{\scriptstyle \theta_{R^n}} & & \downarrow{\scriptstyle \theta_{R^t}} \\
0 & \to & \operatorname{Hom}_{S^{-1}R}(S^{-1}A,M') & \to & \operatorname{Hom}_{S^{-1}R}((S^{-1}R)^n,M') & \to & \operatorname{Hom}_{S^{-1}R}((S^{-1}R)^t,M').
\end{array}$$

Since the vertical maps $\theta_{R^n}$ and $\theta_{R^t}$ are isomorphisms, there is a dotted arrow $\theta_A$ which must be an isomorphism, by Proposition B-1.47 in Part 1. If $\beta \in \operatorname{Hom}_R(A,M)$, then the reader may check that

$$\theta_A(\beta) = \widetilde{\beta} \colon a/s \mapsto \beta(a)/s,$$

from which it follows that the isomorphisms $\theta_A$ are natural.

Construct $\varphi_A \colon S^{-1}\operatorname{Hom}_R(A,M) \to \operatorname{Hom}_R(A, S^{-1}M)$ by defining $\varphi_A \colon g/s \mapsto g_s$, where $g_s(a) = g(a)/s$. Note that $\varphi_A$ is well-defined, for it arises from the $R$-bilinear function $S^{-1}R \times \operatorname{Hom}_R(A,M) \to \operatorname{Hom}_R(A, S^{-1}M)$ given by $(r/s, g) \mapsto rg_s$ (remember that $S^{-1}\operatorname{Hom}_R(A,M) = S^{-1}R \otimes_R \operatorname{Hom}_R(A,M)$, by Proposition C-5.27).

Observe that $\varphi_A$ is an isomorphism when the module $A$ is finitely generated free, and consider the commutative diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & S^{-1}\operatorname{Hom}_R(A,M) & \longrightarrow & S^{-1}\operatorname{Hom}_R(R^n,M) & \longrightarrow & S^{-1}\operatorname{Hom}_R(R^t,M) \\ & & \downarrow{\scriptstyle \varphi_A} & & \downarrow{\scriptstyle \varphi_{R^n}} & & \downarrow{\scriptstyle \varphi_{R^t}} \\ 0 & \longrightarrow & \operatorname{Hom}_R(A,S^{-1}M) & \longrightarrow & \operatorname{Hom}_R(R^n,S^{-1}M) & \longrightarrow & \operatorname{Hom}_R(R^t,S^{-1}M) \end{array}$$

The top row is exact, for it arises from Eq. (2) by first applying the left exact contravariant functor $\operatorname{Hom}_R(\square, M)$, and then applying the exact localization functor. The bottom row is exact, for it arises from Eq. (2) by applying the left exact contravariant functor $\operatorname{Hom}_R(\square, S^{-1}M)$. The Five Lemma shows that $\varphi_A$ is an isomorphism.

Assume first that $A = R^n$ is a finitely generated free $R$-module. If $a_1, \ldots, a_n$ is a basis of $A$, then $a_1/1, \ldots, a_n/1$ is a basis of $S^{-1}A = S^{-1}R \otimes_R R^n$. The map

$$\theta_{R^n}\colon \operatorname{Hom}_R(A, S^{-1}M) \to \operatorname{Hom}_{S^{-1}R}(S^{-1}A, S^{-1}M),$$

given by $f \mapsto \widetilde{f}$, where $\widetilde{f}(a_i/s) = f(a_i)/s$, is easily seen to be a well-defined $R$-isomorphism. •

**Theorem C-5.36.** *Let $R$ be noetherian and let $S \subseteq R$ be multiplicative. If $E$ is an injective $R$-module, then $S^{-1}E$ is an injective $S^{-1}R$-module.*

**Proof.** By Baer's criterion, it suffices to extend any map $I \to S^{-1}E$ to a map $S^{-1}R \to S^{-1}E$, where $I$ is an ideal in $S^{-1}R$; that is, if $i\colon I \to S^{-1}R$ is the inclusion, then the induced map

$$i^*\colon \operatorname{Hom}_{S^{-1}R}(S^{-1}R, S^{-1}E) \to \operatorname{Hom}_{S^{-1}R}(I, S^{-1}E)$$

is a surjection. Now $S^{-1}R$ is noetherian because $R$ is, and so $I$ is finitely generated; say, $I = (r_1/s_1, \ldots, r_n/s_n)$, where $r_i \in R$ and $s_i \in S$. There is an ideal $J$ in $R$; namely, $J = (r_1, \ldots, r_n)$, with $I = S^{-1}J$. Naturality of the isomorphism in Lemma C-5.35 gives a commutative diagram

$$\begin{array}{ccc} S^{-1}\operatorname{Hom}_R(R,E) & \longrightarrow & S^{-1}\operatorname{Hom}_R(J,E) \\ \downarrow & & \downarrow \\ \operatorname{Hom}_{S^{-1}R}(S^{-1}R, S^{-1}E) & \longrightarrow & \operatorname{Hom}_{S^{-1}R}(S^{-1}J, S^{-1}E). \end{array}$$

Now $\operatorname{Hom}_R(R,E) \to \operatorname{Hom}_R(J,E)$ is a surjection, because $E$ is an injective $R$-module, and so $S^{-1} = S^{-1}R \otimes_R \square$ being right exact implies that the top arrow is also a surjection. But the vertical maps are isomorphisms, and so the bottom arrow is a surjection; that is, $S^{-1}E$ is an injective $S^{-1}R$-module. •

**Remark.** Theorem C-5.36 may be false if $R$ is not noetherian. Dade [**49**] showed, for every commutative ring $k$, that if $R = k[X]$, where $X$ is an uncountable set of indeterminates, then there is a multiplicative $S \subseteq R$ and an injective $R$-module $E$ such that $S^{-1}E$ is not an injective $S^{-1}R$-module. If, however, $R = k[X]$, where $k$ is noetherian and $X$ is countable, then $S^{-1}E$ is an injective $S^{-1}R$-module for every injective $R$-module $E$ and every multiplicative $S \subseteq R$. ◄

Here are some globalization tools.

**Notation.** In the special case $S = R - \mathfrak{p}$, where $\mathfrak{p}$ is a prime ideal in $R$, we write

$$S^{-1}M = S^{-1}R \otimes_R M = R_{\mathfrak{p}} \otimes_R M = M_{\mathfrak{p}}.$$

If $f\colon M \to N$ is an $R$-map, write $f_{\mathfrak{p}}\colon M_{\mathfrak{p}} \to N_{\mathfrak{p}}$, where $f_{\mathfrak{p}} = 1_{R_{\mathfrak{p}}} \otimes f$.

We restate Corollary C-5.20(iv) in this notation. The function $f\colon \mathfrak{q} \mapsto \mathfrak{q}_{\mathfrak{p}}$ is a bijection from the family of all prime ideals in $R$ that are contained in $\mathfrak{p}$ to the family of prime ideals in $R_{\mathfrak{p}}$.

**Proposition C-5.37.** *Let $I$ and $J$ be ideals in a domain $R$. If $I_{\mathfrak{m}} = J_{\mathfrak{m}}$ for every maximal ideal $\mathfrak{m}$, then $I = J$.*

**Proof.** Take $b \in J$, and define

$$(I : b) = \{r \in R : rb \in I\}.$$

Let $\mathfrak{m}$ be a maximal ideal in $R$. Since $I_{\mathfrak{m}} = J_{\mathfrak{m}}$, there are $a \in I$ and $s \notin \mathfrak{m}$ with $b/1 = a/s$. As $R$ is a domain, $sb = a \in I$, so that $s \in (I : b)$; but $s \notin \mathfrak{m}$, so that $(I : b) \subsetneq \mathfrak{m}$. Thus, $(I : b)$ cannot be a proper ideal, for it is not contained in any maximal ideal. Therefore, $(I : b) = R$; hence, $1 \in (I : b)$ and $b = 1b \in I$. We have proved that $J \subseteq I$, and the reverse inclusion is proved similarly. •

**Proposition C-5.38.** *Let $R$ be a commutative ring, and let $M, N$ be $R$-modules.*

(i) *If $M_{\mathfrak{m}} = \{0\}$ for every maximal ideal $\mathfrak{m}$, then $M = \{0\}$.*

(ii) *If $f\colon M \to N$ is an $R$-map and $f_{\mathfrak{m}}\colon M_{\mathfrak{m}} \to N_{\mathfrak{m}}$ is an injection for every maximal ideal $\mathfrak{m}$, then $f$ is an injection.*

(iii) *If $f\colon M \to N$ is an $R$-map and $f_{\mathfrak{m}}\colon M_{\mathfrak{m}} \to N_{\mathfrak{m}}$ is a surjection for every maximal ideal $\mathfrak{m}$, then $f$ is a surjection.*

(iv) *If $f\colon M \to N$ is an $R$-map and $f_{\mathfrak{m}}\colon M_{\mathfrak{m}} \to N_{\mathfrak{m}}$ is an isomorphism for every maximal ideal $\mathfrak{m}$, then $f$ is an isomorphism.*

**Proof.**

(i) If $M \neq \{0\}$, then there is $m \in M$ with $m \neq 0$. It follows that the annihilator $I = \{r \in R : rm = 0\}$ is a proper ideal in $R$, for $1 \notin I$, and so there is some maximal ideal $\mathfrak{m}$ containing $I$. Now $1 \otimes m = 0$ in $M_{\mathfrak{m}}$, so that $m \in \ker h_M$. Proposition C-5.28 gives $s \notin \mathfrak{m}$ with $sm = 0$ in $M$. Hence, $s \in I \subseteq \mathfrak{m}$, and this is a contradiction. Therefore, $M = \{0\}$.

(ii) There is an exact sequence $0 \to K \to M \xrightarrow{f} N$, where $K = \ker f$. Since localization is an exact functor, there is an exact sequence

$$0 \to K_{\mathfrak{m}} \to M_{\mathfrak{m}} \xrightarrow{f_{\mathfrak{m}}} N_{\mathfrak{m}}$$

for every maximal ideal $\mathfrak{m}$. By hypothesis, each $f_{\mathfrak{m}}$ is an injection, so that $K_{\mathfrak{m}} = \{0\}$ for all maximal ideals $\mathfrak{m}$. Part (i) now shows that $K = \{0\}$, and so $f$ is an injection.

(iii) There is an exact sequence $M \xrightarrow{f} N \to C \to 0$, where $C = \operatorname{coker} f = N/\operatorname{im} f$. Since tensor product is right exact, $C_{\mathfrak{m}} = \{0\}$ for all maximal ideals $\mathfrak{m}$, and so $C = \{0\}$. But $f$ is surjective if and only if $C = \operatorname{coker} f = \{0\}$.

(iv) This follows at once from parts (ii) and (iii). •

We cannot weaken the hypothesis of Proposition C-5.38(iv) to $M_{\mathfrak{m}} \cong N_{\mathfrak{m}}$ for all maximal ideals $\mathfrak{m}$; we must assume that all the local isomorphisms arise from a given map $f\colon M \to N$. If $G$ is the subgroup of $\mathbb{Q}$ consisting of all $a/b$ with $b$ squarefree, then we saw, in Example C-5.5, that $G_{(p)} \cong \mathbb{Z}_{(p)}$ for all primes $p$, but $G \ncong \mathbb{Z}$.

Localization commutes with Tor, essentially because $S^{-1}R$ is a flat $R$-module.

**Proposition C-5.39.** *If $S$ is a multiplicative subset of a commutative ring $R$, then there are isomorphisms*

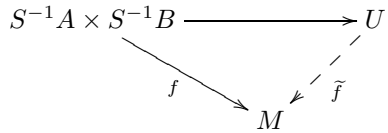
$$S^{-1}\operatorname{Tor}_n^R(A,B) \cong \operatorname{Tor}_n^{S^{-1}R}(S^{-1}A, S^{-1}B)$$

*for all $n \geq 0$ and for all $R$-modules $A$ and $B$.*

**Proof.** First consider the case $n = 0$. For any $R$-module $A$, there is a natural isomorphism

$$\tau_B\colon S^{-1}(A \otimes_R B) \to S^{-1}A \otimes_{S^{-1}R} S^{-1}B,$$

for either side is a solution $U$ of the universal mapping problem




where $M$ is an $(S^{-1}R)$-module, $f$ is $(S^{-1}R)$-bilinear, and $\widetilde{f}$ is an $(S^{-1}R)$-map.

Let $\mathbf{P}_B$ be a deleted projective resolution of $B$. Since the localization functor is exact and preserves projectives, $S^{-1}(\mathbf{P}_B)$ is a deleted projective resolution of $S^{-1}B$. Naturality of the isomorphisms $\tau_A$ gives an isomorphism of complexes

$$S^{-1}(A \otimes_R \mathbf{P}_B) \cong S^{-1}A \otimes_{S^{-1}R} S^{-1}(\mathbf{P}_B),$$

so that their homology groups are isomorphic. Since localization is an exact functor,

$$H_n(S^{-1}(A \otimes_R \mathbf{P}_B)) \cong S^{-1}H_n(A \otimes_R \mathbf{P}_B) \cong S^{-1}\operatorname{Tor}_n^R(A,B).$$

On the other hand, since $S^{-1}(\mathbf{P}_B)$ is a deleted projective resolution of $S^{-1}B$, the definition of Tor gives

$$H_n(S^{-1}A \otimes_{S^{-1}R} S^{-1}(\mathbf{P}_B)) \cong \operatorname{Tor}_n^{S^{-1}R}(S^{-1}A, S^{-1}B). \quad \bullet$$

**Corollary C-5.40.** *Let $A$ be an $R$-module over a commutative ring $R$. If $A_{\mathfrak{m}}$ is a flat $R_{\mathfrak{m}}$-module for every maximal ideal $\mathfrak{m}$, then $A$ is a flat $R$-module.*

**Proof.** The hypothesis, together with Proposition C-3.96, gives $\operatorname{Tor}_n^{R_{\mathfrak{m}}}(A_{\mathfrak{m}}, B_{\mathfrak{m}}) = \{0\}$ for all $n \geq 1$, for every $R$-module $B$, and for every maximal ideal $\mathfrak{m}$. But Proposition C-5.39 gives $\operatorname{Tor}_n^R(A,B)_{\mathfrak{m}} = \{0\}$ for all maximal ideals $\mathfrak{m}$ and all $n \geq 1$.

Finally, Proposition C-5.38 shows that $\operatorname{Tor}_n^R(A,B) = \{0\}$ for all $n \geq 1$. Since this is true for all $R$-modules $B$, we have $A$ flat. •

We must add some hypotheses to get a similar result for Ext (Exercise C-5.24 on page 445).

**Lemma C-5.41.** *If $R$ is a left noetherian ring and $A$ is a finitely generated left $R$-module, then there is a projective resolution $\mathbf{P}$ of $A$ in which each $P_n$ is finitely generated.*

**Proof.** Since $A$ is finitely generated, there exists a finitely generated free left $R$-module $P_0$ and a surjective $R$-map $\varepsilon\colon P_0 \to A$. Since $R$ is left noetherian, $\ker\varepsilon$ is finitely generated, and so there exists a finitely generated free left $R$-module $P_1$ and a surjective $R$-map $d_1\colon P_1 \to \ker\varepsilon$. If we define $D_1\colon P_1 \to P_0$ as the composite $id_1$, where $i\colon \ker\varepsilon \to P_0$ is the inclusion, then there is an exact sequence

$$0 \to \ker D_1 \to P_1 \xrightarrow{D_1} P_0 \xrightarrow{\varepsilon} A \to 0.$$

This construction can be iterated, for $\ker D_1$ is finitely generated, and the proof can be completed by induction. (We remark that we have, in fact, constructed a free resolution of $A$.) •

**Proposition C-5.42.** *Let $S \subseteq R$ be a multiplicative subset of a commutative noetherian ring $R$. If $A$ is a finitely generated $R$-module, then there are isomorphisms*

$$S^{-1}\operatorname{Ext}_R^n(A,B) \cong \operatorname{Ext}_{S^{-1}R}^n(S^{-1}A, S^{-1}B)$$

*for all $n \geq 0$ and for all $R$-modules $B$.*

**Proof.** Since $R$ is noetherian and $A$ is finitely generated, Lemma C-5.41 says there is a projective resolution $\mathbf{P}$ of $A$ each of whose terms is finitely generated. By Lemma C-5.35, there is a natural isomorphism

$$\tau_A\colon S^{-1}\operatorname{Hom}_R(A,B) \to \operatorname{Hom}_{S^{-1}R}(S^{-1}A, S^{-1}B)$$

for every $R$-module $B$ (a finitely generated module over a noetherian ring must be finitely presented). Now $\tau_A$ gives an isomorphism of complexes

$$S^{-1}(\operatorname{Hom}_R(\mathbf{P}_A, B)) \cong \operatorname{Hom}_{S^{-1}R}(S^{-1}(\mathbf{P}_A), S^{-1}B).$$

Taking homology of the left-hand side gives

$$H_n(S^{-1}(\operatorname{Hom}_R(\mathbf{P}_A,B))) \cong S^{-1}H_n(\operatorname{Hom}_R(\mathbf{P}_A,B)) \cong S^{-1}\operatorname{Ext}_R^n(A,B),$$

because localization is an exact functor. On the other hand, homology of the right-hand side is

$$H_n(\operatorname{Hom}_{S^{-1}R}(S^{-1}(\mathbf{P}_A), S^{-1}B)) = \operatorname{Ext}_{S^{-1}R}^n(S^{-1}A, S^{-1}B),$$

because $S^{-1}(\mathbf{P}_A)$ is an $(S^{-1}R)$-projective resolution of $S^{-1}A$. •

**Remark.** An alternative proof of Proposition C-5.42 can be given using a deleted injective resolution $\mathbf{E}_B$ in the second variable. We must still assume that $A$ is finitely generated, in order to use Lemma C-5.35, but we can now use the fact that localization preserves injectives when $R$ is noetherian. ◄

**Corollary C-5.43.** *Let $A$ be a finitely generated $R$-module over a commutative noetherian ring $R$. Then $A_{\mathfrak{m}}$ is a projective $R_{\mathfrak{m}}$-module for every maximal ideal $\mathfrak{m}$ if and only if $A$ is a projective $R$-module.*

**Proof.** Sufficiency follows from localization preserving direct sum, and necessity follows from Proposition C-5.42: for every $R$-module $B$ and maximal ideal $\mathfrak{m}$, we have

$$\operatorname{Ext}^1_R(A,B)_{\mathfrak{m}} \cong \operatorname{Ext}^1_{R_{\mathfrak{m}}}(A_{\mathfrak{m}}, B_{\mathfrak{m}}) = \{0\},$$

because $A_{\mathfrak{m}}$ is projective. By Proposition C-5.38, $\operatorname{Ext}^1_R(A,B) = \{0\}$, which says that $A$ is projective. •

**Example C-5.44.** Let $R$ be a Dedekind ring which is not a PID (Dedekind rings are defined in the next section), and let $\mathfrak{p}$ be a nonzero prime ideal in $R$. Then we shall see, in Proposition C-5.86, that $R_{\mathfrak{p}}$ is a local PID. Hence, if $P$ is a projective $R$-module, then $P_{\mathfrak{p}}$ is a projective $R_{\mathfrak{p}}$-module, and so it is free, by Proposition C-5.26. In particular, if $\mathfrak{b}$ is a nonprincipal ideal in $R$, then $\mathfrak{b}$ is not free even though all its localizations are free. ◄

## Exercises

**C-5.9.** If $R$ is a domain with $Q = \operatorname{Frac}(R)$, determine whether every $R$-subalgebra $A$ of $Q$ is a localization of $R$. If true, prove it; if false, give a counterexample.

**C-5.10.** Prove that every localization of a PID is a PID. Conclude that if $\mathfrak{p}$ is a nonzero prime ideal in a PID $R$, then $R_{\mathfrak{p}}$ is a DVR.

**C-5.11.** If $R$ is a Boolean ring and $\mathfrak{m}$ is a maximal ideal in $R$, prove that $R_{\mathfrak{m}}$ is a field.

**C-5.12.** Let $A$ be an $R$-algebra, and let $N$ be a finitely presented $R$-module. For every $A$-module $M$, prove that

$$\theta\colon \operatorname{Hom}_R(N,M) \to \operatorname{Hom}_A(N \otimes_R A, M),$$

given by $\theta\colon f \mapsto \widetilde{f}$, is a natural isomorphism, where $\widetilde{f}(n \otimes 1) = f(n)$ for all $n \in N$.

**C-5.13.** Let $B$ be a flat $R$-module, and let $N$ be a finitely presented $R$-module. For every $R$-module $M$, prove that

$$\psi\colon B \otimes_R \operatorname{Hom}_R(N,M) \to \operatorname{Hom}_R(N, M \otimes_R B),$$

given by $b \otimes g \mapsto g_b$, is a natural isomorphism, where $g_b(n) = g(n) \otimes b$ for all $n \in N$.

∗ **C-5.14.** Let $S \subseteq R$ be multiplicative, and let $I$ and $J$ be ideals in $R$.

(i) Prove that $S^{-1}(IJ) = (S^{-1}I)(S^{-1}J)$.

(ii) Prove that $S^{-1}(I : J) = (S^{-1}I : S^{-1}J)$.

**C-5.15.** Recall that if $k$ is a commutative ring, then the ring of *Laurent polynomials* over $k$ is the subring of rational functions

$$k[x, x^{-1}] = \Big\{ \sum_{i=m}^{n} a_i x^i \in k(x) : m \le n \text{ and } m, n \in \mathbb{Z} \Big\}.$$

Prove that if $k$ is noetherian, then $k[x, x^{-1}]$ is noetherian.

**Hint.** See Exercise B-1.7 on page 281 in Part 1, and use Corollary C-5.20.

**C-5.16.** A ***valuation ring*** is a domain $R$ such that, for all $a, b \in R$, either $a \mid b$ or $b \mid a$.

(i) Prove that every DVR is a valuation ring.

(ii) Let $R$ be a domain with $F = \mathrm{Frac}(R)$. Prove that $R$ is a valuation ring if and only if $a \in R$ or $a^{-1} \in R$ for each nonzero $a \in F$.

**C-5.17.** (i) Prove that every finitely generated ideal in a valuation ring is principal.

(ii) Prove that every finitely generated ideal in a valuation ring is projective.

∗ **C-5.18.** An abelian group $\Gamma$ is ***ordered*** if it is a partially ordered set in which $a+b \le a'+b'$ whenever $a \le a'$ and $b \le b'$; call $\Gamma$ a ***totally ordered abelian group*** if the partial order is a chain. A ***valuation*** on a field $k$ is a function $v\colon k^\times \to \Gamma$, where $\Gamma$ is a totally ordered abelian group, such that

$$v(ab) = v(a) + v(b),$$
$$v(a+b) \ge \min\{v(a), v(b)\}.$$

(i) If $a/b \in \mathbb{Q}$ is nonzero, write $a = p^m a'$ and $b = p^n b'$, where $m, n \ge 0$ and $(a', p) = 1 = (b', p)$. Prove that $v\colon \mathbb{Q}^\times \to \mathbb{Z}$, defined by $v(a/b) = m - n$, is a valuation.

(ii) If $v\colon k^\times \to \Gamma$ is a valuation on a field $k$, define $R = \{0\} \cup \{a \in k^\times : v(a) \ge 0\}$. Prove that $R$ is a valuation ring. (Every valuation ring arises in this way from a suitable valuation on its fraction field. Moreover, the valuation ring is discrete when the totally ordered abelian group $\Gamma$ is isomorphic to $\mathbb{Z}$.)

(iii) Prove that $a \in R$ is a unit if and only if $v(a) = 0$.

(iv) Prove that every valuation ring is a (not necessarily noetherian) local ring.
**Hint.** Show that $\mathfrak{m} = \{a \in R : v(a) > 0\}$ is the unique maximal ideal in $R$.

**C-5.19.** Let $\Gamma$ be a totally ordered abelian group and let $k$ be a field. Define $k[\Gamma]$ to be the group ring (consisting of all functions $f\colon \Gamma \to k$ almost all of whose values are 0). As usual, if $f(\gamma) = r_\gamma$, we denote $f$ by $\sum_{\gamma\in\Gamma} r_\gamma \gamma$.

(i) Define the ***degree*** of $f = \sum_{\gamma\in\Gamma} r_\gamma\gamma$ to be $\alpha$ if $\alpha$ is the largest index $\gamma$ with $r_\gamma \ne 0$. Prove that $k[\Gamma]$ is a valuation ring, where $v(f)$ is the degree of $f$.

(ii) Give an example of a nonnoetherian valuation ring.

∗ **C-5.20.** A multiplicative subset $S$ of a commutative ring $R$ is ***saturated*** if $ab \in S$ implies $a \in S$ and $b \in S$.

(i) Prove that $U(R)$, the set of all units in $R$, is a saturated subset of $R$.

(ii) An element $r \in R$ is a ***zero-divisor*** on an $R$-module $A$ if there is some nonzero $a \in A$ with $ra = 0$. Prove that $\mathrm{Zer}(A)$, the set of all zero-divisors on an $R$-module $A$, is a saturated subset of $R$.

(iii) If $S \subseteq R$ is multiplicative, prove that there exists a unique smallest saturated subset $S'$ containing $S$ (called the ***saturation*** of $S$) and that $(S')^{-1}R \cong S^{-1}R$.

(iv) Prove that a multiplicative subset $S$ is saturated if and only if its complement $R - S$ is a union of prime ideals.

**C-5.21.** Let $S$ be a multiplicative subset of a commutative ring $R$, and let $M$ be a finitely generated $R$-module. Prove that $S^{-1}M = \{0\}$ if and only if there is $s \in S$ with $sM = \{0\}$.

**C-5.22.** Let $S$ be a multiplicative subset of a commutative ring $R$, and let $A$ be an $R$-module.

(i) If $A$ is finitely generated, prove that $S^{-1}A$ is a finitely generated $(S^{-1}R)$-module.

(ii) If $A$ is finitely presented, prove that $S^{-1}A$ is a finitely presented $(S^{-1}R)$-module.

**C-5.23.** If $\mathfrak{p}$ is a nonzero prime ideal in a commutative ring $R$ and $A$ is a projective $R$-module, prove that $A_{\mathfrak{p}}$ is a free $R_{\mathfrak{p}}$-module.

∗ **C-5.24.** (i) Give an example of an abelian group $B$ for which $\operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Q}, B) \neq \{0\}$.

(ii) Prove that $\mathbb{Q} \otimes_{\mathbb{Z}} \operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Q}, B) \neq \{0\}$ for the abelian group $B$ in part (i).

(iii) Prove that Proposition C-5.42 may be false if $R$ is noetherian but $A$ is not finitely generated.

∗ **C-5.25.** Let $R$ be a commutative $k$-algebra, where $k$ is a commutative ring, and let $M$ be a $k$-module. Prove, for all $n \geq 0$, that

$$R \otimes_k \bigwedge^n(M) \cong \bigwedge^n(R \otimes_k M)$$

(of course, $\bigwedge^n(R \otimes_k M)$ means the $n$th exterior power of the $R$-module $R \otimes_k M$). Conclude, for all maximal ideals $\mathfrak{m}$ in $k$, that

$$\left(\bigwedge^n(M)\right)_{\mathfrak{m}} \cong \bigwedge^n(M_{\mathfrak{m}}).$$

**Hint.** Show that $R \otimes_k \bigwedge^n(M)$ is a solution to the universal mapping problem for alternating $n$-multilinear $R$-functions.

**C-5.26.** Let $R$ be a commutative noetherian ring. If $A$ and $B$ are finitely generated $R$-modules, prove that $\operatorname{Tor}^R_n(A, B)$ and $\operatorname{Ext}^n_R(A, B)$ are finitely generated $R$-modules for all $n \geq 0$.

## C-5.3. Dedekind Rings

A ***Pythagorean triple*** is a triple $(a, b, c)$ of positive integers such that $a^2 + b^2 = c^2$; examples are $(3, 4, 5)$, $(5, 12, 13)$, and $(7, 24, 25)$. All Pythagorean triples were classified by Diophantus, ca. 250 AD. Fermat proved that there do not exist positive integers $(a, b, c)$ with $a^4 + b^4 = c^4$ and, in 1637, he wrote in the margin of his copy of a book by Diophantus that he had a wonderful proof that there are no positive integers $(a, b, c)$ with $a^n + b^n = c^n$ for any $n > 2$. Fermat's proof was never found, and his remark (that was merely a note to himself) became known only several years after his death, when Fermat's son published a new edition of Diophantus in 1670 containing his father's notes. There were other notes of Fermat, many of them true, some of them false, and this statement, the only one unresolved by 1800, was called ***Fermat's Last Theorem***, perhaps in jest. It remained one of the outstanding challenges in number theory until 1995, when Wiles proved Fermat's Last Theorem.

Every positive integer $n > 2$ is a multiple of 4 or of some odd prime $p$. Thus, if there do not exist positive integers $(a, b, c)$ with $a^p + b^p = c^p$ for every odd prime $p$, then Fermat's Last Theorem is true (if $n = pm$, then $a^n + b^n = c^n$ implies $(a^m)^p + (b^m)^p = (c^m)^p$). Over the centuries, there were many attempts to prove

it. For example, Euler published a proof (with gaps, later corrected) for the case $n = 3$, Dirichlet published a proof (with gaps, later corrected) for the case $n = 5$, and Lamé published a correct proof for the case $n = 7$.

The first major progress (not dealing only with particular primes $p$) was due to Kummer, in the middle of the nineteenth century. If $a^p + b^p = c^p$, where $p$ is an odd prime, then a natural starting point of investigation is the identity

$$c^p = a^p + b^p = (a + b)(a + \zeta b)(a + \zeta^2 b) \cdots (a + \zeta^{p-1} b),$$

where $\zeta = \zeta_p$ is a primitive $p$th root of unity (perhaps this idea occurred to Fermat). Kummer proved that if $\mathbb{Z}[\zeta_p]$ is a UFD, where $\mathbb{Z}[\zeta_p] = \{f(\zeta_p) : f(x) \in \mathbb{Z}[x]\}$, then there do not exist positive integers $a, b, c$ with $a^p + b^p = c^p$. On the other hand, he also showed that there do exist primes $p$ for which $\mathbb{Z}[\zeta_p]$ is not a UFD. To restore unique factorization, he invented "ideal numbers" that he adjoined to $\mathbb{Z}[\zeta_p]$. Later, Dedekind recast Kummer's ideal numbers into our present notion of ideal. Thus, Fermat's Last Theorem has served as a catalyst in the development of both modern algebra and algebraic number theory. *Dedekind rings* are the appropriate generalization of rings like $\mathbb{Z}[\zeta_p]$, and we will study them in this section.

## ▪ Integrality

The notion of algebraic integer is a special case of the notion of integral element.

**Definition.** A ***ring extension***[4] $R^*/R$ is a commutative ring $R^*$ containing $R$ as a subring. Let $R^*/R$ be a ring extension. Then an element $a \in R^*$ is ***integral*** over $R$ if it is a root of a monic polynomial in $R[x]$. A ring extension $R^*/R$ is an ***integral extension*** if every $a \in R^*$ is integral over $R$.

**Example C-5.45.** The ***Noether Normalization Theorem*** is often used to prove the Nullstellensatz (see (Matsumura [**150**], p. 262)). It says that if $A$ is a finitely generated $k$-algebra over a field $k$, then there exist algebraically independent elements $a_1, \ldots, a_n$ in $A$ so that $A$ is integral over $k[a_1, \ldots, a_n]$. ◀

Recall that a complex number is an algebraic integer if it is a root of a monic polynomial in $\mathbb{Z}[x]$, so that algebraic integers are integral over $\mathbb{Z}$. The reader should compare the next lemma with Proposition C-2.93.

**Lemma C-5.46.** *If $R^*/R$ is a ring extension, then the following conditions on a nonzero element $u \in R^*$ are equivalent:*

(i) *$u$ is integral over $R$.*

(ii) *There is a finitely generated $R$-submodule $B$ of $R^*$ with $uB \subseteq B$.*

(iii) *There is a finitely generated faithful $R$-submodule $B$ of $R^*$ with $uB \subseteq B$; that is, if $dB = \{0\}$ for some $d \in R$, then $d = 0$.*

[4] We use the same notation for ring extensions as we do for field extensions.

**Proof.**

(i) $\Rightarrow$ (ii). If $u$ is integral over $R$, there is a monic polynomial $f(x) \in R[x]$ with $f(u) = 0$; that is, there are $r_i \in R$ with $u^n = \sum_{i=0}^{n-1} r_i u^i$. Define $B = \langle 1, u, u^2, \ldots, u^{n-1} \rangle$. It is clear that $uB \subseteq B$.

(ii) $\Rightarrow$ (iii). If $B = \langle b_1, \ldots, b_m \rangle$ is a finitely generated $R$-submodule of $R^*$ with $uB \subseteq B$, define $B' = \langle 1, b_1, \ldots, b_m \rangle$. Now $B'$ is finitely generated, faithful (because $1 \in B'$), and $uB' \subseteq B'$.

(iii) $\Rightarrow$ (i). Suppose there is a faithful $R$-submodule of $R^*$ with $uB \subseteq B$, say, $B = \langle b_i, \ldots, b_n \rangle$. There is a system of $n$ equations $ub_i = \sum_{j=1}^{n} p_{ij} b_j$ with $p_{ij} \in R$. If $P = [p_{ij}]$ and if $X = (b_1, \ldots, b_n)^\top$ is an $n \times 1$ column vector, then the $n \times n$ system can be rewritten in matrix notation: $(uI - P)X = 0$. Now $0 = \big(\mathrm{adj}(uI - P)\big)(uI - P)X = dX$, where $d = \det(uI - P)$, by Corollary B-5.53 in Part 1. Since $dX = 0$, we have $db_i = 0$ for all $i$, and so $dB = \{0\}$. Therefore, $d = 0$, because $B$ is faithful. On the other hand, Corollary B-5.47 in Part 1 gives $d = f(u)$, where $f(x) \in R[x]$ is a monic polynomial of degree $n$; hence, $u$ is integral over $R$. •

Being an integral extension is transitive.

**Proposition C-5.47.** *If $T \subseteq S \subseteq R$ are commutative rings with $S$ integral over $T$ and $R$ integral over $S$, then $R$ is integral over $T$.*

**Proof.** If $r \in R$, there is an equation $r^n + s_{n-1} r^{n-1} + \cdots + r_0 = 0$, where $s_i \in S$ for all $i$. By Lemma C-5.46, the subring $S' = T[s_{n-1}, \ldots, s_0]$ is a finitely generated $T$-module. But $r$ is integral over $S'$, so that the ring $S'[r]$ is a finitely generated $S'$-module. Therefore, $S'[r]$ is a finitely generated $T$-module, and so $r$ is integral over $T$. •

**Proposition C-5.48.** *Let $E/R$ be a ring extension.*

(i) *If $u, v \in E$ are integral over $R$, then both $uv$ and $u + v$ are integral over $R$.*

(ii) *The commutative ring $\mathcal{O}_{E/R}$, defined by*

$$\mathcal{O}_{E/R} = \{u \in E : u \text{ is integral over } R\},$$

*is an $R$-subalgebra of $E$.*

**Proof.**

(i) If $u$ and $v$ are integral over $R$, then Lemma C-5.46(ii) says that there are $R$-submodules $B = \langle b_1, \ldots, b_n \rangle$ and $C = \langle c_1, \ldots, c_m \rangle$ of $E$ with $uB \subseteq B$ and $vC \subseteq C$; that is, $ub_i \in B$ for all $i$ and $vc_j \in C$ for all $j$. Define $BC$ to be the $R$-submodule of $E$ generated by all $b_i c_j$; of course, $BC$ is finitely generated. Now $uvBC \subseteq BC$, for $uvb_ic_j = (ub_i)(vc_j)$ is an $R$-linear combination of $b_k c_\ell$'s, and so $uv$ is integral over $R$. Similarly, $u + v$ is integral over $R$, for $(u + v)b_ic_j = (ub_i)c_j + (vc_j)b_i \in BC$.

(ii) Part (i) shows that $\mathcal{O}_{E/R}$ is closed under multiplication and addition. Now $R \subseteq \mathcal{O}_E$, for if $r \in R$, then $r$ is a root of $x - r$. It follows that $1 \in \mathcal{O}_{E/R}$ and that $\mathcal{O}_{E/R}$ is an $R$-subalgebra of $E$. •

Here is a second proof of Proposition C-5.48(i) for a *domain* $E$ using tensor products and linear algebra. Let $f(x) \in R[x]$ be the minimal polynomial of $u$, let $A$ be the companion matrix of $f(x)$, and let $y$ be an eigenvector (over the algebraic closure of $\mathrm{Frac}(E)$): $Ay = uy$. Let $g(x)$ be the minimal polynomial of $v$, let $B$ be the companion matrix of $g(x)$, and let $Bz = vz$. Now

$$(A \otimes B)(y \otimes z) = Ay \otimes Bz = uy \otimes vz = uv(y \otimes z).$$

Therefore, $uv$ is an eigenvalue of $A \otimes B$; that is, $uv$ is a root of the monic polynomial $\det(xI - A \otimes B)$, which lies in $R[x]$ because both $A$ and $B$ have all their entries in $R$. Therefore, $uv$ is integral over $R$. Similarly, the equation

$$(A \otimes I + I \otimes B)(y \otimes z) = Ay \otimes z + y \otimes Bz = (u + v)y \otimes z$$

shows that $u + v$ is integral over $R$.

**Definition.** Let $E/R$ be a ring extension. The $R$-subalgebra $\mathcal{O}_{E/R}$ of $E$, consisting of all those elements integral over $R$, is called the ***integral closure*** of $R$ in $E$. If $\mathcal{O}_{E/R} = R$, then $R$ is called ***integrally closed in*** $E$. If $R$ is a domain and $R$ is integrally closed in $F = \mathrm{Frac}(R)$, that is, $\mathcal{O}_{F/R} = R$, then $R$ is called ***integrally closed***.

Thus, $R$ is integrally closed if, whenever $\alpha \in \mathrm{Frac}(R)$ is integral over $R$, we have $\alpha \in R$.

**Example C-5.49.** The ring $\mathcal{O}_{\mathbb{Q}/\mathbb{Z}} = \mathbb{Z}$, for if a rational number $a$ is a root of a monic polynomial in $\mathbb{Z}[x]$, then Theorem A-3.101 in Part 1 shows that $a \in \mathbb{Z}$. Hence, $\mathbb{Z}$ is integrally closed. ◀

**Proposition C-5.50.** *Every* UFD $R$ *is integrally closed. In particular, every* PID *is integrally closed.*

**Proof.** Let $F = \mathrm{Frac}(R)$, and suppose that $u \in F$ is integral over $R$. Thus, there is an equation

$$u^n + r_{n-1}u^{n-1} + \cdots + r_1 u + r_0 = 0,$$

where $r_i \in R$. We may write $u = b/c$, where $b, c \in R$ and $(b, c) = 1$ (gcd's exist because $R$ is a UFD, and so every fraction can be put in lowest terms). Substituting and clearing denominators, we obtain

$$b^n + r_{n-1}b^{n-1}c + \cdots + r_1 bc^{n-1} + r_0 c^n = 0.$$

Hence, $b^n = -c\big(r_{n-1}b^{n-1} + \cdots + r_1 bc^{n-2} + r_0 c^{n-1}\big)$, so that $c \mid b^n$ in $R$. But $(b, c) = 1$ implies $(b^n, c) = 1$, so that $c$ must be a unit in $R$; that is, $c^{-1} \in R$. Therefore, $u = b/c = bc^{-1} \in R$, and so $R$ is integrally closed. •

We now understand Example A-3.129 in Part 1. If $k$ is a field, the subring $R$ of $k[x]$, consisting of all polynomials $f(x) \in k[x]$ having no linear term, is not integrally closed. It is easy to check that $\mathrm{Frac}(R) = k(x)$, since $x = x^3/x^2 \in \mathrm{Frac}(R)$. But $x \in k(x)$ is a root of the monic polynomial $t^2 - x^2 \in R[t]$, and $x \notin R$. Hence, $R$ is not a UFD.

**Definition.** An ***algebraic number field*** is a finite field extension of $\mathbb{Q}$. If $E$ is an algebraic number field, then $\mathcal{O}_{E/\mathbb{Z}}$ is usually denoted by $\mathcal{O}_E$ instead of by $\mathcal{O}_{E/\mathbb{Z}}$, and it is called the ***ring of integers*** in $E$.

Because of this new use of the word *integers*, algebraic number theorists often speak of the ring of ***rational integers*** when referring to $\mathbb{Z}$.

**Proposition C-5.51.** *Let $E$ be an algebraic number field and let $\mathcal{O}_E$ be its ring of integers.*

(i) *If $\alpha \in E$, then there is a nonzero integer $m$ with $m\alpha \in \mathcal{O}_E$.*

(ii) $\text{Frac}(\mathcal{O}_E) = E$.

(iii) *$\mathcal{O}_E$ is integrally closed.*

**Proof.**

(i) If $\alpha \in E$, then there is a monic polynomial $f(x) \in \mathbb{Q}[x]$ with $f(\alpha) = 0$. Clearing denominators gives an integer $m$ with

$$m\alpha^n + c_{n-1}\alpha^{n-1} + c_{n-2}\alpha^{n-2} + \cdots + c_1\alpha + c_0 = 0,$$

where all $c_i \in \mathbb{Z}$. Multiplying by $m^{n-1}$ gives

$$(m\alpha)^n + c_{n-1}(m\alpha)^{n-1} + mc_{n-2}(m\alpha)^{n-2} + \cdots + c_1m^{n-2}(m\alpha) + m^{n-1}c_0 = 0.$$

Thus, $m\alpha \in \mathcal{O}_E$.

(ii) It suffices to show that if $\alpha \in E$, then there are $a, b \in \mathcal{O}_E$ with $\alpha = a/b$. But $m\alpha \in \mathcal{O}_E$ (by part (i)), $m \in \mathbb{Z} \subseteq \mathcal{O}_E$, and $\alpha = (m\alpha)/m$.

(iii) Suppose that $\alpha \in \text{Frac}(\mathcal{O}_E) = E$ is integral over $\mathcal{O}_E$. By transitivity of integral extensions, Proposition C-5.47, we have $\alpha$ integral over $\mathbb{Z}$. But this means that $\alpha \in \mathcal{O}_E$, which is, by definition, the set of all those elements in $E$ that are integral over $\mathbb{Z}$. Therefore, $\mathcal{O}_E$ is integrally closed. •

**Example C-5.52.** We shall see, in Proposition C-5.65, that if $E = \mathbb{Q}(i)$, then $\mathcal{O}_E = \mathbb{Z}[i]$, the Gaussian integers. Now $\mathbb{Z}[i]$ is a PID, because it is a Euclidean ring, and hence it is a UFD. The generalization of this example which replaces $\mathbb{Q}(i)$ by an algebraic number field $E$ is more subtle. It is true that $\mathcal{O}_E$ is integrally closed, but it may not be true that the elements of $\mathcal{O}_E$ are $\mathbb{Z}$-linear combinations of powers of $\alpha$. Moreover, the rings $\mathcal{O}_E$ may not be UFDs. We will investigate rings of integers at the end of this section. ◄

Given a ring extension $R^*/R$, what is the relation between ideals in $R^*$ and ideals in $R$?

**Definition.** Let $R^*/R$ be a ring extension. If $I$ is an ideal in $R$, define its ***extension*** $I^e$ to be $R^*I$, the ideal in $R^*$ generated by $I$. If $I^*$ is an ideal in $R^*$, define its ***contraction*** $I^{*\,c} = R \cap I^*$.

**Remark.** The definition can be generalized. Let $h\colon R \to R^*$ be a ring homomorphism, where $R$ and $R^*$ are any two commutative rings. Define the extension of an

ideal $I$ in $R$ to be the ideal in $R^*$ generated by $h(I)$; define the contraction of an ideal $I^*$ in $R^*$ to be $h^{-1}(I^*)$. If $R^*/R$ is a ring extension, then taking $h\colon R \to R^*$ to be the inclusion gives the definition above. Another interesting instance is the localization map $h\colon R \to S^{-1}R$. ◀

**Example C-5.53.**

(i) It is easy to see that if $R^*/R$ is a ring extension and $\mathfrak{p}^*$ is a prime ideal in $R^*$, then its contraction $\mathfrak{p}^* \cap R$ is also a prime ideal: if $a, b \in R$ and $ab \in \mathfrak{p}^* \cap R \subseteq \mathfrak{p}^*$, then $\mathfrak{p}^*$ prime gives $a \in \mathfrak{p}^*$ or $b \in \mathfrak{p}^*$; as $a, b \in R$, either $a \in \mathfrak{p}^* \cap R$ or $b \in \mathfrak{p}^* \cap R$. Thus, contraction defines a function $c\colon \operatorname{Spec}(R^*) \to \operatorname{Spec}(R)$.

(ii) By (i), contraction $\mathfrak{p}^* \mapsto \mathfrak{p}^* \cap R$ induces a function $c\colon \operatorname{Spec}(R^*) \to \operatorname{Spec}(R)$; in general, this contraction function is neither an injection nor a surjection. For example, $c\colon \operatorname{Spec}(\mathbb{Q}) \to \operatorname{Spec}(\mathbb{Z})$ is not surjective, while $c\colon \operatorname{Spec}(\mathbb{Q}[x]) \to \operatorname{Spec}(\mathbb{Q})$ is not injective.

(iii) The contraction of a maximal ideal, though necessarily prime, need not be maximal. For example, if $R^*$ is a field, then $(0)^*$ is a maximal ideal in $R^*$, but if $R$ is not a field, then the contraction of $(0)^*$, namely, $(0)$, is not a maximal ideal in $R$. ◀

**Example C-5.54.**

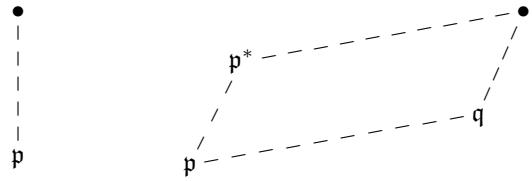
(i) Let $\mathcal{I}(R)$ denote the family of all the ideals in a commutative ring $R$. Extension defines a function $e\colon \mathcal{I}(R) \to \mathcal{I}(R^*)$; in general, it is neither injective nor surjective. If $R^*$ is a field and $R$ is not a field, then $e\colon \mathcal{I}(R) \to \mathcal{I}(R^*)$ is not injective. If $R$ is a field and $R^*$ is not a field, then $e\colon \mathcal{I}(R) \to \mathcal{I}(R^*)$ is not surjective.

(ii) If $R^*/R$ is a ring extension and $\mathfrak{p}$ is a prime ideal in $R$, then its extension $R^*\mathfrak{p}$ need not be a prime ideal. Observe first that if $(a) = Ra$ is a principal ideal in $R$, then its extension is the principal ideal $R^*a$ in $R^*$ generated by $a$. Now let $R = \mathbb{R}[x]$ and $R^* = \mathbb{C}[x]$. The ideal $(x^2+1)$ is prime, because $x^2+1$ is irreducible in $\mathbb{R}[x]$, but its extension is not a prime ideal because $x^2+1$ factors in $\mathbb{C}[x]$. ◀

There are various elementary properties of extension and contraction, such as $I^{*\,ce} \subseteq I^*$ and $I^{ec} \supseteq I$; they are collected in Exercise C-5.29 on page 454.

Is there a reasonable condition on a ring extension $R^*/R$ that will give a good relationship between prime ideals in $R$ and prime ideals in $R^*$? This question was posed and answered by Cohen and Seidenberg. We say that a ring extension $R^*/R$ satisfies ***Lying Over*** if, for every prime ideal $\mathfrak{p}$ in $R$, there exists a prime ideal $\mathfrak{p}^*$ in $R^*$ which contracts to $\mathfrak{p}$; that is, $\mathfrak{p}^* \cap R = \mathfrak{p}$. We say that a ring extension

$R^*/R$ satisfies ***Going Up*** if whenever $\mathfrak{p} \subseteq \mathfrak{q}$ are prime ideals in $R$ and $\mathfrak{p}^*$ lies over $\mathfrak{p}$, there exists a prime ideal $\mathfrak{q}^* \supseteq \mathfrak{p}^*$ which lies over $\mathfrak{q}$; that is, $\mathfrak{q}^* \cap R = \mathfrak{q}$.




Lying Over. Going Up.

We are going to see that extension and contraction are well-behaved in the presence of integral extensions.

**Lemma C-5.55.** *Let $R^*$ be an integral extension of $R$.*

(i) *If $\mathfrak{p}$ is a prime ideal in $R$ and $\mathfrak{p}^*$ lies over $\mathfrak{p}$, then $R^*/\mathfrak{p}^*$ is integral over $R/\mathfrak{p}$.*

(ii) *If $S$ is a multiplicative subset of $R$, then $S^{-1}R^*$ is integral over $S^{-1}R$.*

**Proof.**

(i) We view $R/\mathfrak{p}$ as a subring of $R^*/\mathfrak{p}^*$, by the Second Isomorphism Theorem:

$$R/\mathfrak{p} = R/(\mathfrak{p}^* \cap R) \cong (R + \mathfrak{p}^*)/\mathfrak{p}^* \subseteq R^*/\mathfrak{p}^*.$$

Each element in $R^*/\mathfrak{p}^*$ has the form $\alpha + \mathfrak{p}^*$, where $\alpha \in R^*$. Since $R^*$ is integral over $R$, there is an equation

$$\alpha^n + r_{n-1}\alpha^{n-1} + \cdots + r_0 = 0,$$

where $r_i \in R$. Now view this equation mod $\mathfrak{p}^*$ to see that $\alpha + \mathfrak{p}^*$ is integral over $R/\mathfrak{p}$.

(ii) If $\alpha^* \in S^{-1}R^*$, then $\alpha^* = \alpha/s$, where $\alpha \in R^*$ and $s \in S$. Since $R^*$ is integral over $R$, there is an equation $\alpha^n + r_{n-1}\alpha^{n-1} + \cdots + r_0 = 0$ with $r_i \in R$. Multiplying by $1/s^n$ in $S^{-1}R^*$ gives

$$(\alpha/s)^n + (r_{n-1}/s)(\alpha/s)^{n-1} + \cdots + r_0/s^n = 0,$$

which shows that $\alpha/s$ is integral over $S^{-1}R$. •

When $R^*/R$ is a ring extension and $R$ is a field, every proper ideal in $R^*$ contracts to (0) in $R$. The following proposition eliminates this collapse when $R^*$ is an integral extension of $R$.

**Proposition C-5.56.** *Let $R^*/R$ be a ring extension of domains with $R^*$ integral over $R$. Then $R^*$ is a field if and only if $R$ is a field.*

**Proof.** Assume that $R^*$ is a field. If $u \in R$ is nonzero, then $u^{-1} \in R^*$, and so $u^{-1}$ is integral over $R$. Therefore, there is an equation $(u^{-1})^n + r_{n-1}(u^{-1})^{n-1} + \cdots + r_0 = 0$, where the $r_i \in R$. Multiplying by $u^{n-1}$ gives $u^{-1} = -(r_{n-1} + \cdots + r_0 u^{n-1})$. Therefore, $u^{-1} \in R$ and $R$ is a field.

Conversely, assume that $R$ is a field. If $\alpha \in R^*$ is nonzero, then there is a monic $f(x) \in R[x]$ with $f(\alpha) = 0$. Thus, $\alpha$ is algebraic over $R$, and so we may assume that $f(x) = \mathrm{irr}(\alpha, R)$; that is, $f(x)$ is irreducible. If $f(x) = \sum_{i=0}^{n} r_i x^i$, then

$$\alpha(\alpha^{n-1} + r_{n-1}\alpha^{n-1} + \cdots + r_1) = -r_0.$$

Irreducibility of $f(x)$ gives $r_0 \neq 0$, so that $\alpha^{-1}$ lies in $R^*$. Thus, $R^*$ is a field. •

**Corollary C-5.57.** *Let $R^*/R$ be an integral extension. If $\mathfrak{p}$ is a prime ideal in $R$ and $\mathfrak{p}^*$ is a prime ideal lying over $\mathfrak{p}$, then $\mathfrak{p}$ is a maximal ideal if and only if $\mathfrak{p}^*$ is a maximal ideal.*

**Proof.** By Lemma C-5.55(i), the domain $R^*/\mathfrak{p}^*$ is integral over the domain $R/\mathfrak{p}$. But now Proposition C-5.56 says that $R^*/\mathfrak{p}^*$ is a field if and only if $R/\mathfrak{p}$ is a field; that is, $\mathfrak{p}^*$ is a maximal ideal in $R^*$ if and only if $\mathfrak{p}$ is a maximal ideal in $R$. •

The next corollary gives an important property of rings of integers $\mathcal{O}_E$.

**Corollary C-5.58.** *If $E$ is an algebraic number field, then every nonzero prime ideal in $\mathcal{O}_E$ is a maximal ideal.*

**Proof.** Let $\mathfrak{p}$ be a nonzero prime ideal in $\mathcal{O}_E$. If $\mathfrak{p} \cap \mathbb{Z} \neq (0)$, then there is a prime $p$ with $\mathfrak{p} \cap \mathbb{Z} = (p)$, by Example C-5.53(i). But $(p)$ is a maximal ideal in $\mathbb{Z}$, so that $\mathfrak{p}$ is a maximal ideal, by Corollary C-5.57. It remains to show that $\mathfrak{p} \cap \mathbb{Z} \neq (0)$. Let $\alpha \in \mathfrak{p}$ be nonzero. Since $\alpha$ is integral over $\mathbb{Z}$, there is an equation

$$\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 = 0,$$

where $c_i \in \mathbb{Z}$ for all $i$. If we choose such an equation with $n$ minimal, then $c_0 \neq 0$. Since $\alpha \in \mathfrak{p}$, we have $c_0 = -\alpha(\alpha_{n-1} + c_{n-1}\alpha^{n-2} + \cdots + c_1) \in \mathfrak{p} \cap \mathbb{Z}$, so that $\mathfrak{p} \cap \mathbb{Z}$ is nonzero. •

**Corollary C-5.59.** *Let $R^*$ be integral over $R$, let $\mathfrak{p}$ be a prime ideal in $R$, and let $\mathfrak{p}^*$ and $\mathfrak{q}^*$ be prime ideals in $R^*$ lying over $\mathfrak{p}$. If $\mathfrak{p}^* \subseteq \mathfrak{q}^*$, then $\mathfrak{p}^* = \mathfrak{q}^*$.*

**Proof.** Lemma C-5.55(ii) and Corollary C-5.20(iii) show that the hypotheses are preserved by localizing at $\mathfrak{p}$; that is, $R^*_{\mathfrak{p}}$ is integral over $R_{\mathfrak{p}}$ and $\mathfrak{p}^*R^*_{\mathfrak{p}} \subseteq \mathfrak{q}^*R^*_{\mathfrak{p}}$ are prime ideals. Hence, replacing $R^*$ and $R$ by their localizations, we may assume that $R^*$ and $R$ are local rings and that $\mathfrak{p}$ is a maximal ideal in $R$ (by Theorem C-5.23). But Corollary C-5.57 says that maximality of $\mathfrak{p}$ forces maximality of $\mathfrak{p}^*$. Since $\mathfrak{p}^* \subseteq \mathfrak{q}^*$, we have $\mathfrak{p}^* = \mathfrak{q}^*$. •

Here are the Cohen–Seidenberg Theorems.

**Theorem C-5.60 (Lying Over).** *Let $R^*/R$ be a ring extension with $R^*$ integral over $R$. If $\mathfrak{p}$ is a prime ideal in $R$, then there is a prime ideal $\mathfrak{p}^*$ in $R^*$ lying over $\mathfrak{p}$; that is, $\mathfrak{p}^* \cap R = \mathfrak{p}$.*

**Proof.** There is a commutative diagram

$$\begin{array}{ccc} R & \xrightarrow{\ i\ } & R^* \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h^*} \\ R_{\mathfrak{p}} & \xrightarrow[\ j\ ]{} & S^{-1}R^* \end{array}$$

where $h$ and $h^*$ are localization maps and $i$ and $j$ are inclusions. Let $S = R - \mathfrak{p}$; then $S^{-1}R^*$ is an extension of $R_{\mathfrak{p}}$ (since localization is an exact functor, $R$ contained in $R^*$ implies that $R_{\mathfrak{p}}$ is contained in $S^{-1}R^*$); by Lemma C-5.55, $S^{-1}R^*$ is integral over $R_{\mathfrak{p}}$. Choose a maximal ideal $\mathfrak{m}^*$ in $S^{-1}R^*$. By Corollary C-5.57, $\mathfrak{m}^* \cap R_{\mathfrak{p}}$ is a maximal ideal in $R_{\mathfrak{p}}$. But $R_{\mathfrak{p}}$ is a local ring with unique maximal ideal $\mathfrak{p}R_{\mathfrak{p}}$, so that $\mathfrak{m}^* \cap R_{\mathfrak{p}} = \mathfrak{p}R_{\mathfrak{p}}$. Since the inverse image of a prime ideal (under any ring map) is always prime, the ideal $\mathfrak{p}^* = (h^*)^{-1}(\mathfrak{m}^*)$ is a prime ideal in $R^*$. Now

$$(h^*i)^{-1}(\mathfrak{m}^*) = i^{-1}(h^*)^{-1}(\mathfrak{m}^*) = i^{-1}(\mathfrak{p}^*) = \mathfrak{p}^* \cap R,$$

while

$$(jh)^{-1}(\mathfrak{m}^*) = h^{-1}j^{-1}(\mathfrak{m}^*) = h^{-1}(\mathfrak{m}^* \cap R_{\mathfrak{p}}) = h^{-1}(\mathfrak{p}R_{\mathfrak{p}}) = \mathfrak{p}.$$

Therefore, $\mathfrak{p}^*$ is a prime ideal lying over $\mathfrak{p}$. •

**Theorem C-5.61 (Going Up).** *Let $R^*/R$ be a ring extension with $R^*$ integral over $R$. If $\mathfrak{p} \subseteq \mathfrak{q}$ are prime ideals in $R$ and $\mathfrak{p}^*$ is a prime ideal in $R^*$ lying over $\mathfrak{p}$, then there exists a prime ideal $\mathfrak{q}^*$ lying over $\mathfrak{q}$ with $\mathfrak{p}^* \subseteq \mathfrak{q}^*$.*

**Proof.** Lemma C-5.55 says that $(R^*/\mathfrak{p}^*)/(R/\mathfrak{p})$ is an integral ring extension, where $R/\mathfrak{p}$ is imbedded in $R^*/\mathfrak{p}^*$ as $(R + \mathfrak{p}^*)/\mathfrak{p}^*$. Replacing $R^*$ and $R$ by these quotient rings, we may assume that both $\mathfrak{p}^*$ and $\mathfrak{p}$ are (0). The theorem now follows at once from the Lying Over Theorem. •

There is also a *Going Down Theorem*, but it requires an additional hypothesis.

**Theorem C-5.62 (Going Down).** *Let $R^*/R$ be an integral extension and assume that $R$ is integrally closed. If $\mathfrak{p}_1 \supseteq \mathfrak{p}_2 \supseteq \cdots \supseteq \mathfrak{p}_n$ is a chain of prime ideals in $R$ and, for $m < n$, $\mathfrak{p}_1^* \supseteq \mathfrak{p}_2^* \supseteq \cdots \supseteq \mathfrak{p}_m^*$ is a chain of prime ideals in $R^*$ with each $\mathfrak{p}_i^*$ lying over $\mathfrak{p}_i$, then the chain in $R^*$ can be extended to $\mathfrak{p}_1^* \supseteq \mathfrak{p}_2^* \supseteq \cdots \supseteq \mathfrak{p}_n^*$ with $\mathfrak{p}_i^*$ lying over $\mathfrak{p}_i$ for all $i \leq n$.*

**Proof.** Atiyah–Macdonald [**12**], p. 64. •

## Exercises

∗ **C-5.27.** If $R$ is an integrally closed domain and $S \subseteq R$ is multiplicative, prove that $S^{-1}R$ is also integrally closed.

**C-5.28.** Prove that every valuation ring is integrally closed.

∗ **C-5.29.** Let $R^*/R$ be a ring extension. If $I$ is an ideal in $R$, denote its extension by $I^e$; if $I^*$ is an ideal in $R^*$, denote its contraction by $I^{*\,c}$. Prove each of the following assertions:

(i) Both $e$ and $c$ preserve inclusion: if $I \subseteq J$, then $I^e \subseteq J^e$; if $I^* \subseteq J^*$, then $I^{*\,c} \subseteq J^{*\,c}$.

(ii) $I^{*\,ce} \subseteq I^*$ and $I^{ec} \supseteq I$.

(iii) $I^{*\,cec} = I^{*\,c}$ and $I^{ece} = I^e$.

(iv) $(I^* + J^*)^c \supseteq I^{*\,c} + J^{*\,c}$ and $(I + J)^e = I^e + J^e$.

(v) $(I^* \cap J^*)^c = I^{*\,c} \cap J^{*\,c}$ and $(I \cap J)^e \subseteq I^e \cap J^e$.

(vi) $(I^*J^*)^c \supseteq I^{*\,c}J^{*\,c}$ and $(IJ)^e = I^eJ^e$.

(vii) $\left(\sqrt{I^*}\right)^c = \sqrt{I^{*\,c}}$ and $\left(\sqrt{I}\right)^e \subseteq \sqrt{I^e}$.

(viii) $(J^* : I^*)^c \subseteq (J^{*\,c} : I^{*\,c})$ and $(I : J)^e \subseteq (I^e : J^e)$.

**C-5.30.** If $\mathbb{A}$ is the field of all algebraic numbers, denote the ring of all algebraic integers by $\mathcal{O}_{\mathbb{A}}$. Prove that

$$\mathcal{O}_{\mathbb{A}} \cap \mathbb{Q} = \mathbb{Z}.$$

Conclude, for every algebraic number field $E$, that $\mathcal{O}_E \cap \mathbb{Q} = \mathbb{Z}$.

**C-5.31.** Let $R^*/R$ be an integral ring extension.

(i) If $a \in R$ is a unit in $R^*$, prove that $a$ is a unit in $R$.

(ii) Prove that $J(R) = R \cap J(R^*)$, where $J(R)$ is the Jacobson radical.

**C-5.32.** Generalize Theorem C-5.61 as follows. Let $R$ be integrally closed and let $R^*/R$ be an integral extension. If $\mathfrak{p}_1 \subseteq \mathfrak{p}_2 \subseteq \cdots \subseteq \mathfrak{p}_n$ is a chain of prime ideals in $R$ and, for $m < n$, $\mathfrak{p}_1^* \subseteq \mathfrak{p}_2^* \subseteq \cdots \subseteq \mathfrak{p}_m^*$ is a chain of prime ideals in $R^*$ with each $\mathfrak{p}_i^*$ lying over $\mathfrak{p}_i$, prove that the chain in $R^*$ can be extended to $\mathfrak{p}_1^* \subseteq \mathfrak{p}_2^* \subseteq \cdots \subseteq \mathfrak{p}_n^*$ with $\mathfrak{p}_i^*$ lying over $\mathfrak{p}_i$ for all $i \leq n$.

∗ **C-5.33.** Let $R^*/R$ be an integral extension. If every nonzero prime ideal in $R$ is a maximal ideal, prove that every nonzero prime ideal in $R^*$ is also a maximal ideal.

**Hint.** See the proof of Corollary C-5.58.

∗ **C-5.34.** Let $\alpha$ be algebraic over $\mathbb{Q}$, let $E/\mathbb{Q}$ be a splitting field, and let $G = \mathrm{Gal}(E/\mathbb{Q})$ be its Galois group.

(i) Prove that if $\alpha$ is integral over $\mathbb{Z}$, then, for all $\sigma \in G$, $\sigma(\alpha)$ is also integral over $\mathbb{Z}$.

(ii) Prove that $\alpha$ is an algebraic integer if and only if $\mathrm{irr}(\alpha, \mathbb{Q}) \in \mathbb{Z}[x]$. Compare this proof with that of Corollary C-2.96.

(iii) Let $E$ be an algebraic number field and let $R \subseteq E$ be integrally closed. If $\alpha \in R$, prove that $\mathrm{irr}(\alpha, \mathrm{Frac}(R)) \in R[x]$.
**Hint.** If $\widehat{E}$ is a Galois extension of $\mathrm{Frac}(R)$ containing $\alpha$, then $G = \mathrm{Gal}(\widehat{E}/\,\mathrm{Frac}(R))$ acts transitively on the roots of $\mathrm{irr}(\alpha, \mathrm{Frac}(R))$.

## ▪ Algebraic Integers

We have mentioned that Kummer investigated the ring $\mathbb{Z}[\zeta_p]$, where $p$ is an odd prime and $\zeta_p$ is a primitive $p$th root of unity. We now study rings of integers in algebraic number fields $E$ further. Recall the definition:

$$\mathcal{O}_E = \{\alpha \in E : \alpha \text{ is integral over } \mathbb{Z}\}.$$

We begin with a consequence of Gauss's Lemma. See page 111 of Part 1.

**Lemma C-5.63.** *Let $E$ be an algebraic number field with $[E : \mathbb{Q}] = n$, and let $\alpha \in E$ be an algebraic integer. Then* $\operatorname{irr}(\alpha, \mathbb{Q}) \in \mathbb{Z}[x]$ *and* $\deg(\operatorname{irr}(\alpha, \mathbb{Q})) \mid n$.

**Proof.** By Corollary C-2.96, $\operatorname{irr}(\alpha, \mathbb{Q}) \in \mathbb{Z}[x]$, and so the result follows from Proposition A-3.84 in Part 1. •

**Definition.** A ***quadratic field*** is an algebraic number field $E$ with $[E : \mathbb{Q}] = 2$.

**Proposition C-5.64.** *Every quadratic field $E$ has the form $E = \mathbb{Q}(\sqrt{d})$, where $d$ is a squarefree integer.*

**Proof.** We know that $E = \mathbb{Q}(\alpha)$, where $\alpha$ is a root of a quadratic polynomial; say, $\alpha^2 + b\alpha + c = 0$, where $b, c \in \mathbb{Q}$. If $D = b^2 - 4c$, then the quadratic formula gives $\alpha = -\frac{1}{2}b \pm \frac{1}{2}\sqrt{D}$, and so $E = \mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt{D})$. Write $D$ in lowest terms: $D = U/V$, where $U, V \in \mathbb{Z}$ and $(U, V) = 1$. Now $U = ur^2$ and $V = vs^2$, where $u, v$ are squarefree; hence, $uv$ is squarefree, because $(u, v) = 1$. Therefore, $\mathbb{Q}(\sqrt{D}) = \mathbb{Q}(\sqrt{u/v}) = \mathbb{Q}(\sqrt{uv})$, for $\sqrt{u/v} = \sqrt{uv/v^2} = \sqrt{uv}/v$. •

We now describe the integers in quadratic fields.

**Proposition C-5.65.** *Let $E = \mathbb{Q}(\sqrt{d})$, where $d$ is a squarefree integer (which implies that $d \not\equiv 0 \bmod 4$).*

(i) *If $d \equiv 2 \bmod 4$ or $d \equiv 3 \bmod 4$, then $\mathcal{O}_E = \mathbb{Z}[\sqrt{d}]$.*

(ii) *If $d \equiv 1 \bmod 4$, then $\mathcal{O}_E$ consists of all $\frac{1}{2}\big(u + v\sqrt{d}\big)$ with $u$ and $v$ rational integers having the same parity.*

**Proof.** If $\alpha \in E = \mathbb{Q}(\sqrt{d})$, then there are $a, b \in \mathbb{Q}$ with $\alpha = a + b\sqrt{d}$. We first show that $\alpha \in \mathcal{O}_E$ if and only if

$$2a \in \mathbb{Z} \quad \text{and} \quad a^2 - db^2 \in \mathbb{Z}. \tag{1}$$

If $\alpha \in \mathcal{O}_E$, then Lemma C-5.63 says that $p(x) = \operatorname{irr}(\alpha, \mathbb{Q}) \in \mathbb{Z}[x]$ is quadratic. Now $\operatorname{Gal}(E/\mathbb{Q}) = \langle\sigma\rangle$, where $\sigma\colon E \to E$ carries $\sqrt{d} \mapsto -\sqrt{d}$; that is,

$$\sigma(\alpha) = a - b\sqrt{d}.$$

Since $\sigma$ permutes the roots of $p(x)$, the other root of $p(x)$ is $\sigma(\alpha)$; that is,

$$p(x) = (x - \alpha)(x - \sigma(\alpha)) = x^2 - 2ax + (a^2 - db^2).$$

Hence, the formulas in (1) hold, because $p(x) \in \mathbb{Z}[x]$.

Conversely, if (1) holds, then $\alpha \in \mathcal{O}_E$, because $\alpha$ is a root of a monic polynomial in $\mathbb{Z}[x]$, namely, $x^2 - 2ax + (a^2 - db^2)$.

We now show that $2b \in \mathbb{Z}$. Multiplying the second formula in (1) by 4 gives $(2a)^2 - d(2b)^2 \in \mathbb{Z}$. Since $2a \in \mathbb{Z}$, we have $d(2b)^2 \in \mathbb{Z}$. Write $2b$ in lowest terms: $2b = m/n$, where $(m, n) = 1$. Now $dm^2/n^2 \in \mathbb{Z}$, so that $n^2 \mid dm^2$. But $(n^2, m^2) = 1$ forces $n^2 \mid d$; as $d$ is squarefree, $n = 1$ and $2b = m/n \in \mathbb{Z}$.

We have shown that $a = \frac{1}{2}u$ and $b = \frac{1}{2}v$, where $u, v \in \mathbb{Z}$. Substituting these values into the second formula in (1) gives

$$u^2 \equiv dv^2 \bmod 4. \tag{2}$$

Note that squares are congruent mod 4, either to 0 or to 1. If $d \equiv 2 \bmod 4$, then the only way to satisfy (2) is $u^2 \equiv 0 \bmod 4$ and $v^2 \equiv 0 \bmod 4$. Thus, both $u$ and $v$ must be even, and so $\alpha = \frac{1}{2}u + \frac{1}{2}v\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$. Therefore, $\mathcal{O}_E = \mathbb{Z}[\sqrt{d}]$ in this case, for $\mathbb{Z}[\sqrt{d}] \subseteq \mathcal{O}_E$ is easily seen to be true. A similar argument works when $d \equiv 3 \bmod 4$. However, if $d \equiv 1 \bmod 4$, then $u^2 \equiv v^2 \bmod 4$. Hence, $v$ is even if and only if $u$ is even; that is, $u$ and $v$ have the same parity. If $u$ and $v$ are both even, then $a, b \in \mathbb{Z}$ and $\alpha \in \mathcal{O}_E$. If $u$ and $v$ are both odd, then $u^2 \equiv 1 \equiv v^2 \bmod 4$, and so $u^2 \equiv dv^2 \bmod 4$, because $d \equiv 1 \bmod 4$. Therefore, (1) holds, and so $\alpha$ lies in $\mathcal{O}_E$. •

If $E = \mathbb{Q}(\sqrt{d})$, where $d \in \mathbb{Z}$, then $\mathbb{Z}[\sqrt{d}] \subseteq \mathcal{O}_E$, but we now see that this inclusion may be strict. For example, $\frac{1}{2}\big(1 + \sqrt{5}\big)$ is an algebraic integer (it is a root of $x^2 - x - 1$). Therefore, $\mathbb{Z}[\sqrt{5}] \subsetneq \mathcal{O}_E$, where $E = \mathbb{Q}(\sqrt{5})$.

The coming brief digression into linear algebra will enable us to prove that rings of integers $\mathcal{O}_E$ are noetherian.

**Definition.** Let $E/k$ be a field extension in which $E$ is finite-dimensional. If $u \in E$, then multiplication $\Gamma_u \colon E \to E$, given by $\Gamma_u \colon y \mapsto uy$, is a $k$-map. If $e_1, \ldots, e_n$ is a basis of $E$, then $\Gamma_u$ is represented by a matrix $A = [a_{ij}]$ with entries in $k$; that is,

$$\Gamma_u(e_i) = ue_i = \sum a_{ij} e_j.$$

Define the ***trace*** and ***norm***:

$$\operatorname{tr}(u) = \operatorname{tr}(\Gamma_u) \quad \text{and} \quad N(u) = \det(\Gamma_u).$$

The characteristic polynomial of a linear transformation and, hence, any of its coefficients is independent of any choice of basis of $E/k$, and so the definitions of trace and norm do not depend on the choice of basis. It is easy to see that $\operatorname{tr} \colon E \to k$ is a linear functional and that $N \colon E^\times \to k^\times$ is a (multiplicative) homomorphism.

If $u \in k$, then the matrix of $\Gamma_u$, with respect to any basis of $E/k$, is the scalar matrix $uI$. Hence, if $u \in k$, then

$$\operatorname{tr}(u) = [E : k]u \quad \text{and} \quad N(u) = u^{[E:k]}.$$

**Definition.** The ***trace form*** is the function $t \colon E \times E \to k$ given by

$$t(u, v) = \operatorname{tr}(uv) = \operatorname{tr}(\Gamma_{uv}).$$

It is a routine exercise, left to the reader, to check that the trace form is a symmetric bilinear form.

**Example C-5.66.** If $E = \mathbb{Q}(\sqrt{d})$ is a quadratic field, then a basis for $E/\mathbb{Q}$ is $1, \sqrt{d}$. If $u = a + b\sqrt{d}$, then the matrix of $\Gamma_u$ is

$$\begin{bmatrix} a & bd \\ b & a \end{bmatrix},$$

so that

$$\operatorname{tr}(u) = 2a \quad \text{and} \quad N(u) = a^2 - db^2 = u\overline{u}.$$

Thus, trace and norm arose in the description of the integers in quadratic fields, in relations (1).

We now show that $u = a + b\sqrt{d}$ is a unit in $\mathcal{O}_E$ if and only if $N(u) = \pm 1$. If $u$ is a unit, then there is $v \in \mathcal{O}_E$ with $1 = uv$. Hence, $1 = N(1) = N(uv) = N(u)N(v)$, so that $N(u)$ is a unit in $\mathbb{Z}$; that is, $N(u) = \pm 1$. Conversely, if $N(u) = \pm 1$, then $N(\overline{u}) = N(u) = \pm 1$, where $\overline{u} = a - b\sqrt{d}$. Therefore, $N(u\overline{u}) = 1$. But $u\overline{u} \in \mathbb{Q}$, so that $1 = N(u\overline{u}) = (u\overline{u})^2$. Therefore, $u\overline{u} = \pm 1$, and so $u$ is a unit. ◀

**Lemma C-5.67.** *Let $E/k$ be a field extension of finite degree $n$, and let $u \in E$. If $u = u_1, \ldots, u_s$ are the roots (with multiplicity) of $\operatorname{irr}(u, k)$ in some extension field of $E$, that is, $\operatorname{irr}(u, k) = \prod_{i=1}^{s}(x - u_i)$, then*

$$\operatorname{tr}(u) = [E : k(u)] \sum_{i=1}^{s} u_i \quad \text{and} \quad N(u) = \Big(\prod_{i=1}^{s} u_i\Big)^{[E:k(u)]}.$$

**Remark.** Of course, if $u$ is separable over $k$, then $\operatorname{irr}(u, k)$ has no repeated roots and each $u_i$ occurs exactly once in the formulas. ◀

**Proof.** We sketch the proof. A basis of $k(u)$ over $k$ is $1, u, u^2, \ldots, u^{s-1}$, and the matrix $C_1$ of $\Gamma_u | k(u)$ with respect to this basis is the companion matrix of $\operatorname{irr}(u, k)$. If $1, v_2, \ldots, v_r$ is a basis of $E$ over $k(u)$, then the list

$$1, u, \ldots, u^{s-1}, v_1, v_1u, \ldots, v_1u^{s-1}, \ldots, v_r, v_ru, \ldots, v_ru^{s-1}$$

is a basis of $E$ over $k$. Each of the subspaces $k(u)$ and $\langle v_j, v_ju, \ldots, v_ju^{s-1}\rangle$ for $j \geq 2$ is $\Gamma_u$-invariant, and so the matrix of $\Gamma_u$ relative to the displayed basis of $E$ over $k$ is a direct sum of blocks $C_1 \oplus \cdots \oplus C_r$. In fact, the reader may check that each $C_j$ is the companion matrix of $\operatorname{irr}(u, k)$. The trace and norm formulas now follow from $\operatorname{tr}(C_1 \oplus \cdots \oplus C_r) = \sum_j \operatorname{tr}(C_j)$ and $\det(C_1 \oplus \cdots \oplus C_r) = \prod_j \det(C_j)$. •

If $E/k$ is a field extension and $u \in E$, then a more precise notation for the trace and norm is

$$\operatorname{tr}_{E/k}(u) \quad \text{and} \quad N_{E/k}(u).$$

Indeed, the formulas in Lemma C-5.67 display the dependence on the larger field $E$.

**Proposition C-5.68.** *Let $R$ be a domain with $Q = \operatorname{Frac}(R)$, let $E/Q$ be a field extension of finite degree $[E : Q] = n$, and let $u \in E$ be integral over $R$. If $R$ is integrally closed, then*

$$\operatorname{tr}(u) \in R \quad \textit{and} \quad N(u) \in R.$$

**Proof.** The formulas for $\mathrm{tr}(u)$ and $N(u)$ in Lemma C-5.67 express each as an elementary symmetric function of the roots $u = u_1, \ldots, u_s$ of $\mathrm{irr}(u, Q)$. Since $u$ is integral over $R$, Exercise C-5.34 on page 454 says that $\mathrm{irr}(u, Q) \in R[x]$. Therefore, $\sum_i u_i$ and $\prod_i u_i$ lie in $R$, and hence $\mathrm{tr}(u)$ and $N(u)$ lie in $R$. •

In Example A-5.43 in Part 1, we saw that if $E/k$ is a finite separable extension, then its normal closure $\widehat{E}$ is a Galois extension of $k$. Recall from the Fundamental Theorem of Galois Theory, Theorem A-5.51 in Part 1, that if $G = \mathrm{Gal}(\widehat{E}/k)$ and $H = \mathrm{Gal}(\widehat{E}/E) \subseteq G$, then $[G : H] = [E : k]$.

**Lemma C-5.69.** *Let $E/k$ be a separable field extension of finite degree $n = [E : k]$ and let $\widehat{E}$ be a normal closure of $E$. Write $G = \mathrm{Gal}(\widehat{E}/k)$ and $H = \mathrm{Gal}(\widehat{E}/E)$, and let $T$ be a transversal of $H$ in $G$; that is, $G$ is the disjoint union $G = \bigcup_{\sigma \in T} \sigma H$.*

(i) *For all $u \in E$,*

$$\prod_{\sigma \in T} (x - \sigma(u)) = \mathrm{irr}(u, k)^{[E:k(u)]}.$$

(ii) *For all $u \in E$,*

$$\mathrm{tr}(u) = \sum_{\sigma \in T} \sigma(u) \quad \textit{and} \quad N(u) = \prod_{\sigma \in T} \sigma(u).$$

**Proof.**

(i) Denote $\prod_{\sigma \in T}(x - \sigma(u))$ by $h(x)$; of course, $h(x) \in \widehat{E}[x]$.

We claim that the set $X$, defined by $X = \{\sigma(u) : \sigma \in T\}$, satisfies $\tau(X) = X$ for every $\tau \in G$. If $\sigma \in T$, then $\tau\sigma \in \sigma' H$ for some $\sigma' \in T$, because $T$ is a left transversal; hence, $\tau\sigma = \sigma'\eta$ for some $\eta \in H$. But $\tau\sigma(u) = \sigma'\eta(u) = \sigma'(u)$, because $\eta \in H$, and every element of $H$ fixes $E$. Therefore, $\tau\sigma(u) = \sigma'(u) \in X$. Thus, the function $\varphi_\tau$, defined by $\sigma(u) \mapsto \tau\sigma(u)$, is a function $X \to X$. In fact, $\varphi_\tau$ is a permutation, because $\tau$ is an isomorphism, hence $\varphi_\tau$ is an injection, and hence it is a bijection, by the Pigeonhole Principle. It follows that every elementary symmetric function on $X = \{\sigma(u) : \sigma \in T\}$ is fixed by every $\tau \in G$. Since $\widehat{E}/k$ is a Galois extension, each value of these elementary symmetric functions lies in $k$. We have shown that all the coefficients of $h(x)$ lie in $k$, and so $h(x) \in k[x]$. Now compare $h(x)$ and $\mathrm{irr}(u, k)$. If $\sigma \in G$, then $\sigma$ permutes the roots of $\mathrm{irr}(u, k)$, so that every root $\sigma(u)$ of $h(x)$ is also a root of $\mathrm{irr}(u, k)$. By Exercise A-3.75 in Part 1, we have

$$h(x) = \mathrm{irr}(u, k)^m$$

for some $m \geq 1$, and so it only remains to compute $m$. Now

$$\deg(h) = m \deg(\mathrm{irr}(u, k)) = m[k(u) : k].$$

But $\deg(h) = [G : H] = [E : k]$, and so $m = [E : k]/[k(u) : k] = [E : k(u)]$.

(ii) Recall our earlier notation: $\mathrm{irr}(u, k) = \prod_{i=1}^{s}(x - u_i)$. Since

$$\prod_{\sigma \in T} (x - \sigma(u)) = \mathrm{irr}(u, k)^{[E:k(u)]} = \Big(\prod_{i=1}^{s}(x - u_i)\Big)^{[E:k(u)]},$$

their constant terms are the same,

$$\pm \prod_{\sigma\in T} \sigma(u) = \pm\Big(\prod_{i=1}^{s} u_i\Big)^{[E:k(u)]},$$

and their penultimate coefficients are the same,

$$-\sum_{\sigma\in T} \sigma(u) = -[E : k(u)] \sum_{i=1}^{s} u_i.$$

By Lemma C-5.67, $\operatorname{tr}(u) = [E : k(u)] \sum_{i=1}^{s} u_i$ and $N(u) = \big(\prod_{i=1}^{s} u_i\big)^{[E:k(u)]}$. It follows that

$$\operatorname{tr}(u) = [E : k(u)] \sum_{i=1}^{s} u_i = \sum_{\sigma\in T} \sigma(u)$$

and

$$N(u) = \Big(\prod_{i=1}^{s} u_i\Big)^{[E:k(u)]} = \prod_{\sigma\in T} \sigma(u). \quad \bullet$$

**Definition.** Let $E/k$ be a finite field extension, let $\widehat{E}$ be a normal closure of $E$, and let $T$ be a left transversal of $\operatorname{Gal}(\widehat{E}/E)$ in $\operatorname{Gal}(\widehat{E}/k)$. If $u \in E$, then the elements $\sigma(u)$, where $\sigma \in T$, are called the ***conjugates*** of $u$.

If $E = \mathbb{Q}(i)$, then $\operatorname{Gal}(E/\mathbb{Q}) = \langle\sigma\rangle$, where $\sigma$ is complex conjugation. If $u = a + ib \in E$, then $\sigma(u) = a - ib$, so that we have just generalized the usual notion of conjugate.

If $E/k$ is a separable extension, then the conjugates of $u$ are the roots of $\operatorname{irr}(u, k)$; in the inseparable case, the conjugates may occur with multiplicities.

**Corollary C-5.70.** *If $E/k$ is a finite Galois extension with $G = \operatorname{Gal}(E/k)$, then*

$$\operatorname{tr}(u) = \sum_{\sigma\in G} \sigma(u) \quad \textit{and} \quad N(u) = \prod_{\sigma\in G} \sigma(u).$$

*Moreover,* tr *is nonzero.*

**Proof.** Since $E/k$ is a Galois extension, $E$ is its own normal closure, and so a transversal $T$ of $G$ in itself is just $G$. If $\operatorname{tr} = 0$, then

$$\operatorname{tr}(u) = \sum_{\sigma\in G} \sigma(u) = 0$$

for all $u \in E$, contradicting Independence of Characters, Proposition A-5.38 in Part 1. $\bullet$

This last corollary shows that the norm here coincides with the norm occurring in Chapter C-3 in the proof of Hilbert's Theorem 90.

Let $V$ be a vector space over a field $k$, and let $f\colon V \times V \to k$ be a bilinear form. If $e_1, \ldots, e_n$ is a basis of $V$, recall that the *discriminant* is defined by

$$D(e_1, \ldots, e_n) = \det([f(e_i, e_j)])$$

and that $f$ is *nondegenerate* if there is a basis whose discriminant is nonzero (it then follows that the discriminant of $f$ with respect to any other basis of $V$ is also nonzero).

**Lemma C-5.71.** *If $E/k$ is a finite separable field extension, then the trace form is nondegenerate.*[5]

**Proof.** We compute the discriminant using Lemma C-5.69 (which uses separability). Let $T = \{\sigma_1, \dots, \sigma_n\}$ be a transversal of $\mathrm{Gal}(\widehat{E}/E)$ in $\mathrm{Gal}(\widehat{E}/k)$, where $\widehat{E}$ is a normal closure of $E$. We have

$$\begin{aligned} D(e_1, \dots, e_n) &= \det([t(e_i, e_j)]) \\ &= \det([\mathrm{tr}(e_i e_j)]) \\ &= \det\Big[\sum_\ell \sigma_\ell(e_i e_j)\Big] \qquad \text{(Lemma C-5.69)} \\ &= \det\Big(\Big[\sum_\ell \sigma_\ell(e_i)\sigma_\ell(e_j)\Big]\Big) \\ &= \det\big([\sigma_\ell(e_i)]\det\big([\sigma_\ell(e_j)]\big)\big) \\ &= \det([\sigma_\ell(e_i)])^2. \end{aligned}$$

Suppose that $\det([\sigma_\ell(e_i)]) = 0$. If $[\sigma_\ell(e_i)]$ is singular, there is a column matrix $C = [c_1, \dots, c_n]^\top \in \widehat{E}^n$ with $[\sigma_\ell(e_i)]C = 0$. Hence,

$$c_1\sigma_1(e_j) + \cdots + c_n\sigma_n(e_j) = 0$$

for $j = 1, \dots, n$. It follows that

$$c_1\sigma_1(v) + \cdots + c_n\sigma_n(v) = 0$$

for every linear combination $v$ of the $e_i$. But this contradicts Independence of Characters. •

**Proposition C-5.72.** *Let $R$ be integrally closed, and let $Q = \mathrm{Frac}(R)$. If $E/Q$ is a finite separable field extension of degree $n$ and $\mathcal{O} = \mathcal{O}_{E/R}$ is the integral closure of $R$ in $E$, then $\mathcal{O}$ can be imbedded as a submodule of a free $R$-module of rank $n$.*

**Proof.** Let $e_1, \dots, e_n$ be a basis of $E/Q$. By Proposition C-5.51, for each $i$ there is $r_i \in R$ with $r_i e_i \in \mathcal{O}$; changing notation if necessary, we may assume that each $e_i \in \mathcal{O}$. Now Corollary B-3.97 in Part 1 (which assumes nondegeneracy of bilinear forms) says that there is a basis $f_1, \dots, f_n$ of $E$ with $t(e_i, f_j) = \mathrm{tr}(e_i f_j) = \delta_{ij}$.

Let $\alpha \in \mathcal{O}$. Since $f_1, \dots, f_n$ is a basis, there are $c_j \in Q$ with $\alpha = \sum c_j f_j$. For each $i$, where $1 \le i \le n$, we have $e_i\alpha \in \mathcal{O}$ (because $e_i \in \mathcal{O}$). Therefore, $\mathrm{tr}(e_i\alpha) \in R$, by Proposition C-5.68. But

$$\mathrm{tr}(e_i\alpha) = \mathrm{tr}\Big(\sum_j c_j e_i f_j\Big) = \sum_j c_j \,\mathrm{tr}(e_i f_j) = c_j\delta_{ij} = c_i.$$

Therefore, $c_i \in R$ for all $i$, and so $\alpha = \sum_i c_i f_i$ lies in the free $R$-module with basis $f_1, \dots, f_n$. •

[5] If $E/k$ is inseparable, then the trace form is identically 0. See Zariski–Samuel [**233**], p. 95.

**Definition.** If $E$ is an algebraic number field, then an ***integral basis*** for $\mathcal{O}_E$ is a list $\beta_1, \ldots, \beta_n$ in $\mathcal{O}_E$ such that every $\alpha \in \mathcal{O}_E$ has a unique expression

$$\alpha = c_1\beta_1 + \cdots + c_n\beta_n,$$

where $c_i \in \mathbb{Z}$ for all $i$.

If $B = \beta_1, \ldots, \beta_n$ is an integral basis of $\mathcal{O}_E$, then $\mathcal{O}_E$ is a free abelian group with basis $B$. We now prove that integral bases always exist.

**Proposition C-5.73.** *Let $E$ be an algebraic number field.*

(i) *The ring of integers $\mathcal{O}_E$ has an integral basis, and hence it is a free abelian group of finite rank under addition.*

(ii) *$\mathcal{O}_E$ is a noetherian domain.*

**Proof.**

(i) Since $\mathbb{Q}$ has characteristic 0, the field extension $E/\mathbb{Q}$ is separable. Hence, Proposition C-5.72 applies to show that $\mathcal{O}_E$ is a submodule of a free $\mathbb{Z}$-module of finite rank; that is, $\mathcal{O}_E$ is a subgroup of a finitely generated free abelian group. Thus, $\mathcal{O}_E$ is itself a free abelian group (subgroups of free abelian groups are free abelian). But a basis of $\mathcal{O}_E$ as a free abelian group is an integral basis.

(ii) Any ideal $I$ in $\mathcal{O}_E$ is a subgroup of a finitely generated free abelian group, and hence $I$ is itself a finitely generated abelian group (subgroups of finitely generated abelian groups are finitely generated). A fortiori, $I$ is a finitely generated $\mathcal{O}_E$-module; that is, $I$ is a finitely generated ideal. •

**Example C-5.74.** We show that $\mathcal{O}_E$ need not be a UFD and, hence, it need not be a PID. Let $E = \mathbb{Q}\big(\sqrt{-5}\big)$. Since $-5 \equiv 3 \bmod 4$, Proposition C-5.65 gives $\mathcal{O}_E = \mathbb{Z}\big[\sqrt{-5}\,\big]$. By Example C-5.66, the only units in $\mathcal{O}_E$ are elements $u$ with $N(u) = \pm 1$. If $a^2 + 5b^2 = \pm 1$, where $a, b \in \mathbb{Z}$, then $b = 0$ and $a = \pm 1$, and so the only units in $\mathcal{O}_E$ are $\pm 1$. Consider the factorizations in $\mathcal{O}_E$:

$$2 \cdot 3 = 6 = \big(1 + \sqrt{-5}\big)\big(1 - \sqrt{-5}\big).$$

Note that no two of these factors are associates (the only units are $\pm 1$), and we now show that each of them is irreducible. If $v \in \mathcal{O}_E$ divides any of these four factors (but is not an associate of it), then $N(v)$ is a proper divisor in $\mathbb{Z}$ of $4, 9$, or $6$, for these are the norms of the four factors ($N\big(1 + \sqrt{-5}\big) = 6 = N\big(1 - \sqrt{-5}\big)$). It is quickly checked, however, that there are no such divisors in $\mathbb{Z}$ of the form $a^2 + 5b^2$ other than $\pm 1$. Therefore, $\mathcal{O}_E = \mathbb{Z}\big[\sqrt{-5}\,\big]$ is not a UFD. ◄

Trace and norm can be used to find other rings of integers.

**Definition.** If $n \geq 2$, then a ***cyclotomic field*** is $E = \mathbb{Q}(\zeta_n)$, where $\zeta_n$ is a primitive $n$th root of unity.

Recall that if $p$ is prime, then the cyclotomic polynomial

$$\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1 \in \mathbb{Z}[x]$$

is irreducible, so that $\operatorname{irr}(\zeta_p, \mathbb{Q}) = \Phi_p(x)$ and $[\mathbb{Q}(\zeta_p) : \mathbb{Q}] = p - 1$ (recall that $\Phi_d(x)$ is irreducible for all, not necessarily prime, natural numbers $d$). Moreover,

$$\operatorname{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}) = \{\sigma_1, \ldots, \sigma_{p-1}\},$$

where $\sigma_i \colon \zeta_p \mapsto \zeta_p^i$ for $i = 1, \ldots, p-1$.

We do some elementary calculations in $E = \mathbb{Q}(\zeta_p)$ to enable us to describe $\mathcal{O}_E$.

**Lemma C-5.75.** *Let $p$ be an odd prime, and let $E = \mathbb{Q}(\zeta)$, where $\zeta = \zeta_p$ is a primitive $p$th root of unity.*

(i) $\operatorname{tr}(\zeta^i) = -1$ *for* $1 \le i \le p-1$.

(ii) $\operatorname{tr}(1 - \zeta^i) = p$ *for* $1 \le i \le p-1$.

(iii) $p = \prod_{i=1}^{p-1} (1 - \zeta^i) = N(1 - \zeta)$.

(iv) $\mathcal{O}_E(1 - \zeta) \cap \mathbb{Z} = p\mathbb{Z}$.

(v) $\operatorname{tr}(u(1 - \zeta)) \in p\mathbb{Z}$ *for every* $u \in \mathcal{O}_E$.

**Proof.**

(i) We have $\operatorname{tr}(\zeta) = \sum_{i=1}^{p-1} \zeta^i = \Phi_p(\zeta) - 1$, which is also true for every primitive $p$th root of unity $\zeta^i$. The result follows from $\Phi_p(\zeta) = 0$.

(ii) Since $\operatorname{tr}(1) = [E : \mathbb{Q}] = p - 1$ and tr is a linear functional,

$$\operatorname{tr}(1 - \zeta^i) = \operatorname{tr}(1) - \operatorname{tr}(\zeta^i) = (p-1) - (-1) = p.$$

(iii) Since $\Phi(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$, we have $\Phi_p(1) = p$. On the other hand, the primitive $p$th roots of unity are the roots of $\Phi_p(x)$, so that

$$\Phi_p(x) = \prod_{i=1}^{p-1} (x - \zeta^i).$$

Evaluating at $x = 1$ gives the first equation. The second equation holds because the $1 - \zeta^i$ are the conjugates of $1 - \zeta$.

(iv) The first equation in (iii) shows that $p \in \mathcal{O}_E(1-\zeta) \cap \mathbb{Z}$, so that $\mathcal{O}_E(1-\zeta) \cap \mathbb{Z} \supseteq p\mathbb{Z}$. If this inclusion is strict, then $\mathcal{O}_E(1-\zeta) \cap \mathbb{Z} = \mathbb{Z}$, because $p\mathbb{Z}$ is a maximal ideal in $\mathbb{Z}$. In this case, $\mathcal{O}_E(1 - \zeta) \cap \mathbb{Z} = \mathbb{Z}$, hence $\mathbb{Z} \subseteq \mathcal{O}_E(1 - \zeta)$, and so $1 \in \mathcal{O}_E(1 - \zeta)$. Thus, there is $v \in \mathcal{O}_E$ with $v(1 - \zeta) = 1$; that is, $1 - \zeta$ is a unit in $\mathcal{O}_E$. But if $1 - \zeta$ is a unit, then $N(1 - \zeta) = \pm 1$, contradicting the second equation in (iii).

(v) Each conjugate $\sigma_i(u(1-\zeta)) = \sigma_i(u)(1-\zeta^i)$ is, obviously, divisible by $1-\zeta^i$ in $\mathcal{O}_E$. But $1-\zeta^i$ is divisible by $1-\zeta$ in $\mathcal{O}_E$, because

$$1-\zeta^i = (1-\zeta)(1+\zeta+\zeta^2+\cdots+\zeta^{i-1}).$$

Hence, $\sigma_i(1-\zeta^i) \in \mathcal{O}_E(1-\zeta)$ for all $i$, and so $\sum_i(u(1-\zeta^i)) \in \mathcal{O}_E(1-\zeta)$. By Corollary C-5.70, $\sum_i(u(1-\zeta)) = \operatorname{tr}(u(1-\zeta))$. Therefore, $\operatorname{tr}(u(1-\zeta)) \in \mathcal{O}_E(1-\zeta) \cap \mathbb{Z} = p\mathbb{Z}$, by (iv), for $\operatorname{tr}(u(1-\zeta)) \in \mathbb{Z}$, by Proposition C-5.68. •

**Proposition C-5.76.** *If $p$ is an odd prime and $E = \mathbb{Q}(\zeta_p)$ is a cyclotomic field, then*

$$\mathcal{O}_E = \mathbb{Z}[\zeta_p].$$

**Proof.** Let us abbreviate $\zeta_p$ as $\zeta$. It is always true that $\mathbb{Z}[\zeta] \subseteq \mathcal{O}_E$, and we now prove the reverse inclusion. By Lemma C-5.63, each element $u \in \mathcal{O}_E$ has an expression

$$u = c_0 + c_1\zeta + c_2\zeta^2 + \cdots + c_{p-2}\zeta^{p-2},$$

where $c_i \in \mathbb{Q}$ (remember that $[E:\mathbb{Q}] = p-1$). We must show that $c_i \in \mathbb{Z}$ for all $i$. Multiplying by $1-\zeta$ gives

$$u(1-\zeta) = c_0(1-\zeta) + c_1(\zeta-\zeta^2) + \cdots + c_{p-2}(\zeta^{p-2}-\zeta^{p-1}).$$

By Lemma C-5.75(i), $\operatorname{tr}(\zeta^i - \zeta^{i+1}) = \operatorname{tr}(\zeta^i) - \operatorname{tr}(\zeta^{i+1}) = 0$ for $1 \le i \le p-2$, so that $\operatorname{tr}(u(1-\zeta)) = c_0 \operatorname{tr}(1-\zeta)$; hence, $\operatorname{tr}(u(1-\zeta)) = pc_0$, because $\operatorname{tr}(1-\zeta) = p$, by Lemma C-5.75(ii). On the other hand, $\operatorname{tr}(u(1-\zeta)) \in p\mathbb{Z}$, by Lemma C-5.75(iv). Hence, $pc_0 = mp$ for some $m \in \mathbb{Z}$, and so $c_0 \in \mathbb{Z}$. Now $\zeta^{-1} = \zeta^{p-1} \in \mathcal{O}_E$, so that

$$(u-c_0)\zeta^{-1} = c_1 + c_2\zeta + \cdots + c_{p-2}\zeta^{p-3} \in \mathcal{O}_E.$$

The argument just given shows that $c_1 \in \mathbb{Z}$. Indeed, repetition of this argument shows that all $c_i \in \mathbb{Z}$, and so $u \in \mathbb{Z}[\zeta]$. •

There are other interesting bases of finite Galois extensions.

**Definition.** Let $E/k$ be a finite Galois extension with $\operatorname{Gal}(E/k) = \{\sigma_1, \ldots, \sigma_n\}$. A ***normal basis*** of $E/k$ is a vector space basis $e_1, \ldots, e_n$ of $E$ for which there exists $u \in E$ with $e_i = \sigma_i(u)$ for all $i$.

If $E/k$ is an extension with $\operatorname{Gal}(E/k) = G$, then $E$ is a $kG$-module, where $\sigma e = \sigma(e)$ for all $e \in E$ and $\sigma \in G$.

**Proposition C-5.77.** *Let $E/k$ be a finite Galois extension with $\operatorname{Gal}(E/k) = G$. Then $E \cong kG$ as $kG$-modules if and only if $E/k$ has a normal basis.*

**Proof.** Suppose $\varphi\colon kG \to E$ is a $kG$-isomorphism, where $G = \{1 = \sigma_1, \dots, \sigma_n\}$. Now $kG$ is a vector space over $k$ with basis $1 = \sigma_1, \dots, \sigma_n$. If $u = \varphi(1)$, then $\varphi(\sigma_i) = \sigma_i\varphi(1) = \sigma_i u = \sigma_i(u)$. As $\varphi$ is a $k$-linear isomorphism, $\{\sigma_i(u) : 1 \le i \le n\}$ is a basis of $E$; that is, $E/k$ has a normal basis.

Conversely, assume that there is $u \in E$ with $\{\sigma_i(u) : 1 \le i \le n\}$ a basis of $E/k$. It is easily checked that $\varphi\colon kG \to E$, given by $\sigma_i \mapsto \sigma_i(u)$, is a $kG$-map; $\varphi$ is an isomorphism because it carries a basis of $kG$ onto a basis of $E$. •

**Proposition C-5.78.** *Let $E/k$ be a Galois extension with $\mathrm{Gal}(E/k) = \{\sigma_1, \dots, \sigma_n\}$, and let $e_1, \dots, e_n \in E$. Then $e_1, \dots, e_n$ is a basis of $E$ over $k$ if and only if the matrix $[\sigma_i(e_j)]$ is nonsingular.*

**Proof.** Assume that $M = [\sigma_i(e_j)]$ is nonsingular. Since $\dim_k(E) = n$, it suffices to prove that $e_1, \dots, e_n$ is linearly independent. Otherwise, there are $a_i \in k$ with

$$a_1e_1 + \cdots + a_ne_n = 0.$$

For each $i$, we have

$$a_1\sigma_i(e_1) + \cdots + a_n\sigma_i(e_n) = 0.$$

Thus, $MA = 0$, where $A$ is the column vector $[a_1, \dots, a_n]^\top$ in $k^n \subseteq E^n$. Since $M$ is nonsingular over $E$, we have $A = 0$, and so $e_1, \dots, e_n$ is a basis.

Conversely, if $M = [\sigma_i(e_j)]$, a matrix with entries in $E$, is singular, there is a nontrivial solution $X = [\alpha_1, \dots, \alpha_n]^\top$ of $MX = 0$; that is, for all $i$, there are $\alpha_1, \dots, \alpha_n \in E$ with

$$\alpha_1\sigma_i(e_1) + \alpha_2\sigma_i(e_2) + \cdots + \alpha_n\sigma_i(e_n) = 0.$$

For notational convenience, we may assume that $\alpha_1 \ne 0$. By Corollary C-5.70, there is $u \in E$ with $\mathrm{tr}(u) \ne 0$. Multiply each equation above by $u\alpha_1^{-1}$ to obtain equations, for all $i$,

$$u\sigma_i(e_1) + \beta_2\sigma_i(e_2) + \cdots + \beta_n\sigma_i(e_n) = 0,$$

where $\beta_j = u\alpha_1^{-1}\alpha_j$ for $j \ge 2$. For each $i$, apply $\sigma_i^{-1}$ to the $i$th equation:

$$\begin{aligned}
\sigma_1^{-1}(u)e_1 + \sigma_1^{-1}(\beta_2)e_2 + \cdots + \sigma_1^{-1}(\beta_n)e_n &= 0,\\
\sigma_2^{-1}(u)e_1 + \sigma_2^{-1}(\beta_2)e_2 + \cdots + \sigma_2^{-1}(\beta_n)e_n &= 0,\\
\vdots \qquad\qquad & \qquad\qquad \vdots\\
\sigma_n^{-1}(u)e_1 + \sigma_n^{-1}(\beta_2)e_2 + \cdots + \sigma_n^{-1}(\beta_n)e_n &= 0.
\end{aligned}$$

Note that as $\sigma_i$ varies over all of $G$, so does $\sigma_i^{-1}$. Adding these equations gives

$$\mathrm{tr}(u)e_1 + \mathrm{tr}(\beta_2)e_2 + \cdots + \mathrm{tr}(\beta_n)e_n = 0.$$

This is a nontrivial linear combination of $e_1, \dots, e_n$, because $\mathrm{tr}(u) \ne 0$, and this contradicts linear independence. •

**Corollary C-5.79.** *Let $E/k$ be a Galois extension with $\mathrm{Gal}(E/k) = \{\sigma_1, \dots, \sigma_n\}$, and let $u \in E$. Then $\sigma_1(u), \dots, \sigma_n(u)$ is a normal basis of $E$ over $k$ if and only if the matrix $[\sigma_i\sigma_j(u)]$ is nonsingular.*

**Proof.** Let $e_1 = \sigma_1(u), \dots, e_n = \sigma_n(u)$ in Proposition C-5.78. •

**Remark.** There is another way to view the matrix $[\sigma_i\sigma_j]$. If $\alpha_1, \dots, \alpha_n \in E$, define

$$M(\alpha_1, \dots, \alpha_n) = [\sigma_j(\alpha_i)].$$

If $\beta_1, \dots, \beta_n \in E$, then $M(\alpha_1, \dots, \alpha_n)M(\beta_1, \dots, \beta_n)^\top = [\mathrm{tr}(\alpha_i\beta_j)]$. In particular,

$$M(\alpha_1, \dots, \alpha_n)M(\alpha_1, \dots, \alpha_n)^\top = [\mathrm{tr}(\alpha_i\alpha_j)]. \tag{3}$$

The latter matrix is just the inner product matrix of the trace form defined on page 456. In fact, in light of (3), Lemma C-5.71 (which says that $[\mathrm{tr}(\alpha_1, \dots, \alpha_n)]$ is nonsingular) gives another proof of the nonsingularity of $[M(\alpha_1, \dots, \alpha_n)]$. ◀

**Proposition C-5.80.**

(i) *If $E/k$ is a finite Galois extension with cyclic Galois group $G$, then $E$ has a normal basis.*

(ii) *If $E$ is a finite field and $k$ is a subfield, then $E/k$ has a normal basis.*

**Proof.**

(i) Let $G = \langle\sigma\rangle$, where $\sigma$ has order $n = [E : k]$, and view $\sigma\colon E \to E$ as a $k$-linear transformation. Now $\sigma^n - 1 = 0$; on the other hand, Independence of Characters says that whenever $c_1, \dots, c_n \in k$ and $\sum_i c_i\sigma^i(e) = 0$ for all $e \in E$, all the $c_i = 0$; thus, if $0 \neq f(x) = \sum_{i=0}^{n-1} c_i x^i \in k[x]$, then $f(\sigma) \neq 0$. It follows that the characteristic and minimal polynomials of $\sigma$ coincide, and so Corollary B-3.62 in Part 1 applies: $E$ is a cyclic $k[x]$-module, say, with generator $u \in E$. Thus, every element of $E$ is a $k$-linear combination of $u, \sigma(u), \dots, \sigma^{n-1}(u)$ (for if $m \geq n$, then $\sigma^m(u) = \sigma^i(u)$, where $m \equiv i \bmod n$). Therefore, $E/k$ has a normal basis.

(ii) Exercise A-5.12 on page 200 in Part 1 says that $\mathrm{Gal}(E/k)$ is cyclic. •

**Lemma C-5.81.** *If $k$ is an infinite field and $E/k$ is a finite Galois extension with $G = \mathrm{Gal}(E/k)$, then the elements $\{\sigma_1, \dots, \sigma_n\}$ of $G$ are algebraically independent over $E$: if $f(\sigma_1(v), \dots, \sigma_n(v)) = 0$ for all $v \in E$, then $f$ is the zero polynomial.*

**Proof.** Suppose that $f(x_1, \dots, x_n) \in E[x_1, \dots, x_n]$ and $f(\sigma_1(v), \dots, \sigma_n(v)) = 0$ for all $v \in E$. If $e_1, \dots, e_n$ is a basis of $E/k$, then $v = \sum_q a_q e_q$ for $a_q \in k$, and so

$$\begin{aligned} 0 &= f\Big(\sigma_1\Big(\sum_q a_q e_q\Big), \dots, \sigma_n\Big(\sum_q a_q e_q\Big)\Big) \\ &= f\Big(\sum_q a_q\sigma_1(e_q), \dots, \sum_q a_q\sigma_n(e_q)\Big). \end{aligned}$$

Define

$$g(x_1, \dots, x_n) = f\Big(\sum_q \sigma_1(e_q)x_q, \dots, \sum_q \sigma_n(e_q)x_q\Big).$$

Thus, $g(a_1, \dots, a_n) = 0$ for all $a_1, \dots, a_n \in k$, and so Proposition A-3.58 in Part 1 (whose hypothesis has $k$ infinite) gives $g(x_1, \dots, x_n) = 0$. Since $e_1, \dots, e_n$ is a basis of $E/k$, Proposition C-5.78 says that the matrix $[\sigma_p(e_q)]$ over $E$ is nonsingular. Let

$[w_{iq}]$ be its inverse, so that $\sum_p w_{ip}\sigma_p(e_q)x_q = \delta_{iq}$ (Kronecker delta). For every $i$, we have

$$\sum_{p,q} w_{ip}\sigma_p(e_q)x_q = \sum_q \Big(\sum_p w_{ip}\sigma_p(e_q)\Big)x_q = \sum_q \delta_{iq}x_q = x_i.$$

Thus, $0 = g(\sum_{p,q} w_{1p}\sigma_p(e_q)x_q, \ldots, \sum_{p,q} w_{np}\sigma_p(e_q)x_q) = f(x_1, \ldots, x_n)$, and so $\sigma_1, \ldots, \sigma_n$ are algebraically independent over $E$. •

If $X$ is a set with $n$ elements, then a ***Latin square*** based on $X$ is an $n \times n$ matrix $L$ whose rows and columns are permutations of $X$.

**Lemma C-5.82.** *Let $E$ be a field. If $\{x_1, \ldots, x_n\} \subseteq E[x_1, \ldots, x_n]$ and $L$ is a Latin square based on $\{x_1, \ldots, x_n\}$, then* $\det(L)$ *is a nonzero polynomial in $E[x_1, \ldots, x_n]$.*

**Proof.** If $R = E[x_1, \ldots, x_n]$, then a ring map $\mu\colon R \to E$ induces a ring map $\mu_*\colon \mathrm{Mat}_n(R) \to \mathrm{Mat}_n(E)$; moreover, $\det(\mu_*(L)) = \mu(\det(L))$. In particular, if $\mu$ is given by $x_1 \mapsto 1$ and $x_i \mapsto 0$ for $i \geq 2$, then $\mu_*(L) = P$, where $P$ is the (permutation) matrix obtained from $L$ by setting $x_1 = 1$ and all other $x_i = 0$. Since $\det(P) = \pm 1 \neq 0$, we have $\det(L) \neq 0$. Proposition B-5.46 in Part 1, the complete expansion of $\det(L)$, shows that $\det(L) \in E[x_1, \ldots, x_n]$. •

**Theorem C-5.83 (Normal Basis Theorem).** *Every finite Galois extension $E/k$ has a normal basis.*

**Proof.** By Proposition C-5.80, we may assume that $k$ is an infinite field.

The matrix $L = [x_i x_j]$ over the polynomial ring $E[x_1, \ldots, x_n]$ is a Latin square based on $\{x_1, \ldots, x_n\}$, and so $f(x_1, \ldots, x_n) = \det([x_{i(j)}])$ is a nonzero polynomial, by Lemma C-5.82. Since $G = \mathrm{Gal}(E/k) = \{\sigma_1, \ldots, \sigma_n\}$ is a group, we have $\sigma_i\sigma_j = \sigma_{i(j)} \in G$. Thus, for each $v \in E$, the entries in the $n$-tuple $(\sigma_{i(1)}(v), \ldots, \sigma_{i(n)}(v))$ form a permutation of the entries in $(\sigma_1(v), \ldots, \sigma_n(v))$, and so $f(\sigma_{i(1)}(v), \ldots, \sigma_{i(n)}(v)) = \det([\sigma_i(\sigma_j(v))])$. Lemma C-5.81 gives $u \in E$ with $f(\sigma_{i(1)}(u), \ldots, \sigma_{i(n)}(u)) \neq 0$. Hence, $f(\sigma_{i(1)}(u), \ldots, \sigma_{i(n)}(u)) = \det[\sigma_{i(j)}(u)] = \det[\sigma_i\sigma_j(u)] \neq 0$. Corollary C-5.79 now applies: $\sigma_1(u), \ldots, \sigma_n(u)$ is a normal basis of $E/k$. •

Before we leave this discussion of algebraic number fields, we must mention a beautiful theorem of Dirichlet. In order to state the theorem, we cite two results whose proofs can be found in Samuel [**196**], Chapter 4. An algebraic number field $E$ of degree $n$ has exactly $n$ imbeddings into $\mathbb{C}$. If $r_1$ is the number of such imbeddings with image in $\mathbb{R}$, then $n - r_1$ is even; say, $n - r_1 = 2r_2$.

**Theorem (Dirichlet Unit Theorem).** *Let $E$ be an algebraic number field of degree $n$. Then $n = r_1 + 2r_2$ (where $r_1$ is the number of imbeddings of $E$ into $\mathbb{R}$), and the multiplicative group $U(\mathcal{O}_E)$ of units in $\mathcal{O}_E$ is a finitely generated abelian group. More precisely,*

$$U(\mathcal{O}_E) \cong \mathbb{Z}^{r_1+r_2-1} \times T,$$

*where $T$ is a finite cyclic group consisting of the roots of unity in $E$.*

**Proof.** Borevich–Shafarevich [**23**], p. 112. •

## Exercises

**C-5.35.** (i) If $E = \mathbb{Q}(\sqrt{-3})$, prove that the only units in $\mathcal{O}_E$ are

$$\pm 1, \quad \tfrac{1}{2}(1 \pm \sqrt{-3}), \quad \tfrac{1}{2}(-1 \pm \sqrt{-3}).$$

(ii) Let $d$ be a negative squarefree integer with $d \neq -1$ and $d \neq -3$. If $E = \mathbb{Q}(\sqrt{d})$, prove that the only units in $\mathcal{O}_E$ are $\pm 1$.

**C-5.36.** (i) Prove that if $E = \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{R}$, then there are no units $u \in \mathcal{O}_E$ with $1 < u < 1 + \sqrt{2}$.

(ii) If $E = \mathbb{Q}(\sqrt{2})$, prove that $\mathcal{O}_E$ has infinitely many units.
**Hint.** Use (i) to prove that all powers of $1 + \sqrt{2}$ are distinct.

**Definition.** If $\mathcal{O}_E$ is the ring of integers in an algebraic number field $E$, then a ***discriminant*** of $\mathcal{O}_E$ is

$$\Delta(\mathcal{O}_E) = \det[\mathrm{tr}(\alpha_i \alpha_j)],$$

where $\alpha_1, \ldots, \alpha_n$ is an integral basis of $\mathcal{O}_E$.

**C-5.37.** Let $d$ be a squarefree integer, and let $E = \mathbb{Q}(\sqrt{d})$.

(i) If $d \equiv 2 \bmod 4$ or $d \equiv 3 \bmod 4$, prove that $1, \sqrt{d}$ is an integral basis of $\mathcal{O}_E$, and prove that a discriminant of $\mathcal{O}_E$ is $4d$.

(ii) If $d \equiv 1 \bmod 4$, prove that $1, \frac{1}{2}(1 + \sqrt{d})$ is an integral basis of $\mathcal{O}_E$, and prove that a discriminant of $\mathcal{O}_E$ is $d$.

**C-5.38.** Let $p$ be an odd prime, and let $E = \mathbb{Q}(\zeta_p)$ be the cyclotomic field.

(i) Show that $1, 1 - \zeta_p, (1 - \zeta_p)^2, \ldots, (1 - \zeta_p)^{p-2}$ is an integral basis for $\mathcal{O}_E$.

(ii) Prove that a discriminant of $\mathcal{O}_E$ is $(-1)^{\frac{1}{2}(p-1)} p^{p-2}$.
**Hint.** See Pollard [**175**], p. 67.

∗ **C-5.39.** (i) If $\mathbb{A}$ is the field of all algebraic numbers, prove that $\mathcal{O}_{\mathbb{A}}$ is not noetherian.

(ii) Prove that every nonzero prime ideal in $\mathcal{O}_{\mathbb{A}}$ is a maximal ideal.
**Hint.** Use the proof of Corollary C-5.58.

### ▪ Characterizations of Dedekind Rings

The following definition involves some of the ring-theoretic properties enjoyed by the ring of integers $\mathcal{O}_E$ in an algebraic number field $E$.

**Definition.** A domain $R$ is a ***Dedekind ring*** if it is integrally closed, noetherian, and all its nonzero prime ideals are maximal ideals.

**Example C-5.84.**

(i) The ring $\mathcal{O}_E$ in an algebraic number field $E$ is a Dedekind ring, by Proposition C-5.51, Proposition C-5.73, and CorollaryC-5.58.

(ii) Every PID $R$ is a Dedekind ring (Proposition C-5.50 says that $R$ is integrally closed). ◀

It is shown, in Example C-5.74, that $R = \mathbb{Z}\left[\sqrt{-5}\right]$ is a Dedekind ring that is not a UFD and, hence, it is not a PID. We now characterize discrete valuation rings, and then show that localizations of Dedekind rings are well-behaved.

**Lemma C-5.85.** *A domain $R$ is a* DVR *if and only if it is noetherian, integrally closed, and has a unique nonzero prime ideal.*

**Proof.** If $R$ is a DVR, then it does have the required properties (recall that $R$ is a PID; hence it is integrally closed).

The converse, which requires us to show that $R$ is a PID, is not as simple as we might expect. Let $\mathfrak{p}$ be the nonzero prime ideal, and choose a nonzero $a \in \mathfrak{p}$. Define $M = R/Ra$, and consider the family $\mathcal{A}$ of all the annihilators $\operatorname{ann}(m)$ as $m$ varies over all the nonzero elements of $M$. Since $R$ is noetherian, it satisfies the maximum condition, and so there is a nonzero element $b + Ra \in M$ whose annihilator $\mathfrak{q} = \operatorname{ann}(b + Ra)$ is maximal in $\mathcal{A}$. We claim that $\mathfrak{q}$ is a prime ideal. Suppose that $x, y \in R$, $xy \in \mathfrak{q}$, and $x, y \notin \mathfrak{q}$. Then $y(b + Ra) = yb + Ra$ is a nonzero element of $M$, because $y \notin \mathfrak{q}$. But $\operatorname{ann}(yb + Ra) \supsetneq \operatorname{ann}(b + Ra)$, because $x \notin \operatorname{ann}(b + Ra)$, contradicting the maximality property of $\mathfrak{q}$. Therefore, $\mathfrak{q}$ is a prime ideal. Since $R$ has a unique nonzero prime ideal $\mathfrak{p}$, we have

$$\mathfrak{q} = \operatorname{ann}(b + Ra) = \mathfrak{p}.$$

Note that

$$b/a \notin R;$$

otherwise, $b + Ra = 0 + Ra$, contradicting $b + Ra$ being a nonzero element of $M = R/Ra$.

We now show that $\mathfrak{p}$ is principal, with generator $a/b$ (we do not yet know whether $a/b \in \operatorname{Frac}(R)$ lies in $R$). First, we have $\mathfrak{p}b = \mathfrak{q}b \subseteq Ra$, so that $\mathfrak{p}(b/a) \subseteq R$; that is, $\mathfrak{p}(b/a)$ is an ideal in $R$. If $\mathfrak{p}(b/a) \subseteq \mathfrak{p}$, then $b/a$ is integral over $R$, for $\mathfrak{p}$ is a finitely generated $R$-submodule of $\operatorname{Frac}(R)$, as required in Lemma C-5.46. As $R$ is integrally closed, this puts $b/a \in R$, contradicting what we noted at the end of the previous paragraph. Therefore, $\mathfrak{p}(b/a)$ is not a proper ideal, so that $\mathfrak{p}(b/a) = R$ and $\mathfrak{p} = R(a/b)$. It follows that $a/b \in R$ and $\mathfrak{p}$ is a principal ideal.

Denote $a/b$ by $t$. The proof is completed by showing that the only nonzero ideals in $R$ are the principal ideals generated by $t^n$, for $n \geq 0$. Let $I$ be a nonzero ideal in $R$, and consider the chain of submodules of $\operatorname{Frac}(R)$:

$$I \subseteq It^{-1} \subseteq It^{-2} \subseteq \cdots.$$

We claim that this chain is strictly increasing. If $It^{-n} = It^{-n-1}$, then the finitely generated $R$-module $It^{-1}$ satisfies $t^{-1}(It^{-n}) \subseteq It^{-n}$, so that $t^{-1} = b/a$ is integral over $R$. As above, $R$ integrally closed forces $b/a \in R$, a contradiction. Since $R$ is noetherian, this chain can contain only finitely many ideals in $R$. Thus, there is $n$ with $It^{-n} \subseteq R$ and $It^{-n-1} \not\subseteq R$. If $It^{-n} \subseteq \mathfrak{p} = Rt$, then $It^{-n-1} \subseteq R$, a contradiction. Therefore, $It^{-n} = R$ and $I = Rt^n$, as desired. •

**Proposition C-5.86.** *Let $R$ be a noetherian domain. Then $R$ is a Dedekind ring if and only if the localizations $R_{\mathfrak{p}}$ are* DVR*s for every nonzero prime ideal $\mathfrak{p}$.*

**Remark.** Exercise C-5.39 on page 467 shows that it is necessary to assume that $R$ is noetherian. ◀

**Proof.** If $R$ is a Dedekind ring and $\mathfrak{p}$ is a maximal ideal, Corollary C-5.20(iv) shows that $R_\mathfrak{p}$ has a unique nonzero prime ideal. Moreover, $R_\mathfrak{p}$ is noetherian (Corollary C-5.20(v)), a domain (Corollary C-5.18), and integrally closed (Exercise C-5.27 on page 453). By Lemma C-5.85, $R_\mathfrak{p}$ is a DVR.

For the converse, we must show that $R$ is integrally closed and that its nonzero prime ideals are maximal. Let $u/v \in \text{Frac}(R)$ be integral over $R$. For every nonzero prime ideal $\mathfrak{p}$, the element $u/v$ is integral over $R_\mathfrak{p}$ (note that $\text{Frac}(R_\mathfrak{p}) = \text{Frac}(R)$). But $R_\mathfrak{p}$ is a PID, hence integrally closed, and so $u/v \in R_\mathfrak{p}$. We conclude that $u/v \in \bigcap_\mathfrak{p} R_\mathfrak{p} = R$, by Proposition C-5.22. Therefore, $R$ is integrally closed.

Suppose there are nonzero prime ideals $\mathfrak{p} \subsetneq \mathfrak{q}$ in $R$. By Corollary C-5.20(iv), $\mathfrak{p}_\mathfrak{q} \subsetneq \mathfrak{q}_\mathfrak{q}$ in $R_\mathfrak{q}$, contradicting the fact that DVRs have unique nonzero prime ideals. Therefore, nonzero prime ideals are maximal, and $R$ is a Dedekind ring. •

Let $R$ be a domain with $Q = \text{Frac}(R)$, and let $I = Ra$ be a nonzero principal ideal in $R$. If we define $J = Ra^{-1} \subseteq Q$, the cyclic $R$-submodule generated by $a^{-1}$, then it is easy to see that

$$IJ = \{uv : u \in I \text{ and } v \in J\} = R.$$

**Definition.** If $R$ is a domain with $Q = \text{Frac}(R)$, then a ***fractional ideal*** is a finitely generated nonzero $R$-submodule $I$ of $Q$. If $I$ is a fractional ideal, define

$$I^{-1} = \{v \in Q : vI \subseteq R\}.$$

It is always true that $I^{-1}I \subseteq R$; a fractional ideal $I$ is ***invertible*** if $I^{-1}I = R$.

Every nonzero finitely generated ideal in $R$ is also a fractional ideal. In this context, we often call such ideals (which are the usual ideals!) ***integral ideals*** when we want to contrast them with more general fractional ideals.

We claim that if $I = Ra$ is a nonzero principal ideal in $R$, then $I^{-1} = Ra^{-1}$. Clearly, $(ra^{-1})(r'a) = rr' \in R$ for all $r' \in R$, so that $Ra^{-1} \subseteq I^{-1}$. For the reverse inclusion, suppose that $(u/v)a \in R$, where $u, v \in R$. Then $v \mid ua$ in $R$, so there is $r \in R$ with $rv = ua$. Hence, in $Q$, we have $u = rva^{-1}$, so that $u/v = (rva^{-1})/v = ra^{-1}$. Therefore, every nonzero principal ideal in $R$ is invertible.

**Lemma C-5.87.** *Let $R$ be a domain with $Q = \text{Frac}(R)$. Then a fractional ideal $I$ is invertible if and only if there exist $a_1, \ldots, a_n \in I$ and $q_1, \ldots, q_n \in Q$ with*

(i) $q_i I \subseteq R$ *for* $i = 1, \ldots, n$;
(ii) $1 = \sum_{i=1}^n q_i a_i$.

**Proof.** If $I$ is invertible, then $I^{-1}I = R$. Since $1 \in I^{-1}I$, there are $a_1, \ldots, a_n \in R$ and $q_1, \ldots, q_n \in I^{-1}$ with $1 = \sum_i q_i a_i$. Since $q_i \in I^{-1}$, we have $q_i I \subseteq R$.

To prove the converse, note that the $R$-submodule $J$ of $Q$ generated by $q_1, \ldots, q_n$ is a fractional ideal. Since $1 = \sum_{i=1}^n q_i a_i \in JI$, $JI$ is an $R$-submodule of $R$ containing 1; that is, $JI = R$. To see that $I$ is invertible, it remains to prove that $J = I^{-1}$. Clearly, each $q_i \in I^{-1}$, so that $J \subseteq I^{-1}$. For the reverse inclusion, assume that

$u \in Q$ and $uI \subseteq R$. Since $1 = \sum_i q_i a_i$, we have $u = \sum_i (ua_i)q_i \in J$ because $ua_i \in R$ for all $i$. •

**Remark.** This lemma leads to a homological charactierization of Dedekind rings; see Theorem C-5.94(ii). ◀

**Corollary C-5.88.** *Every invertible ideal $I$ in a domain $R$ is finitely generated.*

**Proof.** Since $I$ is invertible, there exist $a_1, \ldots, a_n \in I$ and $q_1, \ldots, q_n \in Q$ as in the lemma. If $b \in I$, then $b = b1 = \sum_i bq_i a_i \in I$, because $bq_i \in R$. Therefore, $I$ is generated by $a_1, \ldots, a_n \in I$. •

**Proposition C-5.89.** *The following conditions are equivalent for a domain $R$:*

(i) *$R$ is a Dedekind ring.*

(ii) *Every fractional ideal is invertible.*

(iii) *The set of all the fractional ideals $\mathcal{F}(R)$ forms an abelian group under multiplication of ideals.*

**Proof.** The structure of the proof is unusual, for we will need the equivalence of (ii) and (iii) in order to prove (iii) $\Rightarrow$ (i).

(i) $\Rightarrow$ (ii). Let $J$ be a fractional ideal in $R$. Since $R$ is a Dedekind ring, its localization $R_{\mathfrak{p}}$ is a PID, and so $J_{\mathfrak{p}}$, as every nonzero principal ideal, is invertible (in the course of proving that finitely generated torsion-free abelian groups are free abelian, we really proved that fractional ideals of PIDs are cyclic modules). Now Exercise C-5.44 on page 477 gives

$$(J^{-1}J)_{\mathfrak{p}} = (J^{-1})_{\mathfrak{p}} J_{\mathfrak{p}} = (J_{\mathfrak{p}})^{-1} J_{\mathfrak{p}} = R_{\mathfrak{p}}.$$

Proposition C-5.37 gives $J^{-1}J = R$, and so $J$ is invertible.

(ii) $\Leftrightarrow$ (iii). If $I, J \in \mathcal{F}(R)$, then they are finitely generated, by Corollary C-5.88, and

$$IJ = \left\{ \sum a_\ell b_\ell : a_\ell \in I \text{ and } b_\ell \in J \right\}$$

is a finitely generated $R$-submodule of $\mathrm{Frac}(R)$. If $I = (a_1, \ldots, a_n)$ and $J = (b_1, \ldots, b_m)$, then $IJ$ is generated by all $a_i b_j$. Hence, $IJ$ is finitely generated and $IJ \in \mathcal{F}(R)$. Associativity does hold, the identity is $R$, and the inverse of a fractional ideal $J$ is $J^{-1}$, because $J$ is invertible. It follows that $\mathcal{F}(R)$ is an abelian group.

Conversely, if $\mathcal{F}(R)$ is an abelian group and $I \in \mathcal{F}(R)$, then there is $J \in \mathcal{F}(R)$ with $JI = R$. We must show that $J = I^{-1}$. But

$$R = JI \subseteq I^{-1}I \subseteq R,$$

so that $JI = I^{-1}I$. Canceling $I$ in the group $\mathcal{F}(R)$ gives $J = I^{-1}$, as desired.

(iii) $\Rightarrow$ (i). First, $R$ is noetherian, for (iii) $\Rightarrow$ (ii) shows that every nonzero ideal $I$ is invertible, and Corollary C-5.88 shows that $I$ is finitely generated.

Second, we show that every nonzero prime ideal $\mathfrak{p}$ is a maximal ideal. Let $I$ be an ideal with $\mathfrak{p} \subsetneq I$ (we allow $I = R$). Then $\mathfrak{p}I^{-1} \subseteq II^{-1} = R$, so that $\mathfrak{p}I^{-1}$ is an (integral) ideal in $R$. Now $(\mathfrak{p}I^{-1})I = \mathfrak{p}$, because multiplication is associative in $\mathcal{F}(R)$. Since $\mathfrak{p}$ is a prime ideal, Proposition A-3.82 in Part 1

says that either $\mathfrak{p}I^{-1} \subseteq \mathfrak{p}$ or $I \subseteq \mathfrak{p}$. The second option does not hold, so that $\mathfrak{p}I^{-1} \subseteq \mathfrak{p}$. Multiplying by $\mathfrak{p}^{-1}I$ gives $R \subseteq I$. Therefore, $I = R$, and so $\mathfrak{p}$ is a maximal ideal.

Third, if $a \in \mathrm{Frac}(R)$ is integral over $R$, then Lemma C-5.46 gives a finitely generated $R$-submodule $J$ of $\mathrm{Frac}(R)$, i.e., a fractional ideal, with $aJ \subseteq J$. Since $J$ is invertible, there are $q_1, \ldots, q_n \in \mathrm{Frac}(R)$ and $a_1, \ldots, a_n \in J$ with $q_i J \subseteq R$ for all $i$ and $1 = \sum q_i a_i$. Hence, $a = \sum_i q_i a_i a$. But $a_i a \in J$ and $q_i J \subseteq R$ gives $a = \sum_i q_i(a_i a) \in R$. Therefore, $R$ is integrally closed, and hence it is a Dedekind ring. •

**Proposition C-5.90.**

(i) *If $R$ is a* UFD*, then a nonzero ideal $I$ in $R$ is invertible if and only if it is principal.*

(ii) *A Dedekind ring $R$ is a* UFD *if and only if it is a* PID*.*

**Proof.**

(i) We have already seen that every nonzero principal ideal is invertible. Conversely, if $I$ is invertible, there are elements $a_1, \ldots, a_n \in I$ and $q_1, \ldots, q_n \in \mathrm{Frac}(R)$ with $1 = \sum_i q_i a_i$ and $q_i I \subseteq R$ for all $i$. Let $q_i = b_i/c_i$, where $b_i, c_i \in R$. Since $R$ is a UFD, we may assume that $q_i$ is in lowest terms; that is, $(b_i, c_i) = 1$. But $(b_i/c_i)a_j \in R$ says that $c_i \mid b_i a_j$, so that $c_i \mid a_j$ for all $i, j$, by Exercise A-3.101 in Part 1. We claim that $I = Rc$, where $c = \mathrm{lcm}\{c_1, \ldots, c_n\}$. First, $c \in I$, for $cb_i/c_i \in R$ and $c = c1 = \sum_i (cb_i/c_i)a_i$. Hence, $Rc \subseteq I$. For the reverse inclusion, Exercise A-3.101 in Part 1 shows that $c \mid a_j$ for all $j$, so that $a_j \in Rc$, for all $j$, and so $I \subseteq Rc$.

(ii) Since every nonzero ideal in a Dedekind ring is invertible, it follows from (i) that if $R$ is a UFD, then every ideal in $R$ is principal. •

**Definition.** If $R$ is a Dedekind ring, then its ***class group*** $C(R)$ is defined by

$$C(R) = \mathcal{F}(R)/\mathcal{P}(R),$$

where $\mathcal{P}(R)$ is the subgroup of all nonzero principal ideals.

Dirichlet proved, for every algebraic number field $E$, that the class group of $C(\mathcal{O}_E)$ is finite; the order $|C(R)|$ is called the ***class number*** of $\mathcal{O}_E$. The usual proof of finiteness of the class number uses a geometric theorem of Minkowski which says that sufficiently large parallelepipeds in Euclidean space must contain lattice points (see Samuel [**196**], pp. 57–58).

Claborn [**40**] proved, for every (not necessarily finite) abelian group $G$, that there is a Dedekind ring $R$ with $C(R) \cong G$.

We can now prove the result linking Kummer and Dedekind.

**Theorem C-5.91.** *If $R$ is a Dedekind ring, then every proper nonzero ideal has a unique factorization as a product of prime ideals.*

**Remark.** Zariski and Samuel [**233**] define a Dedekind ring to be a domain in which every nonzero ideal is a product of prime ideals (and they then prove that such a

factorization must be unique). They prove the converse of this theorem on page 275 of their book. ◀

**Proof.** Let $\mathcal{S}$ be the family of all proper nonzero ideals in $R$ that are not products of prime ideals. If $\mathcal{S} = \varnothing$, then every nonzero ideal in $R$ is a product of prime ideals. If $\mathcal{S} \neq \varnothing$, then $\mathcal{S}$ has a maximal element $I$, because noetherian rings satisfy the maximum condition (Proposition B-1.10 in Part 1). Now $I$ cannot be a maximal ideal in $R$, for a "product of prime ideals" is allowed to have only one factor. Let $\mathfrak{m}$ be a maximal ideal containing $I$. Since $I \subsetneq \mathfrak{m}$, we have $\mathfrak{m}^{-1}I \subsetneq \mathfrak{m}^{-1}\mathfrak{m} = R$; that is, $\mathfrak{m}^{-1}I$ is a proper ideal properly containing $I$. Neither $\mathfrak{m}$ nor $\mathfrak{m}^{-1}I$ lies in $\mathcal{S}$, for each is strictly larger than a maximal element, namely, $I$, and so each of them is a product of prime ideals. Therefore, $I = \mathfrak{m}(\mathfrak{m}^{-1}I)$ (equality holding because $R$ is a Dedekind ring) is a product of prime ideals, contradicting $I$ being in $\mathcal{S}$. Therefore, $\mathcal{S} = \varnothing$, and every proper nonzero ideal in $R$ is a product of prime ideals.

Suppose that $\mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s$, where the $\mathfrak{p}_i$ and $\mathfrak{q}_j$ are prime ideals. We prove unique factorization by induction on $\max\{r, s\}$. The base step $r = 1 = s$ is obviously true. For the inductive step, note that $\mathfrak{p}_1 \supseteq \mathfrak{q}_1 \cdots \mathfrak{q}_s$, so that Proposition A-3.82 in Part 1 gives $\mathfrak{q}_j$ with $\mathfrak{p}_1 \supseteq \mathfrak{q}_j$. Hence, $\mathfrak{p}_1 = \mathfrak{q}_j$, because prime ideals are maximal. Now multiply the original equation by $\mathfrak{p}_1^{-1}$ and use the inductive hypothesis. •

**Corollary C-5.92.** *If $R$ is a Dedekind ring, then $\mathcal{F}(R)$ is a free abelian group with basis all the nonzero prime ideals.*

**Proof.** Of course, $\mathcal{F}(R)$ is written multiplicatively. That every fractional ideal is a product of primes shows that the set of primes generates $\mathcal{F}(R)$; uniqueness of the factorization says the set of primes is a basis. •

In light of Theorem C-5.91, many of the usual formulas of arithmetic extend to ideals in Dedekind rings. Observe that in $\mathbb{Z}$, the ideal (3) contains (9). In fact, $\mathbb{Z}m \supseteq \mathbb{Z}n$ if and only if $m \mid n$. We will now see that the relation "contains" for ideals is the same as "divides", and that the usual formulas for gcd's and lcm's generalize to Dedekind rings.

**Proposition C-5.93.** *Let $I$ and $J$ be nonzero ideals in a Dedekind ring $R$, and let their prime factorizations be*

$$I = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_n^{e_n} \quad \text{and} \quad J = \mathfrak{p}_1^{f_1} \cdots \mathfrak{p}_n^{f_n},$$

*where $e_i \geq 0$ and $f_i \geq 0$ for all $i$.*

(i) *$J \supseteq I$ if and only if $I = JL$ for some ideal $L$.*

(ii) *$J \supseteq I$ if and only if $f_i \leq e_i$ for all $i$.*

(iii) *If $m_i = \min\{e_i, f_i\}$ and $M_i = \max\{e_i, f_i\}$, then*

$$I \cap J = \mathfrak{p}_1^{M_1} \cdots \mathfrak{p}_n^{M_n} \quad \text{and} \quad I + J = \mathfrak{p}_1^{m_1} \cdots \mathfrak{p}_n^{m_n}.$$

*In particular, $I + J = R$ if and only if $\min\{e_i, f_i\} = 0$ for all $i$.*

(iv) *Let $R$ be a Dedekind ring, and let $I = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_n^{e_n}$ be a nonzero ideal in $R$. Then*

$$R/I = R/\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_n^{e_n} \cong \left(R/\mathfrak{p}_1^{e_1}\right) \times \cdots \times \left(R/\mathfrak{p}_n^{e_n}\right).$$

**Proof.**

(i) If $I \subseteq J$, then $J^{-1}I \subseteq R$, and

$$J(J^{-1}I) = I.$$

Conversely, if $I = JL$, then $I \subseteq J$ because $JL \subseteq JR = J$.

(ii) This follows from (i) and the unique factorization of nonzero ideals as products of prime ideals.

(iii) We prove the formula for $I+J$. Let $I+J = \mathfrak{p}_1^{r_1} \cdots \mathfrak{p}_n^{r_n}$ and let $A = \mathfrak{p}_1^{m_1} \cdots \mathfrak{p}_n^{m_n}$. Since $I \subseteq I + J$ and $J \subseteq I + J$, we have $r_i \leq e_i$ and $r_i \leq f_i$, so that $r_i \leq \min\{e_i, f_i\} = m_i$. Hence, $A \subseteq I + J$. For the reverse inclusion, $A \subseteq I$ and $A \subseteq J$, so that $A = II'$ and $A = JJ'$ for ideals $I'$ and $J'$, by (i). Therefore, $I + J = AI' + AJ' = A(I' + J')$, and so $I + J \subseteq A$. The proof of the formula for $IJ$ is left to the reader.

(iv) This is just the Chinese Remainder Theorem, so that it suffices to verify the hypothesis that $\mathfrak{p}_i^{e_i}$ and $\mathfrak{p}_j^{e_j}$ are coprime when $i \neq j$; that is, $\mathfrak{p}_i^{e_i} + \mathfrak{p}_j^{e_j} = R$. But this follows from (iii). •

Recall Proposition B-4.46 in Part 1: an $R$-module $A$ is projective if and only if it has a *projective basis*: there exist elements $\{a_j : j \in J\} \subseteq A$ and $R$-maps $\{\varphi_j \colon A \to R : j \in J\}$ such that

(i) for each $x \in A$, almost all $\varphi_j(x) = 0$;

(ii) for each $x \in A$, we have $x = \sum_{j \in J} (\varphi_j x) a_j$.

The coming characterizations of Dedekind rings, Theorems C-5.94, C-5.97, and C-5.106, are useful in homological algebra.

**Theorem C-5.94.**

(i) *A nonzero ideal $I$ in a domain $R$ is invertible if and only if $I$ is a projective $R$-module.*

(ii) *A domain $R$ is a Dedekind ring if and only if every ideal in $R$ is projective.*

**Proof.**

(i) If $I$ is invertible, there are elements $a_1, \ldots, a_n \in I$ and $q_1, \ldots, q_n \in \operatorname{Frac}(R)$ with $1 = \sum_i q_i a_i$ and $q_i I \subset R$ for all $i$. Define $\varphi_i \colon I \to R$ by $\varphi_i \colon a \mapsto q_i a$ (note that $\operatorname{im} \varphi_i \subseteq I$ because $q_i I \subseteq R$). If $a \in I$, then

$$\sum_i \varphi_i(a) a_i = \sum_i q_i a a_i = a \sum_i q_i a_i = a.$$

Therefore, $I$ has a projective basis, and so $I$ is a projective $R$-module.

Conversely, if an ideal $I$ is a projective $R$-module, it has a projective basis $\{\varphi_j : j \in J\}, \{a_j : j \in J\}$. If $b \in I$ is nonzero, define $q_j \in \operatorname{Frac}(R)$ by

$$q_j = \varphi_j(b)/b.$$

This element does not depend on the choice of nonzero $b$: if $b' \in I$ is nonzero, then $b'\varphi_j(b) = \varphi_j(b'b) = b\varphi_j(b')$, so that $\varphi_j(b)/b = \varphi_j(b')/b'$. To see that $q_j I \subseteq R$, note that if $b \in I$ is nonzero, then $q_j b = \big(\varphi_j(b)/b\big)b = \varphi_j(b) \in R$. By item (i) in the definition of projective basis, almost all $\varphi_j(b) = 0$, and so there are only finitely many nonzero $q_j = \varphi_j(b)/b$ (remember that $q_j$ does not depend on the choice of nonzero $b \in I$). Item (ii) in the definition of projective basis gives, for $b \in I$,

$$b = \sum_j \varphi_j(b)a_j = \sum_j (q_j b)a_j = b\Big(\sum_j q_j a_j\Big).$$

Canceling $b$ gives $1 = \sum_j q_j a_j$. Finally, the set of those $a_j$ with indices $j$ for which $q_j \neq 0$ completes the data necessary to show that $I$ is an invertible ideal.

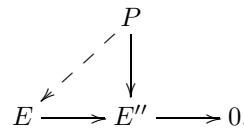
(ii) This follows at once from (i) and Proposition C-5.89. •

**Example C-5.95.** We have seen that $R = \mathbb{Z}\big[\sqrt{-5}\,\big]$ is a Dedekind ring that is not a PID. Any nonprincipal ideal gives an example of a projective $R$-module that is not free. ◄

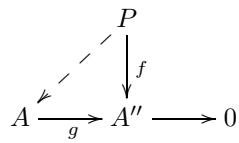
**Remark.** Noncommutative analogs of Dedekind rings are called *left hereditary*. A ring $R$ is ***left hereditary*** if every left ideal is a projective $R$-module. (Small's example of a right noetherian ring that is not left noetherian (Exercise B-1.28 on page 288 in Part 1) is right hereditary but not left hereditary.) Aside from Dedekind rings, some examples of left hereditary rings are semisimple rings, noncommutative principal ideal rings, and FIRs (***free ideal rings***—all left ideals are free $R$-modules). Cohn proved (see [**42**]) that if $k$ is a field, then free $k$-algebras $k\langle X\rangle$ (polynomials in noncommuting variables $X$) are FIRs, and so there exist left hereditary rings that are not left noetherian. ◄

Projective and injective modules over a Dedekind ring are well-behaved.

**Lemma C-5.96.** *A left $R$-module $P$ (over any ring $R$) is projective if and only if every diagram below with $E$ injective can be completed to a commutative diagram. The dual characterization of injective modules is also true,*

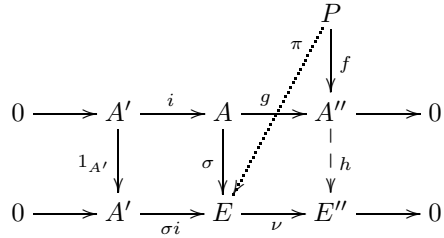



**Proof.** If $P$ is projective, then the diagram can be completed for every not necessarily injective module $E$. Conversely, we must show that the diagram




can be completed for any module $A$ and any surjection $g\colon A \to A''$. By Theorem B-4.64 in Part 1, there is an injective $R$-module $E$ and an injection $\sigma\colon A \to E$.

Define $E'' = \operatorname{coker} \sigma i = E/\operatorname{im} \sigma i$, and consider the commutative diagram with exact rows




where $\nu\colon E \to E'' = \operatorname{coker} \sigma i$ is the natural map and $h\colon A'' \to E''$ exists, by Proposition B-1.46 in Part 1. By hypothesis, there exists a map $\pi\colon P \to E$ with $\nu\pi = hf$. We claim that $\operatorname{im} \pi \subseteq \operatorname{im} \sigma$. For $x \in P$, surjectivity of $g$ gives $a \in A$ with $ga = fx$. Then $\nu\pi x = hfx = hga = \nu\sigma a$, and so $\pi x - \sigma a \in \ker \nu = \operatorname{im} \sigma i$; hence, $\pi x - \sigma a = \sigma i a'$ for some $a' \in A'$, and so $\pi x = \sigma(a + ia') \in \operatorname{im} \sigma$. Therefore, if $x \in P$, there is a unique $a \in A$ with $\sigma a = \pi x$ ($a$ is unique because $\sigma$ is an injection). Thus, there is a well-defined function $\pi'\colon P \to A$, given by $\pi' x = a$, where $\sigma a = \pi x$. The reader may check that $\pi'$ is an $R$-map and that $g\pi' = f$. •

**Theorem C-5.97 (Cartan–Eilenberg).** *The following conditions are equivalent for a domain $R$:*

(i) *$R$ is a Dedekind ring.*

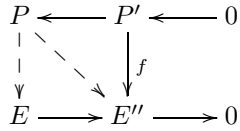
(ii) *Every submodule of a projective $R$-module is projective.*

(iii) *Every quotient of an injective $R$-module is injective.*

**Proof.**

(i) $\Leftrightarrow$ (ii). If $R$ is Dedekind, then we can adapt the proof of Theorem B-2.28 in Part 1 (which proves that every subgroup of a free abelian group is free abelian) to prove that every submodule of a free $R$-module is projective (Exercise C-5.41 on page 476); in particular, every submodule of a projective $R$-module is projective. Conversely, since $R$ itself is a projective $R$-module, its submodules are also projective, by hypothesis; that is, the ideals of $R$ are projective. Theorem C-5.94 now shows that $R$ is a Dedekind ring.

(ii) $\Leftrightarrow$ (iii). Assume (iii), and consider the diagram with exact rows




where $P$ is projective and $E$ is injective; note that the hypothesis gives $E''$ injective. To prove projectivity of $P'$, it suffices, by Lemma C-5.96, to find a map $P' \to E$ making the diagram commute. Since $E''$ is injective, there exists a map $P \to E''$ giving commutativity. Since $P$ is projective, there is a map $P \to E$ also giving commutativity. The composite $P' \to P \to E$ is the desired map.

The converse is the dual of this, using the dual of Lemma C-5.96. •

The next result generalizes Exercise C-3.37 on page 291 from PIDs to Dedekind rings.

**Corollary C-5.98.** *Let $R$ be a Dedekind ring.*

(i) *For all $R$-modules $C$ and $A$ and all $n \geq 2$,*

$$\operatorname{Ext}_R^n(C, A) = \{0\} \quad \text{and} \quad \operatorname{Tor}_n^R(C, A) = \{0\}.$$

(ii) *Let $0 \to A' \to A \to A'' \to 0$ be a short exact sequence. For every module $C$, there are exact sequences*

$$\begin{aligned} 0 \to \operatorname{Hom}(C, A') &\to \operatorname{Hom}(C, A) \to \operatorname{Hom}(C, A'') \\ &\to \operatorname{Ext}^1(C, A') \to \operatorname{Ext}^1(C, A) \to \operatorname{Ext}^1(C, A'') \to 0 \end{aligned}$$

*and*

$$\begin{aligned} 0 \to \operatorname{Tor}_1^R(C, A') &\to \operatorname{Tor}_i^R(C, A) \to \operatorname{Tor}_1^R(C, A'') \\ &\to C \otimes_R A' \to C \otimes_R A \to C \otimes_R A'' \to 0. \end{aligned}$$

**Proof.**

(i) By definition, if $\cdots \to P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \to C \to 0$ is a projective resolution of $C$, then $\operatorname{Ext}^n(C, A) = \ker d_{n+1}^* / \operatorname{im} d_n^*$; moreover, $\operatorname{Ext}^n(C, A)$ does not depend on the choice of projective resolution, by Corollary C-3.74. Now $C$ is a quotient of a free module $F$, and so there is an exact sequence

$$0 \to K \to F \xrightarrow{\varepsilon} C \to 0, \tag{4}$$

where $K = \ker \varepsilon$. Since $R$ is Dedekind, the submodule $K$ of the free $R$-module $F$ is projective, so that (3) defines a projective resolution of $C$ with $P_0 = F$, $P_1 = K$, and $P_n = \{0\}$ for all $n \geq 2$. Hence, $\ker d_{n+1}^* \subseteq \operatorname{Hom}(P_n, A) = \{0\}$ for all $n \geq 2$, and so $\operatorname{Ext}_R^n(C, A) = \{0\}$ for all $A$ and for all $n \geq 2$. A similar argument works for Tor.

(ii) This follows from Corollary C-3.68 and Corollary C-3.57, the long exact sequence for Ext and for Tor, respectively. •

We will use this result in the next subsection to generalize Proposition C-3.93.

## Exercises

∗ **C-5.40.** Let $R$ be a commutative ring and let $M$ be a finitely generated $R$-module. Prove that if $IM = M$ for some ideal $I$ of $R$, then there exists $a \in I$ with $(1 - a)M = \{0\}$.

**Hint.** If $M = \langle m_1 \ldots, m_n \rangle$, then each $m_i = \sum_j a_{ij} m_j$, where $a_{ij} \in I$. Use the adjoint matrix (the matrix of cofactors) as in the proof of Lemma C-5.46.

∗ **C-5.41.** Generalize the proof of Theorem B-2.28 in Part 1 to prove that if $R$ is a left hereditary ring, then every submodule of a free left $R$-module $F$ is isomorphic to a direct sum of ideals and hence is projective.

∗ **C-5.42.** Let $R$ be a Dedekind ring, and let $\mathfrak{p}$ be a nonzero prime ideal in $R$.

(i) If $a \in \mathfrak{p}$, prove that $\mathfrak{p}$ occurs in the prime factorization of $Ra$.

(ii) If $a \in \mathfrak{p}^e$ and $a \notin \mathfrak{p}^{e+1}$, prove that $\mathfrak{p}^e$ occurs in the prime factorization of $Ra$, but that $\mathfrak{p}^{e+1}$ does not occur in the prime factorization of $Ra$.

**C-5.43.** Let $I$ be a nonzero ideal in a Dedekind ring $R$. Prove that if $\mathfrak{p}$ is a prime ideal, then $I \subseteq \mathfrak{p}$ if and only if $\mathfrak{p}$ occurs in the prime factorization of $I$.

∗ **C-5.44.** If $J$ is a fractional ideal of a Dedekind ring $R$, prove that $(J^{-1})_{\mathfrak{p}} = (J_{\mathfrak{p}})^{-1}$ for every maximal ideal $\mathfrak{p}$.

∗ **C-5.45.** Let $I_1, \ldots, I_n$ be ideals in a Dedekind ring $R$. If there is no nonzero prime ideal $\mathfrak{p}$ with $I_i = \mathfrak{p}L_i$ for all $i$ for ideals $L_i$, then

$$I_1 + \cdots + I_n = R.$$

**C-5.46.** Give an example of a projective $\mathbb{Z}\big[\sqrt{-5}\,\big]$-module that is not free.

**Hint.** See Example C-5.95.

∗ **C-5.47.** (i) A commutative ring $R$ is called a ***principal ideal ring*** if every ideal in $R$ is a principal ideal ($R$ would be a PID if it were a domain). For example, $\mathbb{I}_n$ is a principal ideal ring. Prove that $\mathbb{Z} \times \mathbb{Z}$ is a principal ideal ring.

(ii) Let $I_1, \ldots, I_n$ be pairwise coprime ideals in a commutative ring $R$. If $R/I_i$ is a principal ideal ring for each $i$, prove that $R/(I_1 \cdots I_n)$ is a principal ideal ring.
**Hint.** Use the Chinese Remainder Theorem.

**C-5.48.** Let $a$ be a nonzero element in a Dedekind ring $R$. Prove that there are only finitely many ideals $I$ in $R$ containing $a$.

**Hint.** If $a \in I$, then $Ra = IL$ for some ideal $L \subseteq R$.

## ▪ Finitely Generated Modules over Dedekind Rings

We saw in Part 1 that theorems about abelian groups generalize to theorems about modules over PIDs. We are now going to see that such theorems can be further generalized to modules over Dedekind rings.

**Proposition C-5.99.** *Let $R$ be a Dedekind ring.*

(i) *If $I \subseteq R$ is a nonzero ideal, then every ideal in $R/I$ is principal.*

(ii) *Every fractional ideal $J$ can be generated by two elements. More precisely, for any nonzero $a \in J$, there exists $b \in J$ with $J = (a, b)$.*

**Proof.**

(i) Let $I = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_n^{e_n}$ be the prime factorization of $I$. Since the ideals $\mathfrak{p}_i^{e_i}$ are pairwise coprime, it suffices, by Exercise C-5.47, to prove that $R/\mathfrak{p}_i^{e_i}$ is a principal ideal ring for each $i$. Now right exactness of $R_{\mathfrak{p}_i} \otimes_R \square$ shows that $(R/\mathfrak{p}_i^{e_i})_{\mathfrak{p}_i} \cong R_{\mathfrak{p}_i}/(\mathfrak{p}_i^{e_i})_{\mathfrak{p}_i}$. Since $R_{\mathfrak{p}_i}$ is a PID (it is even a DVR, by Proposition C-5.86), any quotient ring of it is a principal ideal ring.

(ii) Assume first that $J$ is an integral ideal. Choose a nonzero $a \in J$. By (i), the ideal $J/Ra$ in $R/Ra$ is principal; say, $J/Ra$ is generated by $b + Ra$, where $b \in J$. It follows that $J = (a, b)$.

For the general case, there is a nonzero $c \in R$ with $cJ \subseteq R$ (if $J$ is generated by $u_1/v_1, \ldots, u_m/v_m$, take $c = \prod_J v_j$). Since $cJ$ is an integral ideal, given any nonzero $a \in J$, there is $cb \in cJ$ with $cJ = (ca, cb)$. It follows that $J = (a, b)$. •

The next corollary says that we can force nonzero ideals to be coprime.

**Corollary C-5.100.** *If $I$ and $J$ are fractional ideals over a Dedekind ring $R$, then there are $a, b \in \mathrm{Frac}(R)$ with*

$$aI + bJ = R.$$

**Proof.** Choose a nonzero $a \in I^{-1}$. Now $aI \subseteq I^{-1}I = R$, so that $aIJ^{-1} \subseteq J^{-1}$. By Proposition C-5.99(ii), there is $b \in J^{-1}$ with

$$J^{-1} = aIJ^{-1} + Rb.$$

Since $b \in J^{-1}$, we have $bJ \subseteq R$, and so

$$R = JJ^{-1} = J(aIJ^{-1} + Rb) = aI + RbJ = aI + bJ. \quad \bullet$$

Let us now investigate the structure of $R$-modules. Recall that if $R$ is a domain and $M$ is an $R$-module, then $\mathrm{rank}(M) = \dim_Q(Q \otimes_R M)$, where $Q = \mathrm{Frac}(R)$. We can restate Lemma C-5.1(iii) in our present language. If $R$ is a domain and $M$ is a nonzero finitely generated torsion-free $R$-module, then $\mathrm{rank}(M) = 1$ if and only if $M$ is isomorphic to a nonzero integral ideal.

**Proposition C-5.101.** *If $R$ is a Dedekind ring and $M$ is a finitely generated torsion-free $R$-module, then*

$$M \cong I_1 \oplus \cdots \oplus I_n,$$

*where $I_i$ is an ideal in $R$.*

**Proof.** The proof is by induction on $\mathrm{rank}(M) \geq 0$. If $\mathrm{rank}(M) = 0$, then $M$ is torsion, by Lemma C-5.1(ii). Since $M$ is torsion-free, $M = \{0\}$. Assume now that $\mathrm{rank}(M) = n + 1$. Choose a nonzero $m \in M$, so that $\mathrm{rank}(Rm) = 1$. The sequence

$$0 \to Rm \to M \xrightarrow{\nu} M'' \to 0$$

is exact, where $M'' = R/Rm$ and $\nu$ is the natural map. Note that $\mathrm{rank}(M'') = n$, by Lemma C-5.1(i). Now $M$ finitely generated implies that $M''$ is also finitely generated. If $T = t(M'')$ is the torsion submodule of $M''$, then $M''/T$ is a finitely generated torsion-free $R$-module with $\mathrm{rank}(M''/T) = \mathrm{rank}(M'') = n$, because $\mathrm{rank}(T) = 0$. By induction, $M''/T$ is a direct sum of ideals, hence is projective, by Theorem C-5.94. Define

$$M' = \nu^{-1}(T) = \{m \in M : rm \in Rm \text{ for some } r \neq 0\} \subseteq M.$$

There is an exact sequence $0 \to M' \to M \to M''/T \to 0$; this sequence splits, by Corollary B-2.26 in Part 1, because $M''/T$ is projective; that is, $M \cong M' \oplus (M''/T)$. Hence

$$\operatorname{rank}(M') = \operatorname{rank}(M) - \operatorname{rank}(M''/T) = 1.$$

Since $R$ is noetherian, every submodule of a finitely generated $R$-module is itself finitely generated; hence, $M'$ is finitely generated. Therefore, $M'$ is isomorphic to an ideal, by Lemma C-5.1(ii), and this completes the proof. •

**Corollary C-5.102.** *If $R$ is a Dedekind ring and $M$ is a finitely generated torsion-free $R$-module, then $M$ is projective.*

**Proof.** By Theorem C-5.94, every ideal in a Dedekind ring is projective. It now follows from Proposition C-5.101 that $M$ is a direct sum of ideals, and hence it is projective. •

Corollary C-5.102 can also be proved by localization. For every maximal ideal $\mathfrak{m}$, the $R_{\mathfrak{m}}$-module $M_{\mathfrak{m}}$ is finitely generated torsion-free. Since $R_{\mathfrak{m}}$ is a PID (even a DVR), $M_{\mathfrak{m}}$ is a free module, and hence it is projective. The result now follows from Corollary C-5.43.

**Corollary C-5.103.** *If $M$ is a finitely generated $R$-module, where $R$ is a Dedekind ring, then the torsion submodule $tM$ is a direct summand of $M$.*

**Proof.** The quotient module $M/tM$ is a finitely generated torsion-free $R$-module, so that it is projective, by Corollary C-5.102. Therefore, $tM$ is a direct summand of $M$, by Corollary B-2.26 in Part 1. •

**Corollary C-5.104.** *Let $R$ be a Dedekind ring. Then an $R$-module $A$ is flat if and only if it is torsion-free.*

**Proof.** By Lemma B-4.103 in Part 1, it suffices to prove that every finitely generated submodule of $A$ is flat. But such submodules are torsion-free, hence projective, and projective modules are always flat, by Lemma B-4.101 in Part 1. The converse is Corollary B-4.104 in Part 1: every flat $R$-module over a domain $R$ is torsion-free. •

**Corollary C-5.105.** *Let $R$ be a Dedekind ring with $Q = \operatorname{Frac}(R)$.*

(i) *If $C$ is a torsion-free $R$-module and $T$ is a torsion module with* $\operatorname{ann}(T) \neq (0)$ *(that is, there is a nonzero $r \in R$ with $rT = \{0\}$), then* $\operatorname{Ext}^1_R(C, T) = \{0\}$.

(ii) *Let $M$ be an $R$-module. If* $\operatorname{ann}(tM) \neq (0)$, *where $tM$ is the torsion submodule of $M$, then $tM$ is a direct summand of $M$.*

**Proof.**

(i) We generalize the proof of Proposition C-3.93. Since $C$ is torsion-free, it is a flat $R$-module, by Corollary C-5.104, so that exactness of $0 \to R \to Q$ gives exactness of $0 \to R \otimes_R C \to Q \otimes_R C$. Thus, $C \cong R \otimes_R C$ can be imbedded in

a vector space $V$ over $Q$; namely, $V = Q \otimes_R C$. Applying the contravariant functor $\mathrm{Hom}_R(\square, T)$ to $0 \to C \to V \to V/C \to 0$ gives an exact sequence

$$\mathrm{Ext}^1_R(V,T) \to \mathrm{Ext}^1_R(C,T) \to \mathrm{Ext}^2(V/C,T).$$

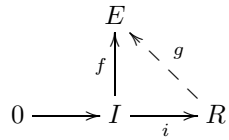
Now the last term is $\{0\}$, by Corollary C-5.98, and $\mathrm{Ext}^1_R(V,T)$ is (torsion-free) divisible, by (a straightforward generalization of) Example C-3.70; hence, $\mathrm{Ext}^1_R(C,T)$ is divisible. Since $\mathrm{ann}(T) \neq (0)$, Exercise C-3.39 on page 292 gives $\mathrm{Ext}^1_R(C,T) = \{0\}$.

(ii) To prove that the extension $0 \to tM \to M \to M/tM \to 0$ splits, it suffices to prove that $\mathrm{Ext}^1_R(M/tM, tM) = \{0\}$. Since $M/tM$ is torsion-free, this follows from part (i) and Corollary C-3.90. •

The next result generalizes Proposition B-4.61 in Part 1.

**Theorem C-5.106.** *Let $R$ be a domain. Then $R$ is Dedekind if and only if divisible $R$-modules are injective.*

**Proof.** Let $R$ be a Dedekind ring and let $E$ be a divisible $R$-module. By the Baer criterion, Theorem B-4.57 in Part 1, it suffices to complete the diagram




where $I$ is an ideal and $i\colon I \to R$ is the inclusion. Of course, we may assume that $I$ is nonzero, and so $I$ is invertible: there are elements $a_1, \ldots, a_n \in I$ and elements $q_1, \ldots, q_n \in Q$ with $q_i I \subseteq R$ and $1 = \sum_i q_i a_i$. Since $E$ is divisible, there are elements $e_i \in E$ with $f(a_i) = a_i e_i$. Note, for every $b \in I$, that

$$f(b) = f\Big(\sum_i q_i a_i b\Big) = \sum_i (q_i b) f(a_i) = \sum_i (q_i b) a_i e_i = b \sum_i (q_i a_i) e_i.$$

Hence, if we define $e = \sum_i (q_i a_i) e_i$, then $e \in E$ and $f(b) = be$ for all $b \in I$. Defining $g\colon R \to E$ by $g(r) = re$ shows that the diagram can be completed, and so $E$ is injective.

Conversely, if $E$ is an injective $R$-module, then $E$ is divisible (that every injective $R$-module is divisible was proved (for arbitrary domains $R$) in Lemma B-4.60 in Part 1). If $E'$ is a quotient of $E$, then $E'$ is divisible and, hence, is injective, by hypothesis. Therefore, every quotient of an injective module is injective, and so $R$ is Dedekind, by Theorem C-5.97. •

Having examined torsion-free modules, let us now look at torsion modules.

**Proposition C-5.107.** *Let $\mathfrak{p}$ be a nonzero prime ideal in a Dedekind ring $R$. If $M$ is an $R$-module with $\mathrm{ann}(M) = \mathfrak{p}^e$ for some $e > 0$, then the localization map $M \to M_{\mathfrak{p}}$ is an isomorphism (and hence $M$ may be regarded as an $R_{\mathfrak{p}}$-module).*

**Proof.** It suffices to prove that $M \cong R_{\mathfrak{p}} \otimes_R M$. If $m \in M$ is nonzero and $s \in R$ with $s \notin \mathfrak{p}$, then

$$\mathfrak{p}^e + Rs = R,$$

by Proposition C-5.93. Hence, there exist $u \in \mathfrak{p}^e$ and $r \in R$ with $1 = u + rs$, and so

$$m = um + rsm = rsm.$$

If $1 = u' + r's$, where $u' \in \mathfrak{p}$ and $r' \in R$, then $s(r - r')m = 0$, so that

$$s(r - r') \in \operatorname{ann}(m) = \mathfrak{p}^e.$$

Since $s \notin \mathfrak{p}^e$, it follows that $r - r' \in \mathfrak{p}^e$ (see Exercise C-5.42 on page 477; the prime factorization of $Rs$ does not contain $\mathfrak{p}^e$; if the prime factorization of $R(r - r')$ does not contain $\mathfrak{p}^e$, then neither does the prime factorization of $Rs(r - r')$). Hence, $rm = r'm$. Define $s^{-1}m = rm$, and define $f\colon R_{\mathfrak{p}} \times M \to M$ by $f(r/s, m) = s^{-1}(rm)$. It is straightforward to check that $f$ is $R$-bilinear, and so there is an $R$-map $\widetilde{f}\colon R_{\mathfrak{p}} \otimes M \to M$ with $\widetilde{f}\colon (r/s \otimes m) = s^{-1}rm$. In particular, $\widetilde{f}(1 \otimes m) = m$, so that $\widetilde{f}$ is surjective. On the other hand, the localization map $h_M\colon M \to R_{\mathfrak{p}} \otimes_R M$, defined by $h_M(m) = 1 \otimes m$, is easily seen to be the inverse of $\widetilde{f}$. •

**Definition.** Let $\mathfrak{p}$ be a nonzero prime ideal in a Dedekind ring $R$. An $R$-module $M$ is called $\mathfrak{p}$-***primary*** if, for each $m \in M$, there is $e > 0$ with $\operatorname{ann}(m) = \mathfrak{p}^e$.

**Theorem C-5.108 (Primary Decomposition).** *Let $R$ be a Dedekind ring, and let $T$ be a finitely generated torsion $R$-module. If $I = \operatorname{ann}(T) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_n^{e_n}$, then*

$$T = T[\mathfrak{p}_1] \oplus \cdots \oplus T[\mathfrak{p}_n],$$

*where*

$$T[\mathfrak{p}_i] = \{m \in T : \operatorname{ann}(m) \text{ is a power of } \mathfrak{p}_i\}.$$

*$T[\mathfrak{p}_i]$ is called the $\mathfrak{p}_i$-**primary component** of $T$.*

**Proof.** It is easy to see that the $\mathfrak{p}$-primary components $T[\mathfrak{p}]$ are submodules of $T$. By Proposition C-5.93(iv), there is an isomorphism of rings

$$f : \frac{R}{I} \cong \frac{R}{\mathfrak{p}_1^{e_1}} \oplus \cdots \oplus \frac{R}{\mathfrak{p}_n^{e_n}}.$$

Let $a_i \in R$ be an element such that $f(a_i + I)$ has 1 in the $i^{\text{th}}$ coordinate and 0 in all other coordinates. Then $a_i a_j \in I$ for all $j \neq i$ and there is $r_i \in I$ with $a_i = 1 + r_i$. We will show that $a_i T = T[\mathfrak{p}_i]$. Since $a_i \in \mathfrak{p}_j^{e_j}$ for $j \neq i$ it follows that $\mathfrak{p}_i^{e_i} a_i \subseteq I$ and so $\mathfrak{p}_i^{e_i}$ annihilates $a_i T$. Thus, $a_i T \subseteq T[\mathfrak{p}_i]$.

To prove the reverse inclusion, let $x \in T[\mathfrak{p}_i]$. We may write $1 \in R$ as

$$1 = \sum_1^n a_i + r \quad \text{with } r \in I.$$

It follows that $x = 1x = \sum a_i x$ since $r \in I$. But for $i \neq j$, $a_j x \in T[\mathfrak{p}_j] \cap T[\mathfrak{p}_i]$ since $x \in T[\mathfrak{p}_i]$ and $a_j T \subseteq T[\mathfrak{p}_j]$. The intersection is (0) as $\mathfrak{p}_i$ and $\mathfrak{p}_j$ are coprime. Thus, $x = a_i x$ so $x \in a_i T$ and $T[\mathfrak{p}_i] = a_i T$.

The direct sum decomposition of $R/I$ gives rise to the direct sum

$$T = a_1 T \oplus \cdots \oplus a_n T = T[\mathfrak{p}_1] \oplus \cdots \oplus T[\mathfrak{p}_n]. \quad \bullet$$

**Theorem C-5.109.** *Let $R$ be a Dedekind ring.*

(i) *Two finitely generated torsion $R$-modules $T$ and $T'$ are isomorphic if and only if $T[\mathfrak{p}] \cong T'[\mathfrak{p}]$ for all primes $\mathfrak{p}$.*

(ii) *Every finitely generated $\mathfrak{p}$-primary $R$-module $T$ is a direct sum of cyclic $R$-modules, and the number of summands of each type is an invariant of $T$.*

**Proof.**

(i) The result follows easily from the observation that if $f\colon T \to T'$ is an isomorphism, then $\operatorname{ann}(t) = \operatorname{ann}(f(t))$ for all $t \in T$.

(ii) The Primary Decomposition Theorem shows that $T$ is the direct sum of its primary components $T[\mathfrak{p}_i]$. By Proposition C-5.107, $T[\mathfrak{p}_i]$ is an $R_{\mathfrak{p}_i}$-module. But $R_{\mathfrak{p}_i}$ is a PID, and so the Basis Theorem and the Fundamental Theorem hold: each $T[\mathfrak{p}_i]$ is a direct sum of cyclic modules, and the numbers and isomorphism types of the cyclic summands are uniquely determined. •

We now know, when $R$ is Dedekind, that every finitely generated $R$-module $M$ is a direct sum of cyclic modules and ideals. What uniqueness is there in such a decomposition? Since the torsion submodule is a fully invariant direct summand, we may focus on torsion-free modules.

Recall Proposition C-5.3: two ideals $J$ and $J'$ in a domain $R$ are isomorphic if and only if there is $a \in \operatorname{Frac}(R)$ with $J' = aJ$.

**Lemma C-5.110.** *Let $M$ be a finitely generated torsion-free $R$-module, where $R$ is a Dedekind ring, so that $M \cong I_1 \oplus \cdots \oplus I_n$, where the $I_i$ are ideals. Then*

$$M \cong R^{n-1} \oplus J,$$

*where $J = I_1 \cdots I_n$.*

**Remark.** We call $R^{n-1} \oplus J$ a ***Steinitz normal form*** for $M$. We will prove, in Theorem C-5.112, that $J$ is unique up to isomorphism. ◀

**Proof.** It suffices to prove that $I \oplus J \cong R \oplus IJ$, for the result then follows easily by induction on $n \geq 2$. By Corollary C-5.100, there are nonzero $a, b \in \operatorname{Frac}(R)$ with $aI + bJ = R$. Since $aI \cong I$ and $bJ \cong J$, we may assume that $I$ and $J$ are coprime integral ideals. There is an exact sequence

$$0 \to I \cap J \xrightarrow{\delta} I \oplus J \xrightarrow{\alpha} I + J \to 0,$$

where $\delta\colon x \mapsto (x, x)$ and $\alpha\colon (u, v) \mapsto u - v$. Since $I$ and $J$ are coprime, however, we have $I \cap J = IJ$ and $I + J = R$, by Proposition C-5.93. As $R$ is projective, this sequence splits; that is, $I \oplus J \cong R \oplus IJ$. •

The following cancellation lemma, while true for Dedekind rings, can be false for some other rings. In Example C-4.89(ii), we described an example of Kaplansky showing that if $R = \mathbb{R}[x, y, z]/(x^2 + y^2 + z^2 - 1)$ is the real coordinate ring of the 2-sphere, then there is a finitely generated stably free $R$-module $M$ that is not free. Hence, there are free $R$-modules $F$ and $F'$ with $M \oplus F \cong F' \oplus F$ but $M \ncong F'$.

**Lemma C-5.111.** *Let $R$ be a Dedekind ring. If $R \oplus G \cong R \oplus H$, where $G$ and $H$ are $R$-modules, then $G \cong H$.*

**Proof.** We may assume that there is a module $E = A \oplus G = B \oplus H$, where $A \cong R \cong B$. Let $p\colon E \to B$ be the projection $p\colon (b,h) \mapsto b$, and let $p' = p|G$. Now

$$\ker p' = G \cap H \quad \text{and} \quad \operatorname{im} p' \subseteq B \cong R.$$

Thus, $\operatorname{im} p' \cong L$, where $L$ is an ideal in $R$.

If $\operatorname{im} p' = \{0\}$, then $G \subseteq \ker p = H$. Since $E = A \oplus G$, Corollary B-2.16 in Part 1 gives $H = G \oplus (H \cap A)$. On the one hand, $E/G = (A \oplus G)/G \cong A \cong R$; on the other hand, $E/G = (B \oplus H)/G \cong B \oplus (H/G) \cong R \oplus (H/G)$. Thus, $R \cong R \oplus (H/G)$. Since $R$ is a domain, this forces $H/G = \{0\}$: if $R = X \oplus Y$, then $X$ and $Y$ are ideals; if $x \in X$ and $y \in Y$ are both nonzero, then $xy \in X \cap Y = (0)$, giving zero-divisors in $R$. It follows that $H/G = \{0\}$ and $G = H$.

We may now assume that $L = \operatorname{im} p'$ is a nonzero ideal. The First Isomorphism Theorem gives $G/(G \cap H) \cong L$. Since $R$ is a Dedekind ring, $L$ is a projective module, and so

$$G = I \oplus (G \cap H),$$

where $I \cong L$. Similarly,

$$H = J \oplus (G \cap H),$$

where $J$ is isomorphic to an ideal. Therefore,

$$E = A \oplus G = A \oplus I \oplus (G \cap H),$$
$$E = B \oplus H = B \oplus J \oplus (G \cap H).$$

It follows that

$$A \oplus I \cong E/(G \cap H) \cong B \oplus J.$$

If we can prove that $I \cong J$, then

$$G = I \oplus (G \cap H) \cong J \oplus (G \cap H) = H.$$

Therefore, we have reduced the theorem to the special case where $G$ and $H$ are nonzero ideals.

We will prove that if $\alpha\colon R \oplus I \to R \oplus J$ is an isomorphism, then $I \cong J$. As in our discussion of generalized matrices on page 142, $\alpha$ determines a $2 \times 2$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

where $a_{11}\colon R \to R$, $a_{21}\colon R \to J$, $a_{12}\colon I \to R$, and $a_{22}\colon I \to J$. Indeed, as maps between ideals are just multiplications by elements of $Q = \operatorname{Frac}(R)$, we may regard $A$ as a matrix in $\mathrm{GL}(2, Q)$. Now $a_{21} \in J$ and $a_{22} I \subseteq J$, so that if $d = \det(A)$, then

$$dI = (a_{11}a_{22} - a_{12}a_{12})I \subseteq J.$$

Similarly, $\beta = \alpha^{-1}$ determines a $2 \times 2$ matrix $B = A^{-1}$. But

$$d^{-1}J = \det(B)J \subseteq I,$$

so that $J \subseteq dI$. We conclude that $J = dI$, and so $J \cong I$. •

Let us sketch another proof of Lemma C-5.111 using exterior algebra. Let $R$ be a Dedekind ring, and let $I$ and $J$ be fractional ideals; we show that $R \oplus I \cong R \oplus J$ implies $I \cong J$. The fact that $2 \times 2$ determinants are used in the original proof suggests that second exterior powers may be useful. By Theorem B-5.35 in Part 1,

$$\bigwedge^2 (R \oplus I) \cong \Big(R \otimes \bigwedge^2 I\Big) \oplus \Big(\bigwedge^1 R \otimes_R \bigwedge^1 I\Big) \oplus \Big(\bigwedge^2 R \otimes_R I\Big).$$

Now $\bigwedge^2 R = \{0\}$, by Corollary B-5.30 in Part 1, and $\bigwedge^1 R \otimes_R \bigwedge^1 I \cong R \otimes_R I \cong I$. We now show, for every maximal ideal $\mathfrak{m}$, that $\big(\bigwedge^2 I\big)_{\mathfrak{m}} = \{0\}$. By Exercise C-5.25 on page 445,

$$\Big(\bigwedge^n I\Big)_{\mathfrak{m}} \cong \bigwedge^n I_{\mathfrak{m}}.$$

But $R_{\mathfrak{m}}$ is a PID, so that $I_{\mathfrak{m}}$ is a principal ideal, and hence $\bigwedge^2 I_{\mathfrak{m}} = \{0\}$, by Corollary B-5.30 in Part 1. It now follows from Proposition C-5.38(i) that $\bigwedge^2 I = \{0\}$. Therefore, $\bigwedge^2 (R \oplus I) \cong I$. Similarly, $\bigwedge^2 (R \oplus J) \cong J$, and so $I \cong J$.

**Theorem C-5.112 (Steinitz).** *Let $R$ be a Dedekind ring, and let $M \cong I_1 \oplus \cdots \oplus I_n$ and $M' \cong I'_1 \oplus \cdots \oplus I'_\ell$ be finitely generated torsion-free $R$-modules. Then $M \cong M'$ if and only if $n = \ell$ and $I_1 \cdots I_n \cong I'_1 \cdots I'_n$.*

**Proof.** Lemma C-5.1(iii) shows that $\operatorname{rank}(I_i) = 1$ for all $i$, and Lemma C-5.1(i) shows that $\operatorname{rank}(M) = n$; similarly, $\operatorname{rank}(M') = \ell$. Since $M \cong M'$, we have $Q \otimes_R M \cong Q \otimes_R M'$, so that $\operatorname{rank}(M) = \operatorname{rank}(M')$ and $n = \ell$. By Lemma C-5.110, it suffices to prove that if $R^n \oplus I \cong R^n \oplus J$, then $I \cong J$. But this follows at once from repeated use of Lemma C-5.111. •

Recall that two $R$-modules $A$ and $B$ over a commutative ring $R$ are *stably isomorphic* if there exists an $R$-module $C$ with $A \oplus C \cong B \oplus C$.

**Corollary C-5.113.** *Let $R$ be a Dedekind ring, and let $M$ and $M'$ be finitely generated torsion-free $R$-modules. Then $M$ and $M'$ are stably isomorphic if and only if they are isomorphic.*

**Proof.** Isomorphic modules are always stably isomorphic. To prove the converse, assume that there is a finitely generated torsion-free $R$-module $X$ with

$$M \oplus X \cong M' \oplus X.$$

By Lemma C-5.110, there are ideals $I, J, L$ with $M \cong R^{n-1} \oplus I$, $M' \cong R^{n-1} \oplus J$, and $X \cong R^{m-1} \oplus L$, where $n = \operatorname{rank}(M) = \operatorname{rank}(M')$. Hence,

$$M \oplus X \cong R^{n-1} \oplus I \oplus R^{m-1} \oplus L \cong R^{n+m-1} \oplus IL.$$

Similarly,

$$M' \oplus X \cong R^{n+m-1} \oplus JL.$$

By Theorem C-5.112, $IL \cong JL$, and so there is a nonzero $a \in \operatorname{Frac}(R)$ with $aIL = JL$. Multiplying by $L^{-1}$ gives $aI = J$, and so $I \cong J$. Therefore,

$$M \cong R^{n-1} \oplus I \cong R^{n-1} \oplus J \cong M'. \quad \bullet$$

If $R$ is a ring, then $R$-**Proj** is the full subcategory of ${}_R\mathbf{Mod}$ generated by all finitely generated projective left $R$-modules. Recall that the *Grothendieck group* $K_0(R)$ has generators all $[P]$ and relations $[P] + [P'] = [P \oplus P']$, where $[P]$ is the isomorphism class of $P$. There is a relation between the class group $C(R)$ of a Dedekind ring $R$ and its Grothendieck group $K_0(R)$. If $I$ is a nonzero ideal in a Dedekind ring $R$, denote the corresponding element in $C(R)$ by $\mathrm{cls}(I)$. Corollary C-4.88 says that two $R$-modules determine the same element of $K_0(R)$ if and only if they are stably isomorphic. Thus, if $R$ is Dedekind, then Corollary C-5.113 says that two finitely generated projective $R$-modules are isomorphic if and only if they determine the same element of $K_0(R)$.

**Theorem C-5.114.** *If $R$ is a Dedekind ring, then*

$$K_0(R) \cong C(R) \oplus \mathbb{Z},$$

*where $C(R)$ is the class group of $R$.*

**Proof.** If $P$ is a nonzero finitely generated projective $R$-module, then $P$ has a Steinitz normal form: by Lemma C-5.110, $P \cong R^{n-1} \oplus J$, where $\mathrm{rank}(P) = n$ and $J$ is a nonzero finitely generated ideal. Moreover, the isomorphism class of $J$ is uniquely determined by $[P]$, by Theorem C-5.112. Therefore, the function $\varphi\colon K_0(R) \to C(R) \oplus \mathbb{Z}$, given by $\varphi\colon [P] \mapsto (\mathrm{cls}(J), n)$, is a well-defined homomorphism. Clearly, $\varphi$ is surjective, and it is injective by Corollary C-5.113. •

**Corollary C-5.115.** *If $R$ is a Dedekind ring, then $\widetilde{K}_0(R) \cong C(R)$.*

**Proof.** By definition, the reduced Grothendieck group $\widetilde{K}_0(R) = \ker \rho$, where $\rho\colon K_0(R) \to \mathbb{Z}$ is defined by $[P] \mapsto \mathrm{rank}(P)$. Here, $\ker \rho = C(R)$. •

## Exercises

**C-5.49.** Let $R$ be a Dedekind ring, and let $I \subseteq R$ be a nonzero ideal. Prove that there exists an ideal $J \subseteq R$ with $I + J = R$ and $IJ$ principal.

**Hint.** Let $I = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_n^{e_n}$, and choose $r_i \in \mathfrak{p}_i^{e_i} - \mathfrak{p}_i^{e_i+1}$. Use the Chinese Remainder Theorem to find an element $a \in R$ with $a \in \mathfrak{p}_i^{e_i}$ and $a \notin \mathfrak{p}_i^{e_i+1}$, and consider the prime factorization of $Ra$.

**C-5.50.** (i) If $R$ is a nonzero commutative ring, prove that $R^n \cong R^m$ implies $n = m$. Conclude that the rank of a free $R$-module is well-defined.

**Hint.** Corollary B-4.38 in Part 1.

(ii) If $R$ is any commutative ring, prove that $\mathbb{Z}$ is a direct summand of $K_0(R)$.

**C-5.51.** If $I$ is a fractional ideal in a Dedekind ring $R$, prove that $I \otimes_R I^{-1} \cong R$.

**Hint.** Use invertibility of $I$.

## C-5.4. Homological Dimensions

**Rings in this section need not be commutative**.

Semisimple and Dedekind rings can be characterized in terms of their projective modules: Proposition C-2.23 says that a ring $R$ is semisimple if and only if every $R$-module is projective; Theorem C-5.94(ii) says that a domain $R$ is Dedekind if and only if every ideal is projective. The notion of global dimension allows us to classify arbitrary rings in this spirit.

**Definition.** Let $R$ be a ring and let $A$ be a left $R$-module. If there is a finite projective resolution

$$0 \to P_n \to \cdots \to P_1 \to P_0 \to A \to 0,$$

then we write $\operatorname{pd}(A) \le n$.[6] If $n \ge 0$ is the smallest integer such that $\operatorname{pd}(A) \le n$, then we say that $A$ has ***projective dimension*** $n$; if there is no finite projective resolution of $A$, then $\operatorname{pd}(A) = \infty$.

**Example C-5.116.**

(i) A module $A$ is projective if and only if $\operatorname{pd}(A) = 0$. We may thus regard $\operatorname{pd}(A)$ as a measure of how far away $A$ is from being projective.

(ii) If $R$ is a Dedekind ring, we claim that $\operatorname{pd}(A) \le 1$ for every $R$-module $A$. By Theorem C-5.97, every submodule of a free $R$-module is projective. Hence, if $F$ is a free $R$-module and $\varepsilon\colon F \to A$ is a surjection, then $\ker \varepsilon$ is projective, and

$$0 \to \ker \varepsilon \to F \xrightarrow{\varepsilon} A \to 0$$

is a projective resolution of $A$. ◀

**Definition.** Let $\mathbf{P} = \to P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \xrightarrow{\varepsilon} A \to 0$ be a projective resolution of a module $A$. If $n \ge 0$, then the $n$th ***syzygy***[7] is

$$\Omega_n(A, \mathbf{P}) = \begin{cases} \ker \varepsilon & \text{if } n = 0, \\ \ker d_n & \text{if } n \ge 1. \end{cases}$$

**Proposition C-5.117.** *For every $n \ge 1$, for all left $R$-modules $A$ and $B$, and for every projective resolution $\mathbf{P}$ of $B$, there is an isomorphism*

$$\operatorname{Ext}_R^{n+1}(A, B) \cong \operatorname{Ext}_R^1(\Omega_{n-1}(A, \mathbf{P}), B).$$

**Proof.** The proof is by induction on $n \ge 1$. Exactness of the projective resolution

$$\mathbf{P} = \to P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \xrightarrow{\varepsilon} A \to 0$$

gives exactness of

$$\to P_2 \xrightarrow{d_2} P_1 \to \Omega_0(A, \mathbf{P}) \to 0,$$

which is a projective resolution $\mathbf{P}^+$ of $\Omega_0(A, \mathbf{P})$. In more detail, define

$$P_n^+ = P_{n+1} \qquad \text{and} \qquad d_n^+ = d_{n+1}.$$

[6] Sometimes it is convenient to display the ring $R$, and we will write $\operatorname{pd}_R(M)$ in this case.

[7] The term *syzygy* was introduced into mathematics by Sylvester in 1850. It is a Greek word meaning *yoke*, and it was used in astronomy to mean an alignment of three celestial figures.

Since $\operatorname{Ext}^1$ is independent of the choice of projective resolution of the first variable,

$$\operatorname{Ext}^1_R(\Omega_0(A,\mathbf{P}),B) = \frac{\ker(d_2^+)^*}{\operatorname{im}(d_1^+)^*} = \frac{\ker(d_3)^*}{\operatorname{im}(d_2)^*} = \operatorname{Ext}^2_R(A,B).$$

The inductive step is proved in the same way, noting that

$$\to P_{n+2} \xrightarrow{d_{n+2}} P_{n+1} \to \Omega_n(A,\mathbf{P}) \to 0$$

is a projective resolution of $\Omega_n(A,\mathbf{P})$. •

**Corollary C-5.118.** *For all left $R$-modules $A$ and $B$, for all $n \geq 0$, and for any projective resolutions $\mathbf{P}$ and $\mathbf{P}'$ of $A$, there is an isomorphism*

$$\operatorname{Ext}^1_R(\Omega_n(A,\mathbf{P}),B) \cong \operatorname{Ext}^1_R(\Omega_n(A,\mathbf{P}'),B).$$

**Proof.** By Proposition C-5.117, both sides are isomorphic to $\operatorname{Ext}^{n+1}_R(A,B)$. •

Two left $R$-modules $\Omega$ and $\Omega'$ are called ***projectively equivalent*** if there exist projective left $R$-modules $P$ and $P'$ with $\Omega \oplus P \cong \Omega' \oplus P'$. Exercise C-5.52 on page 495 shows that any two $n$th syzygies of a left $R$-module $A$ are projectively equivalent. We often abuse notation and speak of *the* $n$th syzygy of a module, writing $\Omega_n(A)$ instead of $\Omega_n(A,\mathbf{P})$.

Syzygies help compute projective dimension.

**Lemma C-5.119.** *The following conditions are equivalent for a left $R$-module $A$:*

(i) $\operatorname{pd}(A) \leq n$.

(ii) $\operatorname{Ext}^k_R(A,B) = \{0\}$ *for all left $R$-modules $B$ and all $k \geq n+1$.*

(iii) $\operatorname{Ext}^{n+1}_R(A,B) = \{0\}$ *for all left $R$-modules $B$.*

(iv) *For every projective resolution $\mathbf{P}$ of $A$, the $(n-1)$st syzygy $\Omega_{n-1}(A,\mathbf{P})$ is projective.*

(v) *There exists a projective resolution $\mathbf{P}$ of $A$ with $\Omega_{n-1}(A,\mathbf{P})$ projective.*

**Proof.**

(i) $\Rightarrow$ (ii). By hypothesis, there is a projective resolution $\mathbf{P}$ of $A$ with $P_k = \{0\}$ for all $k \geq n+1$. Necessarily, all the maps $d_k\colon P_k \to P_{k-1}$ are zero for $k \geq n+1$, and so

$$\operatorname{Ext}^k_R(A,B) = \frac{\ker(d_{k+1})^*}{\operatorname{im}(d_k)^*} = \{0\}.$$

(ii) $\Rightarrow$ (iii). Obvious.

(iii) $\Rightarrow$ (iv). If $\mathbf{P} = \to P_n \to P_{n-1} \to \cdots \to P_1 \to P_0 \to A \to 0$ is a projective resolution of $A$, then $\operatorname{Ext}^{n+1}_R(A,B) \cong \operatorname{Ext}^1_R(\Omega_{n-1}(A,\mathbf{P}),B)$, by Proposition C-5.117. But the last group is $\{0\}$, by hypothesis, so that $\Omega_{n-1}(A)$ is projective, by Corollary C-3.86.

(iv) $\Rightarrow$ (v). Obvious.

(v) $\Rightarrow$ (i). If

$$\to P_n \to P_{n-1} \to \cdots \to P_1 \to P_0 \to A \to 0$$

is a projective resolution of $A$, then

$$0 \to \Omega_{n-1}(A) \to P_{n-1} \to \cdots \to P_1 \to P_0 \to A \to 0$$

is an exact sequence. Since $\Omega_{n-1}(A)$ is projective, the last sequence is a projective resolution of $A$, and so $\mathrm{pd}(A) \le n$. •

**Example C-5.120.** Let $G$ be a finite cyclic group with $|G| > 1$. If $\mathbb{Z}$ is viewed as a trivial $\mathbb{Z}G$-module, then $\mathrm{pd}(\mathbb{Z}) = \infty$, because Corollary C-3.110 gives $H^n(G, \mathbb{Z}) = \mathrm{Ext}^n_{\mathbb{Z}G}(\mathbb{Z}, \mathbb{Z}) \neq \{0\}$ for all odd $n$. ◀

The following definition will soon be simplified.

**Definition.** If $R$ is a ring, then its ***left projective global dimension*** is defined by

$$\ell\mathrm{pD}(R) = \sup\{\mathrm{pd}(A) : A \in \mathrm{obj}({}_R\mathbf{Mod})\}.$$

Of course, if $R$ is commutative, we write $\mathrm{pD}(R)$ instead of $\ell\mathrm{pD}(R)$.

**Proposition C-5.121.** *For any ring $R$,*

$$\ell\mathrm{pD}(R) \le n \text{ if and only if } \mathrm{Ext}^{n+1}_R(A, B) = \{0\}$$

*for all left $R$-modules $A$ and $B$.*

**Proof.** This follows from the equivalence of (i) and (iii) in Lemma C-5.119. •

A ring $R$ is semisimple if and only if $\ell\mathrm{pD}(R) = 0$. Thus, global dimension is a measure of how far a ring is from being semisimple.

A similar discussion can be given using injective resolutions.

**Definition.** Let $R$ be a ring and let $B$ be a left $R$-module. If there is an injective resolution

$$0 \to B \to E^0 \to E^1 \to \cdots \to E^n \to 0,$$

then we write $\mathrm{id}(B) \le n$. If $n \ge 0$ is the smallest integer such that $\mathrm{id}(B) \le n$, then we say that $B$ has ***injective dimension*** $n$; if there is no finite injective resolution of $B$, then $\mathrm{id}(B) = \infty$.

**Example C-5.122.**

(i) A left $R$-module $B$ is injective if and only if $\mathrm{id}(B) = 0$. We may thus regard $\mathrm{id}(B)$ as a measure of how far away $B$ is from being injective.

(ii) The injective and projective dimensions of a left $R$-module $A$ can be distinct. For example, the abelian group $A = \mathbb{Z}$ has $\mathrm{pd}(A) = 0$ and $\mathrm{id}(A) = 1$.

(iii) If $R$ is a Dedekind ring, then Theorem C-5.97 says that every quotient module of an injective $R$-module is injective. Hence, if $\eta\colon B \to E$ is an imbedding of an $R$-module $B$ into an injective $R$-module $E$, then $\mathrm{coker}\,\eta$ is injective, and

$$0 \to B \xrightarrow{\eta} E \to \mathrm{coker}\,\eta \to 0$$

is an injective resolution of $B$. It follows that $\mathrm{id}(B) \le 1$. ◀

**Definition.** Let $\mathbf{E} = 0 \to B \xrightarrow{\eta} E^0 \xrightarrow{d^0} E^1 \xrightarrow{d^1} E^2 \to \cdots$ be an injective resolution of a left $R$-module $B$. If $n \geq 0$, then the $n$th ***cosyzygy*** is

$$\mho^n(B, \mathbf{E}) = \begin{cases} \operatorname{coker} \eta & \text{if } n = 0, \\ \operatorname{coker} d^{n-1} & \text{if } n \geq 1. \end{cases}$$

**Proposition C-5.123.** *For every $n \geq 1$, for all left $R$-modules $A$ and $B$, and for every injective resolution $\mathbf{E}$ of $A$, there is an isomorphism*

$$\operatorname{Ext}_R^{n+1}(A, B) \cong \operatorname{Ext}_R^1(A, \mho^{n-1}(B, \mathbf{E})).$$

**Proof.** Dual to the proof of Proposition C-5.117. •

**Corollary C-5.124.** *For all left $R$-modules $A$ and $B$, for all $n \geq 0$, and for any injective resolutions $\mathbf{E}$ and $\mathbf{E}'$ of $B$, there is an isomorphism*

$$\operatorname{Ext}_R^1(A, \mho^n(B, \mathbf{E})) \cong \operatorname{Ext}_R^1(A, \mho^n(B, \mathbf{E}')).$$

**Proof.** Dual to the proof of Corollary C-5.118. •

Two left $R$-modules $\mho$ and $\mho'$ are called ***injectively equivalent*** if there exist injective left $R$-modules $E$ and $E'$ with $\mho \oplus E \cong \mho' \oplus E'$. Exercise C-5.53 on page 495 shows that any two $n$th cosyzygies of a left $R$-module $B$ are injectively equivalent. We often abuse notation and speak of *the* $n$th cosyzygy of a module, writing $\mho^n(B)$ instead of $\mho^n(B, \mathbf{E})$.

Cosyzygies help compute injective dimension.

**Lemma C-5.125.** *The following conditions are equivalent for a left $R$-module $B$:*

(i) $\operatorname{id}(B) \leq n$.

(ii) $\operatorname{Ext}_R^k(A, B) = \{0\}$ *for all left $R$-modules $A$ and all $k \geq n+1$.*

(iii) $\operatorname{Ext}_R^{n+1}(A, B) = \{0\}$ *for all left $R$-modules $A$.*

(iv) *For every injective resolution $\mathbf{E}$ of $B$, the $(n-1)$st cosyzygy $\mho^{n-1}(B, \mathbf{E})$ is injective.*

(v) *There exists an injective resolution $\mathbf{E}$ of $B$ with $\mho^{n-1}(B, \mathbf{E})$ injective.*

**Proof.** Dual to that of Lemma C-5.119, using Exercise C-3.48 on page 308 •

**Definition.** If $R$ is a ring, then its ***left injective global dimension*** is defined by

$$\ell\operatorname{iD}(R) = \sup\{\operatorname{id}(B) : B \in \operatorname{obj}({}_R\mathbf{Mod})\}.$$

**Proposition C-5.126.** *For any ring $R$,*

$$\ell\operatorname{iD}(R) \leq n \text{ if and only if } \operatorname{Ext}_R^{n+1}(A, B) = \{0\}$$

*for all left $R$-modules $A$ and $B$.*

**Proof.** This follows from the equivalence of (i) and (iii) in Lemma C-5.125. •

**Theorem C-5.127.** *For every ring $R$,*

$$\ell\operatorname{pD}(R) = \ell\operatorname{iD}(R).$$

**Proof.** This follows at once from Propositions C-5.121 and C-5.126, for each number is equal to the smallest $n$ for which $\operatorname{Ext}_R^{n+1}(A,B) = \{0\}$ for all left $R$-modules $A$ and $B$. •

**Definition.** The ***left global dimension*** of a ring $R$ is the common value of the left projective global dimension and the left injective global dimension:

$$\ell\mathrm{D}(R) = \ell\mathrm{pD}(R) = \ell\mathrm{iD}(R).$$

If $R$ is commutative, then we denote its global dimension by $\mathrm{D}(R)$.

There is also a ***right global dimension*** $r\mathrm{D}(R) = \ell\mathrm{D}(R^{\mathrm{op}})$ of a ring $R$; thus, we consider projective and injective dimensions of right $R$-modules. If $R$ is commutative, then $\ell\mathrm{D}(R) = r\mathrm{D}(R)$ and we write $\mathrm{D}(R)$. Since left semisimple rings are also right semisimple, by Corollary C-2.36, we have $\ell\mathrm{D}(R) = 0$ if and only if $r\mathrm{D}(R) = 0$. On the other hand, these two dimensions can differ. Jategaonkar [**115**] proved that if $1 \le m \le n \le \infty$, then there exists a ring $R$ with $\ell\mathrm{D}(R) = m$ and $r\mathrm{D}(R) = n$.

**Theorem C-5.128.** *A domain $R$ is a Dedekind ring if and only if $D(R) \le 1$.*

**Proof.** Let $R$ be a Dedekind ring. If $A$ is an $R$-module, there is a free $R$-module $F$ and an exact sequence $0 \to K \to F \to A \to 0$. But $K$ is projective, by Theorem C-5.97 (which says that a domain $R$ is Dedekind if and only if every submodule of a projective $R$-module is projective). Thus, $\mathrm{pd}(A) \le 1$ and $D(R) \le 1$.

Conversely, assume that $D(R) \le 1$. If $S$ is a submodule of a projective $R$-module $F$, then there is an exact sequence $0 \to S \to F \to F/S \to 0$. But $\mathrm{pd}(F/S) \le 1$, so that $S$ is projective, by the equivalence of parts (i) and (iv) of Lemma C-5.119. Theorem C-5.97 shows that $R$ is Dedekind. •

The next theorem says that $\ell\mathrm{D}(R)$ can be computed from $\mathrm{pd}(M)$ for finitely generated $R$-modules $M$; in fact, $\ell\mathrm{D}(R)$ can even be computed from $\mathrm{pd}(M)$ for $M$ cyclic.

**Lemma C-5.129.** *A left $R$-module $B$ is injective if and only if $\operatorname{Ext}_R^1(R/I, B) = \{0\}$ for every left ideal $I$.*

**Proof.** If $B$ is injective, then $\operatorname{Ext}_R^1(A,B)$ vanishes for every right $R$-module $A$. Conversely, suppose that $\operatorname{Ext}_R^1(R/I,B) = \{0\}$ for every left ideal $I$. Applying $\operatorname{Hom}_R(\square, B)$ to the exact sequence $0 \to I \to R \to R/I \to 0$ gives exactness of

$$\operatorname{Hom}_R(R,B) \to \operatorname{Hom}_R(I,B) \to \operatorname{Ext}_R^1(R/I,B) = 0.$$

That is, every $R$-map $f\colon I \to B$ can be extended to an $R$-map $R \to B$ (Proposition B-4.51 in Part 1). But this is precisely the Baer criterion, Theorem B-4.57 in Part 1, and so $B$ is injective. •

**Theorem C-5.130 (Auslander).** *For any ring $R$,*

$$\ell\mathrm{D}(R) = \sup\{\mathrm{pd}(R/I) : I \text{ is a left ideal}\}.$$

**Proof (Matlis).** If $\sup\{\mathrm{pd}(R/I)\} = \infty$, we are done. Therefore, we may assume there is an integer $n \geq 0$ with $\mathrm{pd}(R/I) \leq n$ for every left ideal $I$. By Lemma C-5.125, $\mathrm{Ext}_R^{n+1}(R/I, B) = \{0\}$ for every left $R$-module $B$. But $\ell\mathrm{pD}(R) = \ell\mathrm{iD}(R)$, by Theorem C-5.127, so that it suffices to prove that $\mathrm{id}(B) \leq n$ for every $B$. Let $\mathbf{E}$ be an injective resolution of $B$, with $(n-1)$st cosyzygy $\mho^{n-1}(B)$. By Proposition C-5.123, $\{0\} = \mathrm{Ext}_R^{n+1}(R/I, B) \cong \mathrm{Ext}_R^1(R/I, \mho^{n-1}(B))$. Now Lemma C-5.129 gives $\mho^{n-1}(B)$ injective, and so Lemma C-5.125 gives $\mathrm{id}(B) \leq n$, as desired. •

This theorem explains why every ideal in a Dedekind ring being projective is such a strong condition.

Just as Ext defines the global dimension of a ring $R$, we can use Tor to define the *weak dimension* (or Tor-*dimension*) of a ring $R$.

**Definition.** Let $R$ be a ring and let $A$ be a right $R$-module. A ***flat resolution*** of $A$ is an exact sequence

$$\to F_n \to \cdots \to F_1 \to F_0 \to A \to 0$$

in which each $F_n$ is a flat right $R$-module.

If there is a finite flat resolution

$$0 \to F_n \to \cdots \to F_1 \to F_0 \to A \to 0,$$

then we write $\mathrm{fd}(A) \leq n$. If $n \geq 0$ is the smallest integer such that $\mathrm{fd}(A) \leq n$, then we say that $A$ has ***flat dimension*** $n$; if there is no finite flat resolution of $A$, then $\mathrm{fd}(A) = \infty$.

**Example C-5.131.**

(i) A right $R$-module $A$ is flat if and only if $\mathrm{fd}(A) = 0$. We may thus regard $\mathrm{fd}(A)$ as a measure of how far away $A$ is from being flat.

(ii) Since projective right $R$-modules are flat, every projective resolution of $A$ is a flat resolution. It follows that if $R$ is any ring, then $\mathrm{fd}(A) \leq \mathrm{pd}(A)$ for every right $R$-module $A$. The same argument applies to left $R$-modules.

(iii) If $R$ is a Dedekind ring and $A$ is an $R$-module, then $\mathrm{fd}(A) \leq \mathrm{pd}(A) \leq 1$, by (ii). Corollary C-5.104 says that $A$ is flat if and only if $A$ is torsion-free. Hence, $\mathrm{fd}(A) = 1$ if and only if $A$ is not torsion-free.

(iv) Schanuel's Lemma (Proposition B-4.48 in Part 1) does not hold for flat resolutions. Recall that a $\mathbb{Z}$-module is flat if and only if it is torsion-free. Consider the flat resolutions

$$0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0,$$
$$0 \to K \to F \to \mathbb{Q}/\mathbb{Z} \to 0,$$

where $F$ (and its subgroup $K$) is a free abelian group. There is no isomorphism $K \oplus \mathbb{Q} \cong \mathbb{Z} \oplus F$, for $\mathbb{Q}$ is not projective. ◄

**Definition.** Let $\mathbf{F} = \to F_2 \xrightarrow{d_2} F_1 \xrightarrow{d_1} F_0 \xrightarrow{\varepsilon} B \to 0$ be a flat resolution of a left $R$-module $B$. If $n \geq 0$, then the $n$th ***yoke*** is

$$Y_n(B, \mathbf{F}) = \begin{cases} \ker \varepsilon & \text{if } n = 0, \\ \ker d_n & \text{if } n \geq 1. \end{cases}$$

The term *yoke* is not standard; it is a translation of the Greek word *syzygy*.

**Proposition C-5.132.** *For every $n \geq 1$, for all right $R$-modules $A$ and left $R$-modules $B$, and for every flat resolution $\mathbf{F}$ of $B$, there is an isomorphism*

$$\operatorname{Tor}_{n+1}^R(A, B) \cong \operatorname{Tor}_1^R(A, Y_{n-1}(B, \mathbf{F})).$$

**Proof.** The same as the proof of Proposition C-5.117. •

**Corollary C-5.133.** *For every right $R$-module $A$ and left $R$-module $B$, for all $n \geq 0$, and for any flat resolutions $\mathbf{F}$ and $\mathbf{F}'$ of $B$, there is an isomorphism*

$$\operatorname{Tor}_1^R(A, Y_n(B, \mathbf{F})) \cong \operatorname{Tor}_1^R(A, Y_n(B, \mathbf{F})).$$

**Proof.** The same as the proof of Corollary C-5.118. •

**Lemma C-5.134.** *The following conditions are equivalent for a left $R$-module $B$:*

(i) $\operatorname{fd}(B) \leq n$.

(ii) $\operatorname{Tor}_k^R(A, B) = \{0\}$ *for all* $k \geq n+1$ *and all right $R$-modules $A$.*

(iii) $\operatorname{Tor}_{n+1}^R(A, B) = \{0\}$ *for all right $R$-modules $A$.*

(iv) *For every flat resolution $\mathbf{F}$ of $B$, the $(n-1)$st yoke $Y_{n-1}(B, \mathbf{F})$ is flat.*

(v) *There exists a flat resolution $\mathbf{F}$ of $B$ with flat $(n-1)$st yoke $Y_{n-1}(B, \mathbf{F})$.*

**Proof.** The same as the proof of Lemma C-5.119. •

**Definition.** The ***right weak dimension*** of a ring $R$ is defined by

$$r\operatorname{wD}(R) = \sup\{\operatorname{fd}(A) : A \in \operatorname{obj}(\mathbf{Mod}_R)\}.$$

**Proposition C-5.135.** *For any ring $R$, $r\operatorname{wD}(R) \leq n$ if and only if $\operatorname{Tor}_{n+1}^R(A, B) = \{0\}$ for every left $R$-module $B$.*

**Proof.** This follows at once from Lemma C-5.134. •

We define the flat dimension of left $R$-modules in the obvious way, in terms of its flat resolutions, and taking the supremum over all left $R$-modules gives the left weak dimension.

**Definition.** The ***left weak dimension*** of a ring $R$ is

$$\ell\operatorname{wD}(R) = \sup\{\operatorname{fd}(B) : B \in \operatorname{obj}({}_R\mathbf{Mod})\}.$$

**Theorem C-5.136.** *For any ring $R$,*

$$r\operatorname{wD}(R) = \ell\operatorname{wD}(R).$$

**Proof.** If either dimension is finite, then the left or right weak dimension is the smallest $n \geq 0$ with $\operatorname{Tor}_{n+1}^R(A, B) = \{0\}$ for all right $R$-modules $A$ and all left $R$-modules $B$. •

**Definition.** The ***weak dimension*** of a ring $R$, denoted by $\operatorname{wD}(R)$, is the common value of $r\operatorname{wD}(R)$ and $\ell\operatorname{wD}(R)$.

As we have remarked earlier, there are (noncommutative) rings whose left global dimension and right global dimension can be distinct. In contrast, weak dimension has no left/right distinction, because tensor and Tor involve both left and right modules simultaneously.

**Example C-5.137.** A ring $R$ has $\operatorname{wD}(R) = 0$ if and only if every left or right module is flat. These rings turn out to be ***von Neumann regular***: for each $a \in R$, there exists $a' \in R$ with $aa'a = a$. Examples of such rings are Boolean rings (rings $R$ in which $r^2 = r$ for all $r \in R$) and $\operatorname{End}_k(V)$, where $V$ is a (possibly infinite-dimensional) vector space over a field $k$ (Rotman [**187**], pp. 159–160). For these rings $R$, we have $\operatorname{wD}(R) = 0 < \ell\operatorname{D}(R)$. ◄

The next proposition explains why weak dimension is so called.

**Proposition C-5.138.** *For any ring $R$,*

$$\operatorname{wD}(R) \leq \min\{\ell\operatorname{D}(R), r\operatorname{D}(R)\}.$$

**Proof.** It suffices to prove that $\operatorname{fd}(A) \leq \operatorname{pd}(A)$ for any right or left $R$-module $A$; but we saw this in Example C-5.131(ii). •

**Corollary C-5.139.** *Suppose that $\operatorname{Ext}_R^n(A, B) = \{0\}$ for all left $R$-modules $A$ and $B$. Then $\operatorname{Tor}_n^R(C, D) = \{0\}$ for all right $R$-modules $C$ and all left $R$-modules $D$.*

**Proof.** If $\operatorname{Ext}_R^n(A, B) = \{0\}$ for all $A$, $B$, then $\ell\operatorname{D}(R) \leq n-1$; if $\operatorname{Tor}_n^R(C, D) \neq \{0\}$ for some $C$, $D$, then $n \leq \operatorname{wD}(DR)$. Hence,

$$\ell\operatorname{D}(R) \leq n - 1 < n \leq \operatorname{wD}(R),$$

contradicting Proposition C-5.138. •

**Lemma C-5.140.** *A left $R$-module $B$ is flat if and only if $\operatorname{Tor}_1^R(R/I, B) = \{0\}$ for every right ideal $I$.*

**Proof.** Exactness of $0 \to I \xrightarrow{i} R \to R/I \to 0$ gives exactness of

$$0 = \operatorname{Tor}_1^R(R, B) \to \operatorname{Tor}_1^R(R/I, B) \to I \otimes_R B \xrightarrow{i\otimes 1} R \otimes_R B.$$

Therefore, $i \otimes 1$ is an injection if and only if $\operatorname{Tor}_1^R(R/I, B) = \{0\}$.

On the other hand, $B$ is flat if and only if $i \otimes 1$ is an injection for every right ideal $I$, by Corollary B-4.109 in Part 1. •

As global dimension, weak dimension can be computed from cyclic modules.

**Corollary C-5.141.** *For any ring $R$,*

$$\begin{aligned}\mathrm{wD}(R) &= \sup\{\mathrm{fd}(R/I) : I \text{ is a right ideal of } R\}\\ &= \sup\{\mathrm{fd}(R/J) : J \text{ is a left ideal of } R\}.\end{aligned}$$

**Proof.** This proof is similar to the proof of Theorem C-5.130, using Lemma C-5.140 instead of Lemma C-5.129. •

**Proposition C-5.142.** *If $R$ is a commutative ring and $S \subseteq R$ is multiplicative, then*

$$\mathrm{wD}(S^{-1}R) \leq \mathrm{wD}(R).$$

**Proof.** We may assume that $\mathrm{wD}(R) = n < \infty$. If $A$ is an $S^{-1}R$-module, we must prove that $\mathrm{fd}(A) \leq n$. By Proposition C-5.27(i), $A \cong S^{-1}M$ for some $R$-module $M$ (indeed, $M = {}_hA$, the $S^{-1}R$-module $A$ viewed as an $R$-module). By hypothesis, there is an $R$-flat resolution

$$0 \to F_n \to \cdots \to F_0 \to M \to 0.$$

But localization is an exact functor, so that

$$0 \to S^{-1}F_n \to \cdots \to S^{-1}F_0 \to S^{-1}M \to 0$$

is an exact sequence of $S^{-1}R$-modules. Finally, each $S^{-1}F_i$ is flat, by Exercise B-4.93 on page 542 in Part 1, so that $\mathrm{fd}(S^{-1}M) = \mathrm{fd}(A) \leq n$. •

**Theorem C-5.143.** *Let $R$ be a left noetherian ring.*

(i) *If $A$ is a finitely generated left $R$-module, then*

$$\mathrm{pd}(A) = \mathrm{fd}(A).$$

(ii) $\mathrm{wD}(R) = \ell\mathrm{D}(R)$.

(iii) *If $R$ is a commutative noetherian ring, then*

$$\mathrm{wD}(R) = \mathrm{D}(R).$$

**Proof.**

(i) It is always true that $\mathrm{fd}(A) \leq \mathrm{pd}(A)$, since every projective resolution is a flat resolution (Example C-5.131(ii)). For the reverse inequality, it is enough to prove that if $\mathrm{fd}(A) \leq n$, then $\mathrm{pd}(A) \leq n$. By Lemma C-5.41, there is a projective resolution of $A$,

$$\to P_n \to \cdots \to P_0 \to A \to 0,$$

in which every $P_i$ is finitely generated. Now this is also a flat resolution, so that, by Lemma C-5.134, $\mathrm{fd}(A) \leq n$ implies that $Y_n = \ker(P_{n-1} \to P_{n-2})$ is flat. But every finitely generated flat left $R$-module is projective, by Corollary B-4.112 in Part 1 (because $R$ is left noetherian), and so

$$0 \to Y_{n-1} \to P_{n-1} \to \cdots \to P_0 \to A \to 0$$

is a projective resolution. Therefore, $\mathrm{pd}(A) \leq n$.

(ii) By Theorem C-5.130, we have that $\ell\mathrm{D}(R)$ is the supremum of projective dimensions of cyclic left $R$-modules, and by Corollary C-5.141, we have that $\mathrm{wd}(R)$ is the supremum of flat dimensions of cyclic left $R$-modules. But part (i) gives $\mathrm{fd}(A) = \mathrm{pd}(A)$ for every finitely generated right $R$-module $A$, and this suffices to prove the result.

(iii) This is a special case of (ii). •

## Exercises

∗ **C-5.52.** (i) If $A \to B \xrightarrow{f} C \to D$ is an exact sequence and $X$ is any module, prove that there is an exact sequence

$$A \to B \oplus X \xrightarrow{f \oplus 1_X} C \oplus X \to D.$$

(ii) Let

$$\mathbf{P} = \to P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \xrightarrow{\varepsilon} A \to 0$$

and

$$\mathbf{P'} = \to P_2' \xrightarrow{d_2'} P_1' \xrightarrow{d_1'} P_0' \xrightarrow{\varepsilon'} A \to 0$$

be projective resolutions of a left $R$-module $A$. For all $n \geq 0$, prove that there are projective modules $Q_n$ and $Q_n'$ with

$$\Omega_n(A, \mathbf{P}) \oplus Q_n \cong \Omega_n(A, \mathbf{P'}) \oplus Q_n'.$$

**Hint.** Proceed by induction on $n \geq 0$, using Schanuel's Lemma, Proposition B-4.48 in Part 1.

∗ **C-5.53.** Let

$$0 \to B \to E^0 \to E^1 \to E^2 \to$$

and

$$0 \to B \to E'^0 \to E'^1 \to E'^2 \to$$

be injective resolutions of a left $R$-module $B$. For all $n \geq 0$, prove that there are injective modules $I_n$ and $I_n'$ with

$$\mathcal{O}^n(B, \mathbf{E}) \oplus I_n \cong \mathcal{O}^n(B, \mathbf{E'}) \oplus I_n'.$$

**Hint.** The proof is dual to that of Exercise C-5.52.

∗ **C-5.54.** Let $0 \to M' \to M \to M'' \to 0$ be an exact sequence of left $R$-modules (over some ring $R$). Use the long exact Ext sequence to prove the following statements:

(i) If $\mathrm{pd}(M') < \mathrm{pd}(M)$, prove that $\mathrm{pd}(M'') = \mathrm{pd}(M)$.

(ii) If $\mathrm{pd}(M') > \mathrm{pd}(M)$, prove that $\mathrm{pd}(M'') = \mathrm{pd}(M') + 1$.

(iii) If $\mathrm{pd}(M') = \mathrm{pd}(M)$, prove that $\mathrm{pd}(M'') \leq \mathrm{pd}(M') + 1$.

**C-5.55.** (i) If $0 \to A' \to A \to A'' \to 0$ is an exact sequence, prove that

$$\mathrm{pd}(A) \leq \max\{\mathrm{pd}(A'), \mathrm{pd}(A'')\}.$$

(ii) If the sequence in part (i) is not split and $\mathrm{pd}(A') = \mathrm{pd}(A'') + 1$, prove that

$$\mathrm{pd}(A) = \max\{\mathrm{pd}(A'), \mathrm{pd}(A'')\}.$$

∗ **C-5.56.** If $A$ is an $R$-module with $\mathrm{pd}(A) = n$, prove that there exists a free $R$-module $F$ with $\mathrm{Ext}^n_R(A, F) \neq \{0\}$.

**Hint.** Every module is a quotient of a free module.

**C-5.57.** If $G$ is a finite cyclic group of order not 1, prove that

$$\ell\mathrm{D}(\mathbb{Z}G) = \infty = r\mathrm{D}(\mathbb{Z}G).$$

**Hint.** Use Theorem C-3.109.

**C-5.58. (Auslander)** If $R$ is both left noetherian and right noetherian, prove that

$$\ell\mathrm{D}(R) = r\mathrm{D}(R).$$

**Hint.** Use weak dimension.

**C-5.59.** Prove that a noetherian von Neumann regular ring is semisimple.

**Hint.** See Example C-5.137.

∗ **C-5.60.** If $(M_\alpha)_{\alpha\in A}$ is a family of left $R$-modules, prove that

$$\mathrm{pd}\Big(\bigoplus_{\alpha\in A} M_\alpha\Big) = \sup_{\alpha\in A}\big\{\mathrm{pd}(M_\alpha)\big\}.$$

∗ **C-5.61.** Let $R$ be a commutative ring, let $c \in R$, let $R^* = R/cR$, and let $\nu\colon R \to R^*$ be the natural map.

(i) If $M$ is an $R$-module, define $M^* = M/cM$. Prove that $M^* \cong R^* \otimes_R M$.

(ii) If $A^*$ is an $R^*$-module, define ${}_\nu A^*$ to be $A^*$ viewed as an $R$-module, as on page 434. Prove that ${}_\nu A^* \cong \mathrm{Hom}_R(R^*, A^*)$. Conclude that $M \mapsto M^*$ and $A^* \mapsto {}_\nu A^*$ form an adjoint pair of functors.

∗ **C-5.62.** If $\nu\colon R \to R^*$ is a ring homomorphism and $A^*$ and $B^*$ are $R^*$-modules, prove that there is a natural isomorphism

$$\mathrm{Hom}_{R^*}(A^*, B^*) \to \mathrm{Hom}_R({}_\nu A^*, {}_\nu B^*),$$

where ${}_\nu A^*$ is $A^*$ viewed as an $R$-module via change of rings.

∗ **C-5.63.** (i) If $I$ is a left ideal in a ring $R$, prove that either $R/I$ is projective or $\mathrm{pd}(R/I) = 1 + \mathrm{pd}(I)$ (we agree that $1 + \infty = \infty$).

(ii) If $0 \to M' \to M \to M'' \to 0$ is exact and two of the modules have finite projective (or injective) dimension, prove that the third module has finite projective (or injective) dimension as well.

## C-5.5. Hilbert's Theorem on Syzygies

We are going to compute the global dimension $\mathrm{D}(k[x_1, \ldots, x_n])$ of polynomial rings. The key result is that $\mathrm{D}(R[x]) = \mathrm{D}(R)+1$ for any, possibly noncommutative, ring $R$ (where $R[x]$ is the polynomial ring in which the indeterminate $x$ commutes with all the constants in $R$). However, we will assume that $R$ is commutative for clarity of exposition.

**Henceforth, all rings are assumed commutative unless we say otherwise.**

**Lemma C-5.144.** *If $0 \to A' \to A \to A'' \to 0$ is a short exact sequence, then*

$$\operatorname{pd}(A'') \le 1 + \max\{\operatorname{pd}(A), \operatorname{pd}(A')\}.$$

**Proof.** We may assume that the right side is finite, or there is nothing to prove; let $\operatorname{pd}(A) \le n$ and $\operatorname{pd}(A') \le n$. Applying $\operatorname{Hom}(\square, B)$, where $B$ is any module, to the short exact sequence gives the long exact sequence

$$\to \operatorname{Ext}^{n+1}(A', B) \to \operatorname{Ext}^{n+2}(A'', B) \to \operatorname{Ext}^{n+2}(A, B) \to .$$

The two outside terms are $\{0\}$, by Lemma C-5.119(ii), so that exactness forces $\operatorname{Ext}^{n+2}(A'', B) = \{0\}$ for all $B$. The same lemma gives $\operatorname{pd}(A'') \le n + 1$. •

We wish to compare global dimension of $R$ and $R[x]$, and so we consider the $R[x]$-module $R[x] \otimes_R M$ arising from an $R$-module $M$.

**Definition.** If $M$ is an $R$-module over a commutative ring $R$, define

$$M[x] = \bigoplus_{i \ge 0} M_i,$$

where $M_i \cong M$ for all $i$. The $R$-module $M[x]$ is an $R[x]$-module if we define

$$x\Big(\sum_i x^i m_i\Big) = \sum_i x^{i+1} m_i.$$

Note that $M[x] \cong R[x] \otimes_R M$ via $(x^n m_n) \mapsto 1 \otimes (\sum m_n)$. Thus, if $M$ is a free $R$-module over a commutative ring $R$, then $M[x]$ is a free $R[x]$-module, for tensor product commutes with direct sum. The next result generalizes this from $\operatorname{pd}_R(M) = 0$ to higher dimensions.

**Lemma C-5.145.** *For every $R$-module $M$, where $R$ is a commutative ring,*

$$\operatorname{pd}_R(M) = \operatorname{pd}_{R[x]}(M[x]).$$

**Proof.** It suffices to prove that if one of the dimensions is finite and at most $n$, then so is the other.

If $\operatorname{pd}_R(M) \le n$, then there is an $R$-projective resolution

$$0 \to P_n \to \cdots \to P_0 \to M \to 0.$$

Since $R[x]$ is a free $R$-module, it is flat, and there is an exact sequence of $R[x]$-modules

$$0 \to R[x] \otimes_R P_n \to \cdots \to R[x] \otimes_R P_0 \to R[x] \otimes_R M \to 0.$$

But $R[x] \otimes_R M \cong M[x]$ and $R[x] \otimes_R P_n$ is a projective $R[x]$-module (for a projective is a direct summand of a free module). Therefore, $\operatorname{pd}_{R[x]}(M[x]) \le n$.

If $\operatorname{pd}_{R[x]}(M[x]) \le n$, then there is an $R[x]$-projective resolution

$$0 \to Q_n \to \cdots \to Q_0 \to M[x] \to 0.$$

As an $R$-module, $M[x] \cong \bigoplus_{n \ge 1} M_n$, where $M_n \cong M$ for all $n$. By Exercise C-5.60 on page 496, $\operatorname{pd}_R(M[x]) = \operatorname{pd}_R(M)$. Each projective $R[x]$-module $Q_i$ is an $R[x]$-summand of a free $R[x]$-module $F_i$; a fortiori, $Q_i$ is an $R$-direct summand of $F_i$.

But $R[x]$ is a free $R$-module, so that $F_i$ is also a free $R$-module. Therefore, $Q_i$ is projective as an $R$-module, and so $\mathrm{pd}_R(M) \le n = \mathrm{pd}_{R[x]}(M[x])$. •

**Corollary C-5.146.** *If $R$ is a commutative ring and $\mathrm{D}(R) = \infty$, then $\mathrm{D}(R[x]) = \infty$.*

**Proof.** If $\mathrm{D}(R) = \infty$, then for every integer $n$, there exists an $R$-module $M_n$ with $n < \mathrm{pd}(M_n)$. By Lemma C-5.145, $n < \mathrm{pd}_{R[x]}(M_n[x])$ for all $n$. Therefore, $\mathrm{D}(R[x]) = \infty$. •

**Proposition C-5.147.** *For every commutative ring $R$,*

$$D(R[x]) \le \mathrm{D}(R) + 1.$$

**Proof.** Recall the characteristic sequence, Theorem B-3.77 in Part 1: if $M$ is an $R$-module and $S\colon M \to M$ is an $R$-map, then there is an exact sequence of $R[x]$-modules

$$0 \to M[x] \to M[x] \to M^S \to 0,$$

where $M^S$ is the $R[x]$-module $M$ with scalar multiplication given by $ax^i m = aS^i(m)$. If $M$ is already an $R[x]$-module and $S\colon M \to M$ is the $R$-map $m \mapsto xm$, then $M^S = M$. By Lemma C-5.144,

$$\mathrm{pd}_{R[x]}(M) \le 1 + \mathrm{pd}_{R[x]}(M[x]) = 1 + \mathrm{pd}_R(M) \le 1 + D(R). \quad \bullet$$

We now work toward the reverse inequality.

**Definition.** If $M$ is an $R$-module, where $R$ is a commutative ring, then a nonzero element $c \in R$ is ***regular*** on $M$ (or is $M$-***regular***) if the $R$-map $M \to M$, given by $m \mapsto cm$, is injective. Otherwise, $c$ is a ***zero-divisor*** on $M$; that is, there is some nonzero $m \in M$ with $cm = 0$.

Before stating the next theorem, let us explain the notation. Suppose that $R$ is a commutative ring, $c \in R$, and $R^* = R/Rc$. If $M$ is an $R$-module, then $M/cM$ is an $(R/Rc)$-module; that is, $M/cM$ is an $R^*$-module. On the other hand, there is the notion of *change of rings* in Exercise B-4.25 on page 475 in Part 1. If $\nu\colon R \to R/Rc$ is the natural map, $r \in R$, and $a^* \in A^*$, define

$$ra^* = \nu(r)a^*.$$

This makes $A^*$ into an $R$-module, denoted by ${}_\nu A^*$. Exercise C-5.61 on page 496 asks you to prove that ${}_\nu A^* \cong \mathrm{Hom}_R(R/cR, A^*)$, so that these constructions involve an adjoint pair of functors, namely, $(\Box^*, {}_\nu\Box)$.

**Proposition C-5.148 (Rees Lemma).** *Let $R$ be a commutative ring, let $c \in R$ be neither a unit nor a zero-divisor, and let $R^* = R/Rc$. If $c$ is regular on an $R$-module $M$, then there are natural isomorphisms, for every $R^*$-module $A^*$ and all $n \ge 0$,*

$$\mathrm{Ext}^n_{R^*}(A^*, M/cM) \cong \mathrm{Ext}^{n+1}_R({}_\nu A^*, M),$$

*where ${}_\nu A^*$ is the $R^*$-module $A^*$ viewed as an $R$-module.*

**Proof.** Recall Theorem C-3.45, the axioms characterizing Ext functors. Given a sequence of contravariant functors $G^n\colon {}_{R^*}\mathbf{Mod} \to \mathbf{Ab}$, for $n \geq 0$, such that

(i) for every short exact sequence $0 \to A^* \to B^* \to C^* \to 0$ of $R^*$-modules, there is a long exact sequence with natural connecting homomorphisms

$$\to G^n(C^*) \to G^n(B^*) \to G^n(A^*) \to G^{n+1}(C^*) \to,$$

(ii) $G^0$ and $\mathrm{Hom}_{R^*}(\square, L^*)$ are naturally equivalent, for some $R^*$-module $L^*$,

(iii) $G^n(P^*) = 0$ for all projective $R^*$-modules $P^*$ and all $n \geq 1$,

then $G^n$ is naturally equivalent to $\mathrm{Ext}^n_{R^*}(\square, L^*)$ for all $n \geq 0$.

Define contravariant functors $G^n\colon {}_{R^*}\mathbf{Mod} \to \mathbf{Ab}$ by $G^n = \mathrm{Ext}^{n+1}_R({}_\nu\square, M)$; that is, for all $R^*$-modules $A^*$,

$$G^n(A^*) = \mathrm{Ext}^{n+1}_R({}_\nu A^*, M).$$

Since axiom (i) holds for the functors $\mathrm{Ext}^n$, it also holds for the functors $G^n$. Let us prove axiom (ii). The map $\mu_c\colon M \to M$, defined by $m \mapsto cm$, is an injection, because $c$ is $M$-regular, and so the sequence $0 \to M \xrightarrow{\mu_c} M \to M/cM \to 0$ is exact. Consider the portion of the long exact sequence, where $A^*$ is an $R^*$-module:

$$\mathrm{Hom}_R({}_\nu A^*, M) \to \mathrm{Hom}_R({}_\nu A^*, M/cM) \xrightarrow{\partial} \mathrm{Ext}^1_R({}_\nu A^*, M) \xrightarrow{\mu_{c*}} \mathrm{Ext}^1_R({}_\nu A^*, M).$$

We claim that $\partial$ is an isomorphism. If $a \in {}_\nu A^*$, then $ca = 0$, because $A^*$ is an $R^*$-module (remember that $R^* = R/cR$). Hence, if $f \in \mathrm{Hom}_R({}_\nu A^*, M)$, then $cf(a) = f(ca) = f(0) = 0$. Since $\mu_c\colon M \to M$ is an injection, $f(a) = 0$ for all $a \in A^*$. Thus, $f = 0$, $\ker\partial = \mathrm{Hom}_R({}_\nu A^*, M) = \{0\}$, and $\partial$ is an injection. The map $\mu_{c*}\colon \mathrm{Ext}^1_R({}_\nu A^*, M) \to \mathrm{Ext}^1_R({}_\nu A^*, M)$ is multiplication by $c$, by Example C-3.60. On the other hand, Example C-3.70 shows that if $\mu'_c\colon {}_\nu A^* \to {}_\nu A^*$ is multiplication by $c$, then the induced map ${\mu'_c}^*$ on Ext is also multiplication by $c$. But $\mu'_c = 0$, because $A^*$ is an $(R/cR)$-module, and so ${\mu'_c}^* = 0$. Hence, $\mu_{c*} = {\mu'_c}^* = 0$. Therefore, $\mathrm{im}\,\partial = \ker(\mu_{c*}) = \mathrm{Ext}^1_R({}_\nu A^*, M)$, and so $\partial$ is a surjection. It follows that

$$\partial\colon \mathrm{Hom}_R({}_\nu A^*, M/cM) \to \mathrm{Ext}^1_R({}_\nu A^*, M)$$

is an isomorphism, natural because it is the connecting homomorphism. By Exercise C-5.62 on page 496, there is a natural isomorphism

$$\mathrm{Hom}_{R^*}(A^*, M/cM) \to \mathrm{Hom}_R({}_\nu A^*, M/cM).$$

The composite

$$\mathrm{Hom}_{R^*}(A^*, M/cM) \to \mathrm{Hom}_R({}_\nu A^*, M/cM) \to \mathrm{Ext}^1_R({}_\nu A^*, M) = G^0(A^*)$$

is a natural isomorphism; hence, its inverse defines a natural isomorphism

$$G^0 \to \mathrm{Hom}_{R^*}(\square, M/cM).$$

Setting $L^* = M/cM$ completes the verification of axiom (ii).

It remains to verify axiom (iii): $G^n(P^*) = \{0\}$ whenever $P^*$ is a projective $R^*$-module and $n \geq 1$. In fact, since $G^n$ is an additive functor and every projective is a direct summand of a free module, we may assume that $P^*$ is a free $R^*$-module

with basis, say, $E$. If $Q = \bigoplus_{e\in E} Re$ is the free $R$-module with basis $E$, then there is an exact sequence of $R$-modules, where $\lambda_c$ is multiplication by $c$,

$$0 \to Q \xrightarrow{\lambda_c} Q \to P^* \to 0. \tag{1}$$

The first arrow is an injection because $c$ is not a zero-divisor on $R$; the last arrow is a surjection because

$$Q/cQ = \Big(\bigoplus_{e\in E} Re\Big)/\Big(\bigoplus_{e\in E} Rce\Big) \cong \bigoplus_{e\in E}(R/Rc)e = \bigoplus_{e\in E} R^*e = P^*.$$

The long exact sequence arising from (1) contains

$$\to \operatorname{Ext}^n_R(Q, M) \to \operatorname{Ext}^{n+1}_R({}_\nu P^*, M) \to \operatorname{Ext}^{n+1}_R(Q, M) \to .$$

Since $Q$ is $R$-free and $n \geq 1$, the outside terms are $\{0\}$, and exactness gives $G^n(P^*) = \operatorname{Ext}^{n+1}_R({}_\nu P^*, M) = \{0\}$. Therefore,

$$\operatorname{Ext}^{n+1}_R({}_\nu A^*, M) = G^n(A^*) \cong \operatorname{Ext}^n_{R^*}(A^*, M/cM). \quad \bullet$$

**Theorem C-5.149.** *For every commutative ring $k$,*

$$\mathrm{D}(k[x]) = \mathrm{D}(k) + 1.$$

**Proof.** We have proved that $\mathrm{D}(k[x]) \leq \mathrm{D}(k) + 1$ in Proposition C-5.147, and so it suffices to prove the reverse inequality. We may assume that $\mathrm{D}(k) = n < \infty$, by Corollary C-5.146.

In the notation of the Rees Lemma, Proposition C-5.148, let us write $R = k[x]$, $c = x$, and $R^* = k$. Let $A$ be a $k$-module with $\operatorname{pd}_k(A) = n$. By Exercise C-5.56 on page 496, there is a free $k$-module $F$ with $\operatorname{Ext}^n_k(A, F) \neq \{0\}$; of course, multiplication by $x$ is an injection $F \to F$. As in the proof of the Rees Lemma, there is a free $k[x]$-module $Q = k[x] \otimes_k F$ with $Q/xQ \cong F$. The Rees Lemma gives

$$\operatorname{Ext}^{n+1}_{k[x]}(A, Q) \cong \operatorname{Ext}^n_k({}_\nu A, Q/xQ) \cong \operatorname{Ext}^n_k({}_\nu A, F) \neq \{0\}$$

($A$ is viewed as a $k[x]$-module ${}_\nu A$ via $k[x] \to k$). Therefore, $\operatorname{pd}_{k[x]}(A) \geq n+1$, and so $\mathrm{D}(k[x]) \geq n + 1 = \mathrm{D}(k) + 1$. $\bullet$

**Corollary C-5.150 (Hilbert's Theorem on Syzygies).** *If $k$ is a field, then*

$$\mathrm{D}(k[x_1, \ldots, x_n]) = n.$$

**Proof.** Since $\mathrm{D}(k) = 0$ and $\mathrm{D}(k[x]) = 1$ for every field $k$, the result follows from Theorem C-5.149 by induction on $n \geq 0$. $\bullet$

Hilbert's Theorem on Syzygies implies that if $R = k[x_1, \ldots, x_n]$, where $k$ is a field, then every finitely generated $R$-module $M$ has a resolution

$$0 \to P_n \to P_{n-1} \to \cdots \to P_0 \to M \to 0,$$

where $P_i$ is free for all $i < n$ and $P_n$ is projective.

**Definition.** We say that a (necessarily finitely generated) $R$-module $M$, over an arbitrary commutative ring $R$, has **FFR** (***finite free resolution***) if it has a resolution in which every $P_i$, including the last $P_n$, is a finitely generated free module.

Hilbert's Theorem on Syzygies can be improved to the theorem that if $k$ is a field, then every finitely generated $k[x_1, \ldots, x_n]$-module has FFR (Rotman [**187**], p. 480). (Of course, this result also follows from the more difficult Quillen–Suslin Theorem, which says that every projective $k[x_1, \ldots, x_n]$-module, where $k$ is a field, is free ([**187**], p. 209).)

The next corollary is another application of the Rees Lemma.

**Corollary C-5.151.** *Let $R$ be a commutative ring, let $x \in R$ not be a zero-divisor, and let $x$ be regular on an $R$-module $M$. If $\mathrm{pd}_R(M) = n < \infty$, then*

$$\mathrm{pd}_{R/xR}(M/xM) \leq n - 1.$$

**Proof.** Since $\mathrm{pd}_R(M) = n < \infty$, we have $\mathrm{Ext}_R^{n+1}(L, M) = \{0\}$ for all left $R$-modules $L$; in particular, $\mathrm{Ext}_R^{n+1}(L^*, M) = \{0\}$ for all left $R^*$-modules $L^*$ (by Exercise C-5.61 on page 496, $L^*$ can be viewed as a left $R$-module). By the Rees Lemma with $R^* = R/xR$, we have $\mathrm{Ext}_{R^*}^n(L^*, M/xM) \cong \mathrm{Ext}_R^{n+1}(L^*, M)$. Therefore, $\mathrm{Ext}_{R^*}^n(L^*, M/xM) = \{0\}$ for all $R^*$-modules $L^*$, and so $\mathrm{pd}_{R^*}(M/xM) \leq n-1$; that is, $\mathrm{pd}_{R/xR}(M/xM) \leq n - 1$. •

Here is a change of rings theorem, simpler than the Rees Lemma, which gives another proof of the inequality $\mathrm{D}(k[x]) \geq \mathrm{D}(k)+1$ in the proof of Theorem C-5.149: let $R = k[x]$, so that $R^* = R/(x) = k[x]/(x) = k$.

**Proposition C-5.152 (Kaplansky).** *Let $R$ be a (not necessarily commutative) ring, let $x \in Z(R)$ be neither a unit nor a zero-divisor, let $R^* = R/(x)$, and let $M^*$ be a left $R^*$-module. If $\mathrm{pd}_{R^*}(M^*) = n < \infty$, then $\mathrm{pd}_R(M^*) = n + 1$.*

**Proof.** We note that a left $R^*$-module is merely a left $R$-module $M$ with $xM = \{0\}$.

The proof is by induction on $n \geq 0$. If $n = 0$, then $M^*$ is a projective left $R^*$-module. Since $x$ is not a zero-divisor, there is an exact sequence of left $R$-modules

$$0 \to R \xrightarrow{x} R \to R^* \to 0,$$

so that $\mathrm{pd}_R(R^*) \leq 1$. Now $M^*$, being a projective left $R^*$-module, is a direct summand of a free left $R^*$-module $F^*$. But $\mathrm{pd}_R(F^*) \leq 1$, for it is a direct sum of copies of $R^*$, and so $\mathrm{pd}_R(M^*) \leq 1$. Finally, if $\mathrm{pd}_R(M^*) = 0$, then $M^*$ would be a projective left $R$-module; but this contradicts Exercise B-4.36 on page 490 in Part 1, for $xM = \{0\}$. Therefore, $\mathrm{pd}_R(M^*) = 1$.

If $n \geq 1$, then there is an exact sequence of left $R^*$-modules

$$0 \to K^* \to F^* \to M^* \to 0 \tag{2}$$

with $F^*$ free. Now $\mathrm{pd}_{R^*}(K^*) = n - 1$, so induction gives $\mathrm{pd}_R(K^*) = n$.

If $n = 1$, then $\mathrm{pd}_R(K^*) = 1$ and $\mathrm{pd}_R(K^*) \leq 2$, by Exercise C-5.54 on page 495. There is an exact sequence of left $R$-modules

$$0 \to L \to F \to M^* \to 0 \tag{3}$$

with $F$ free. Since $xM^* = \{0\}$, we have $xF \subseteq \ker(F \to M^*) = L$, and this gives an exact sequence of $R^*$-modules (each term is annihilated by $x$)

$$0 \to L/xF \to F/xF \to M^* \to 0.$$

Thus, $\operatorname{pd}_{R^*}(L/xF) = \operatorname{pd}_{R^*}(M^*) - 1 = 0$, because $F/xF$ is a free $R^*$-module, so the exact sequence of $R^*$-modules

$$0 \to xF/xL \to L/xL \to L/xF \to 0$$

splits. Since $M^* \cong F/L \cong xF/xL$, we see that $M^*$ is a direct summand of $L/xL$. Were $L$ a projective left $R$-module, then $L/xL$ and, hence, $M^*$ would be projective left $R^*$-modules, contradicting $\operatorname{pd}_{R^*}(M^*) = 1$. Exact sequence (3) shows that $\operatorname{pd}_R(M^*) = 1 + \operatorname{pd}_R(L) \geq 2$, and so $\operatorname{pd}_R(M^*) = 2$.

Finally, assume that $n \geq 2$. Exact sequence (2) gives $\operatorname{pd}_{R^*}(K^*) = n - 1 > 1 \geq \operatorname{pd}_R(F^*)$; hence, Exercise C-5.54 gives

$$\operatorname{pd}_R(M^*) = \operatorname{pd}_R(K^*) + 1 = n + 1. \quad \bullet$$

## C-5.6. Commutative Noetherian Rings

We now investigate relations between ideals and prime ideals. These results will be used in the next section to prove the theorems of Auslander–Buchsbaum and Serre about regular local rings. Unless we say otherwise,

**all rings in this section are commutative and noetherian**.

Recall that a prime ideal $\mathfrak{p}$ is *minimal* if there is no prime ideal $\mathfrak{q}$ with $\mathfrak{q} \subsetneq \mathfrak{p}$. In a domain, (0) is the only minimal prime ideal.

**Definition.** A ***prime chain*** of ***length*** $n$ in a commutative ring $R$ is a strictly decreasing chain of prime ideals

$$\mathfrak{p}_0 \supsetneq \mathfrak{p}_1 \supsetneq \cdots \supsetneq \mathfrak{p}_n.$$

The ***height*** $\operatorname{ht}(\mathfrak{p})$ of a prime ideal $\mathfrak{p}$ is the length of the longest prime chain with $\mathfrak{p} = \mathfrak{p}_0$. Thus, $\operatorname{ht}(\mathfrak{p}) \leq \infty$.

**Example C-5.153.**

(i) If $\mathfrak{p}$ is a prime ideal, then $\operatorname{ht}(\mathfrak{p}) = 0$ if and only if $\mathfrak{p}$ is a minimal prime ideal. If $R$ is a domain, then $\operatorname{ht}(\mathfrak{p}) = 0$ if and only if $\mathfrak{p} = (0)$.

(ii) If $R$ is a Dedekind ring and $\mathfrak{p}$ is a nonzero prime ideal in $R$, then $\operatorname{ht}(\mathfrak{p}) = 1$.

(iii) Let $k$ be a field and let $R = k[X]$ be the polynomial ring in infinitely many variables $X = \{x_1, x_2, \dots\}$. If $\mathfrak{p}_i = (x_i, x_{i+1}, \dots)$, then $\mathfrak{p}_i$ is a prime ideal (because $R/\mathfrak{p}_i \cong k[x_1, \dots, x_{i-1}]$ is a domain) and, for every $n \geq 1$,

$$\mathfrak{p}_1 \supsetneq \mathfrak{p}_2 \supsetneq \cdots \supsetneq \mathfrak{p}_{n+1}$$

is a prime chain of length $n$. It follows that $\operatorname{ht}(\mathfrak{p}_1) = \infty$. ◀

**Definition.** The ***Krull dimension*** of a ring $R$ is

$$\dim(R) = \sup\{\operatorname{ht}(\mathfrak{p}) : \mathfrak{p} \in \operatorname{Spec}(R)\};$$

that is, $\dim(R)$ is the length of a longest prime chain in $R$.

Before proceeding further, we present a corollary of Theorem B-6.51 in Part 1; this corollary might well have appeared in Part 1.

**Corollary C-5.154.** *Let $I$ be an ideal in a noetherian ring $R$.*

(i) *Any two normal primary decompositions of $I$ have the same set of isolated prime ideals, and so the isolated prime ideals are uniquely determined by $I$.*

(ii) *$I$ has only finitely many minimal prime ideals.*

(iii) *A noetherian ring has only finitely many minimal prime ideals.*

**Proof.**

(i) Let $I = Q_1 \cap \cdots \cap Q_n$ be a normal primary decomposition. If $P$ is any prime ideal containing $I$, then

$$P \supseteq I = Q_1 \cap \cdots \cap Q_n \supseteq Q_1 \cdots Q_n.$$

Now $P \supseteq Q_i$ for some $i$, by Proposition A-3.82 in Part 1, and so $P \supseteq \sqrt{Q_i} = P_i$. In other words, any prime ideal containing $I$ must contain an isolated associated prime ideal. Hence, the isolated primes are the minimal elements in the set of associated primes of $I$; by Theorem B-6.51 in Part 1, they are uniquely determined by $I$.

(ii) As in part (i), any prime ideal $P$ containing $I$ must contain an isolated prime of $I$. Hence, if $P$ is minimal over $I$, then $P$ must equal an isolated prime ideal of $I$. The result follows, for $I$ has only finitely many isolated prime ideals.

(iii) This follows from part (ii) taking $I = (0)$. •

If $R$ is a Dedekind ring, then $\dim(R) = 1$, for every nonzero prime is a maximal ideal; if $R$ is a domain, then $\dim(R) = 0$ if and only if $R$ is a field. If $R = k[X]$ is a polynomial ring in infinitely many variables, then $\dim(R) = \infty$. The next proposition characterizes the noetherian rings of Krull dimension 0.

**Proposition C-5.155.** *Let $R$ be a commutative ring. Then $R$ is noetherian with* $\dim(R) = 0$ *if and only if every finitely generated $R$-module $M$ has a composition series.*

**Proof.** Assume that $R$ is noetherian with Krull dimension 0. Since $R$ is noetherian, Corollary C-5.154(iii) says that there are only finitely many minimal prime ideals. Since $\dim(R) = 0$, every prime ideal is a minimal prime ideal (as well as a maximal ideal). We conclude that $R$ has only finitely many prime ideals, say, $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$. Now $\operatorname{nil}(R) = \bigcap_{i=1}^{n} \mathfrak{p}_i$ is nilpotent, by Exercise B-6.10 on page 614 in Part 1; say, $\operatorname{nil}(R)^m = (0)$. Define

$$N = \mathfrak{p}_1 \cdots \mathfrak{p}_n \subseteq \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n = \operatorname{nil}(R),$$

so that

$$N^m = (\mathfrak{p}_1 \cdots \mathfrak{p}_n)^m \subseteq \operatorname{nil}(R)^m = (0).$$

Let $M$ be a finitely generated $R$-module, and consider the chain

$$M \supseteq \mathfrak{p}_1 M \supseteq \mathfrak{p}_1\mathfrak{p}_2 M \supseteq \cdots \supseteq NM.$$

The factor module $\mathfrak{p}_1 \cdots \mathfrak{p}_{i-1} M / \mathfrak{p}_1 \cdots \mathfrak{p}_i M$ is an $(R/\mathfrak{p}_i)$-module; that is, it is a vector space over the field $R/\mathfrak{p}_i$ (for $\mathfrak{p}_i$ is a maximal ideal). Since $M$ is finitely generated, the factor module is finite-dimensional, and so the chain can be refined

so that all the factor modules be simple. Finally, repeat this argument for the chains

$$N^jM \supseteq \mathfrak{p}_1N^jM \supseteq \mathfrak{p}_1\mathfrak{p}_2N^jM \supseteq \cdots \supseteq N^{j+1}M.$$

Since $N^m = \{0\}$, we have constructed a composition series for $M$.

Conversely, if every finitely generated $R$-module has a composition series, then the cyclic $R$-module $R$ has a composition series, say, of length $\ell$. It follows that any ascending chain of ideals has length at most $\ell$, and so $R$ is noetherian. To prove that $\dim(R) = 0$, we must show that $R$ does not contain any pair of prime ideals with $\mathfrak{p} \supsetneq \mathfrak{q}$. Passing to the quotient ring $R/\mathfrak{q}$, we may restate the hypotheses: $R$ is a domain having a nonzero prime ideal as well as a composition series $R \supseteq I_1 \supseteq \cdots \supseteq I_d \neq (0)$. The last ideal $I_d$ is a minimal ideal; choose a nonzero element $x \in I_d$. Of course, $xI_d \subseteq I_d$; since $R$ is a domain, $xI_d \neq (0)$, so that minimality of $I_d$ gives $xI_d = I_d$. Hence, there is $y \in I_d$ with $xy = x$; that is, $1 = y \in I_d$, and so $I_d = R$. We conclude that $R$ is a field, contradicting its having a nonzero prime ideal. •

Low-dimensional cases become more interesting once we recognize that they can be used in tandem with formation of quotient rings and localization, for these methods often reduce a more general case to these cases. We shall see this in the proof of the next theorem.

We are going to prove a theorem of Krull, the *Principal Ideal Theorem*, which implies that every prime ideal (in a noetherian ring) has finite height. Our proof is Kaplansky's adaptation of a proof of Rees.

**Lemma C-5.156.** *Let $a$ and $b$ be nonzero elements in a domain $R$. If there exists $c \in R$ such that $ca^2 \in (b)$ implies $ca \in (b)$, then the series $(a, b) \supseteq (a) \supseteq (a^2)$ and $(a^2, b) \supseteq (a^2, ab) \supseteq (a^2)$ have isomorphic factor modules.*[8]

**Proof.** Now $(a, b)/(a) \cong (a^2, ab)/(a^2)$, for multiplication by $a$ sends $(a, b)$ onto $(a^2, ab)$ and $(a)$ onto $(a^2)$.

The module $(a)/(a^2)$ is cyclic with annihilator $(a)$; that is, $(a)/(a^2) \cong R/(a)$. The module $(a^2, b)/(a^2, ab)$ is also cyclic, for the generator $a^2$ lies in $(a^2, ab)$. Now $A = \operatorname{ann}\big((a^2, b)/(a^2, ab)\big)$ contains $(a)$, and so it suffices to prove that $A = (a)$; that is, if $cb = ua^2 + vab$, then $c \in (a)$. This equation gives $ua^2 \in (b)$, and so the hypothesis gives $ua = rb$ for some $r \in R$. Substituting, $cb = rab + vab$, and canceling $b$ gives $c = ra + va \in (a)$. Therefore, $(a)/(a^2) \cong (a^2, b)/(a^2, ab)$. •

**Definition.** A prime ideal $\mathfrak{p}$ is ***minimal over an ideal*** $I$ if $I \subseteq \mathfrak{p}$ and there is no prime ideal $\mathfrak{q}$ with $I \subseteq \mathfrak{q} \subsetneq \mathfrak{p}$; equivalently, $\mathfrak{p}/I$ is a minimal prime ideal in $R/I$.

**Theorem C-5.157 (Principal Ideal Theorem).** *Let $(r)$ be a principal ideal in a commutative noetherian ring $R$. If $(r)$ is proper (that is, if $r$ is not a unit) and $\mathfrak{p}$ is a prime ideal minimal over $(r)$, then* $\operatorname{ht}(\mathfrak{p}) \leq 1$.

[8] Our notation for ideals is not consistent. The principal ideal generated by an element $a \in R$ is sometimes denoted by $(a)$ and sometimes denoted by $Ra$.

**Proof.** If, on the contrary, $\mathrm{ht}(\mathfrak{p}) \geq 2$, then there is a prime chain

$$\mathfrak{p} \supsetneq \mathfrak{p}_1 \supsetneq \mathfrak{p}_2.$$

We normalize the problem in two ways. First, replace $R$ by $R/\mathfrak{p}_2$; second, localize at $\mathfrak{p}/\mathfrak{p}_2$. The hypotheses are modified accordingly: $R$ is now a local domain whose maximal ideal $\mathfrak{m}$ is minimal over a proper principal ideal $(x)$, and there is a prime ideal $\mathfrak{q}$ with

$$R \supsetneq \mathfrak{m} \supsetneq \mathfrak{q} \supsetneq (0).$$

Choose a nonzero element $b \in \mathfrak{q}$, and define

$$I_i = ((b) : x^i) = \{c \in R : cx^i \in (b)\}.$$

The ascending chain $I_1 \subseteq I_2 \subseteq \cdots$ must stop, because $R$ is noetherian: say, $I_n = I_{n+1} = \cdots$. It follows that if $c \in I_{2n}$, then $c \in I_n$; that is, if $cx^{2n} \in (b)$, then $cx^n \in (b)$. If we set $a = x^n$, then $ca^2 \in (b)$ implies $ca \in (b)$.

If $R^* = R/(a^2)$, then $\dim(R^*) = 0$, for it has exactly one prime ideal. By Proposition C-5.155, the $R^*$-module $(a, b)/(a^2)$ (as every finitely generated $R^*$-module) has finite length $\ell$ (the length of its composition series). But Lemma C-5.156 implies that both $(a, b)$ and its submodule $(a^2, b)$ have length $\ell$. The Jordan–Hölder Theorem says that this can happen only if $(a^2, b) = (a, b)$, which forces $a \in (a^2, b)$: thus, there are $s, t \in R$ with $a = sa^2 + tb$. Since $sa \in \mathfrak{m}$, the element $1 - sa$ is a unit (for $R$ is a local ring with maximal ideal $\mathfrak{m}$). Hence, $-a(1 - sa) = tb \in (b)$ gives $a \in (b) \subseteq \mathfrak{q}$. But $a = x^n$ gives $x \in \mathfrak{q}$, contradicting $\mathfrak{m}$ being a prime ideal minimal over $(x)$. •

We now generalize the Principal Ideal Theorem to finitely generated ideals.

**Theorem C-5.158 (Generalized Principal Ideal Theorem).** *If $I = (a_1, \ldots, a_n)$ is a proper ideal in a ring $R$ and $\mathfrak{p}$ is a prime ideal minimal over $I$, then $\mathrm{ht}(\mathfrak{p}) \leq n$.*

**Proof.** The hypotheses still hold after localizing at $\mathfrak{p}$, so we may now assume that $R$ is a local ring with $\mathfrak{p}$ as its maximal ideal.

The proof is by induction on $n \geq 1$, the base step being the Principal Ideal Theorem. Let $I = (a_1, \ldots, a_{n+1})$, and assume, by way of contradiction, that $\mathrm{ht}(\mathfrak{p}) > n + 1$; thus, there is a prime chain

$$\mathfrak{p} = \mathfrak{p}_0 \supsetneq \mathfrak{p}_1 \supsetneq \cdots \supsetneq \mathfrak{p}_{n+1}.$$

We may assume that there are no prime ideals strictly between $\mathfrak{p}$ and $\mathfrak{p}_1$ (if there were such a prime ideal, insert it, thereby lengthening the chain; such insertions can be done only finitely many times, for the module $\mathfrak{p}/\mathfrak{p}_1$ has ACC). Now $I \not\subseteq \mathfrak{p}_1$, because $\mathfrak{p}$ is a prime ideal minimal over $I$. Reindexing the generators of $I$ if necessary, $a_1 \notin \mathfrak{p}_1$. Hence, $(a_1, \mathfrak{p}_1) \supsetneq \mathfrak{p}_1$. We claim that $\mathfrak{p}$ is the only prime ideal containing $(a_1, \mathfrak{p}_1)$; there can be no prime ideal $\mathfrak{p}'$ with $(a_1, \mathfrak{p}_1) \subseteq \mathfrak{p}' \subseteq \mathfrak{p}$ (the second inclusion holds because $\mathfrak{p}$ is the only maximal ideal in $R$) because there are no prime ideals strictly between $\mathfrak{p}$ and $\mathfrak{p}_1$. Therefore, in the ring $R/(a_1, \mathfrak{p}_1)$, the image of $\mathfrak{p}$ is the unique nonzero prime ideal. As such, it must be the nilradical, and hence it is

nilpotent, by Exercise B-6.10 on page 614 in Part 1. There is an integer $m$ with $\mathfrak{p}^m \subseteq (a_1, \mathfrak{p}_1)$, and so there are equations

$$a_i^m = r_i a_1 + b_i, \qquad r_i \in R, \quad b_i \in \mathfrak{p}_1, \quad \text{and} \quad i \geq 2. \tag{1}$$

Define $J = (b_2, \ldots, b_{n+1})$. Now $J \subseteq \mathfrak{p}_1$, while $\operatorname{ht}(\mathfrak{p}_1) > n$. By induction, $\mathfrak{p}_1$ cannot be a prime ideal minimal over $J$, and so there exists a prime ideal $\mathfrak{q}$ minimal over $J$:

$$J \subseteq \mathfrak{q} \subsetneq \mathfrak{p}_1.$$

Now $a_i^m \in (a_1, \mathfrak{q})$ for all $i$, by (1). Thus, any prime ideal $\mathfrak{p}'$ containing $(a_1, \mathfrak{q})$ must contain all $a_i^m$, hence all $a_i$, and hence $I$. As $\mathfrak{p}$ is the unique maximal ideal, $I \subseteq \mathfrak{p}' \subseteq \mathfrak{p}$. But $\mathfrak{p}$ is a prime ideal minimal over $I$, and so $\mathfrak{p}' = \mathfrak{p}$. Therefore, $\mathfrak{p}$ is the unique prime ideal containing $(a_1, \mathfrak{q})$. If $R^* = R/\mathfrak{q}$, then $\mathfrak{p}^* = \mathfrak{p}/\mathfrak{q}$ is a prime ideal minimal over the principal ideal $(a_1 + \mathfrak{q})$. On the other hand, $\operatorname{ht}(\mathfrak{p}^*) \geq 2$, for $\mathfrak{p}^* \supsetneq \mathfrak{p}_1^* \supsetneq (0)$ is a prime chain, where $\mathfrak{p}_1^* = \mathfrak{p}_1/\mathfrak{q}$. This contradiction to the Principal Ideal Theorem completes the proof. •

**Corollary C-5.159.** *Every prime ideal in a (noetherian) ring $R$ has finite height, and so* $\operatorname{Spec}(R)$ *has* DCC.

**Proof.** Every prime ideal $\mathfrak{p}$ is finitely generated, because $R$ is noetherian; say, $\mathfrak{p} = (a_1, \ldots, a_n)$. But $\mathfrak{p}$ is a minimal prime ideal over itself, so that Theorem C-5.158 gives $\operatorname{ht}(\mathfrak{p}) \leq n$. •

A noetherian ring may have infinite Krull dimension, for there may be no uniform bound on the length of prime chains. We will see that this cannot happen for local rings.

The Generalized Principal Ideal Theorem bounds the height of a prime ideal that is minimal over an ideal; more generally, the next result bounds the height of a prime ideal that merely contains an ideal, but which may not be minimal over it.

**Corollary C-5.160.** *Let $I = (a_1, \ldots, a_n)$ be an ideal in $R$, and let $\mathfrak{p}$ be a prime ideal in $R$ with $\mathfrak{p} \supseteq I$. If $\operatorname{ht}(\mathfrak{p}/I)$ denotes the height of $\mathfrak{p}/I$ in $R/I$, then*

$$\operatorname{ht}(\mathfrak{p}) \leq n + \operatorname{ht}(\mathfrak{p}/I).$$

**Proof.** The proof is by induction on $h = \operatorname{ht}(\mathfrak{p}/I) \geq 0$. If $h = 0$, then Exercise C-5.68 on page 510 says that $\mathfrak{p}$ is minimal over $I$, and so the base step is the Generalized Principal Ideal Theorem. For the inductive step $h > 0$, $\mathfrak{p}$ is not minimal over $I$. By Corollary C-5.154(iii), there are only finitely many minimal primes in $R/I$, and so Exercise C-5.68 on page 510 says that there are only finitely many prime ideals minimal over $I$, say, $\mathfrak{q}_1, \ldots, \mathfrak{q}_s$. Since $\mathfrak{p}$ is not minimal over $I$, $\mathfrak{p} \not\subseteq \mathfrak{q}_i$ for any $i$; hence, Proposition A-3.82 in Part 1 says that $\mathfrak{p} \not\subseteq \mathfrak{q}_1 \cup \cdots \cup \mathfrak{q}_s$, and so there is $y \in \mathfrak{p}$ with $y \notin \mathfrak{q}_i$ for any $i$. Define $J = (I, y)$.

We now show, in $R/J$, that $\operatorname{ht}(\mathfrak{p}/J) \leq h - 1$. Let

$$\mathfrak{p}/J \supsetneq \mathfrak{p}_1/J \supsetneq \cdots \supsetneq \mathfrak{p}_r/J$$

be a prime chain in $R/J$. Since $I \subsetneq J$, there is a surjective ring map $R/I \to R/J$. The prime chain lifts to a prime chain in $R/I$:

$$\mathfrak{p}/I \supsetneq \mathfrak{p}_1/I \supsetneq \cdots \supsetneq \mathfrak{p}_r/I.$$

Now $\mathfrak{p}_r \supseteq J \supsetneq I$, and $J = (I, y)$ does not contain any $\mathfrak{q}_i$. But the ideals $\mathfrak{q}_i/I$ are the minimal prime ideals in $R/I$, by Exercise C-5.68 on page 510, so that $\mathfrak{p}_r$ is not a minimal prime ideal in $R$. Therefore, there is a prime chain starting at $\mathfrak{p}$ of length $r+1$. We conclude that $r+1 \leq h$, and so $\operatorname{ht}(\mathfrak{p}/J) \leq h-1$.

Since $J = (I, y) = (a_1, \ldots, a_n, y)$ is generated by $n+1$ elements, the inductive hypothesis gives

$$\operatorname{ht}(\mathfrak{p}) \leq n+1+\operatorname{ht}(\mathfrak{p}/J) = (n+1)+(h-1) = n+h = n+\operatorname{ht}(\mathfrak{p}/I). \quad \bullet$$

Recall that a nonzero element $c \in R$ is *regular* on $R$ if the multiplication map $\mu_c\colon R \to R$, given by $r \mapsto cr$, is an injection; that is, $c$ is not a zero-divisor.

**Definition.** A sequence $x_1, \ldots, x_n$ in a ring $R$ is an ***R-sequence*** if $x_1$ is regular on $R$, $x_2$ is regular on $R/(x_1)$, $x_3$ is regular on $R/(x_1, x_2), \ldots, x_n$ is regular on $R/(x_1, \ldots, x_{n-1})$.

For example, if $R = k[x_1, \ldots, x_n]$ is a polynomial ring over a field $k$, then it is easy to see that $x_1, \ldots, x_n$ is an $R$-sequence. Exercise C-5.67 on page 510 gives an example of a permutation of an $R$-sequence that is not an $R$-sequence. However, if $R$ is local, then every permutation of an $R$-sequence is also an $R$-sequence (Kaplansky [**118**], p. 86).

The Generalized Principal Ideal Theorem gives an upper bound on the height of a prime ideal; the next lemma gives a lower bound.

**Lemma C-5.161.**

(i) *If $x$ is not a zero-divisor in a ring $R$, then $x$ lies in no minimal prime ideal.*

(ii) *If $\mathfrak{p}$ is a prime ideal in $R$ and $x \in \mathfrak{p}$ is not a zero-divisor, then*

$$1+\operatorname{ht}(\mathfrak{p}/(x)) \leq \operatorname{ht}(\mathfrak{p}).$$

(iii) *If a prime ideal $\mathfrak{p}$ in $R$ contains an $R$-sequence $x_1, \ldots, x_d$, then*

$$d \leq \operatorname{ht}(\mathfrak{p}).$$

**Proof.**

(i) Suppose, on the contrary, that $\mathfrak{p}$ is a nonzero minimal prime ideal containing $x$. Now $R_{\mathfrak{p}}$ is a ring with only one nonzero prime ideal, namely, $\mathfrak{p}_{\mathfrak{p}}$, which must be the nilradical. Thus, $x/1$, as every element in $\mathfrak{p}_{\mathfrak{p}}$, is nilpotent. If $x^m/1 = 0$ in $R_{\mathfrak{p}}$, then there is $s \notin \mathfrak{p}$ (so that $s \neq 0$) with $sx = 0$, contradicting $x$ not being a zero-divisor.

(ii) If $h = \operatorname{ht}(\mathfrak{p}/(x))$, then there is a prime chain in $R/(x)$:

$$\mathfrak{p}/(x) \supsetneq \mathfrak{p}_1/(x) \supsetneq \cdots \supsetneq \mathfrak{p}_h/(x).$$

Lifting back to $R$, there is a prime chain $\mathfrak{p} \supsetneq \mathfrak{p}_1 \supsetneq \cdots \supsetneq \mathfrak{p}_h$ with $\mathfrak{p}_h \supseteq (x)$. Since $x$ is not a zero-divisor, part (i) says that $\mathfrak{p}_h$ is not a minimal prime. Therefore, there exists a prime ideal $\mathfrak{p}_{h+1}$ properly contained in $\mathfrak{p}_h$, which shows that $\operatorname{ht}(\mathfrak{p}) \geq 1+h$.

(iii) The proof is by induction on $d \geq 1$. For the base step $d = 1$, suppose, on the contrary, that $\operatorname{ht}(\mathfrak{p}) = 0$; then $\mathfrak{p}$ is a minimal prime ideal, and this contradicts part (i). For the inductive step, part (ii) gives $\operatorname{ht}(\mathfrak{p}/(x_1)) + 1 \leq \operatorname{ht}(\mathfrak{p})$. Now $\mathfrak{p}/(x_1)$ contains an $(R/(x_1))$-sequence $x_2 + (x_1), \ldots, x_d + (x_1)$, by Exercise C-5.72(ii) on page 524, so that the inductive hypothesis gives $d - 1 \leq \operatorname{ht}(\mathfrak{p}/(x_1))$. Therefore, part (ii) gives $d \leq \operatorname{ht}(\mathfrak{p})$. •

**Definition.** A subset $X$ of an $R$-module $M$ is ***scalar-closed*** if $x \in X$ implies that $rx \in X$ for all $r \in R$.

Every submodule of a module $M$ is scalar-closed. An example of a scalar-closed subset (of $R$) that is not a submodule is

$$\operatorname{Zer}(R) = \{r \in R : r = 0 \text{ or } r \text{ is a zero-divisor}\}.$$

**Definition.** Let $X \subseteq M$ be a scalar-closed subset. The ***annihilator*** of $x \in X$ is

$$\operatorname{ann}(x) = \{r \in R : rx = 0\},$$

the ***annihilator*** of $X$ is

$$\operatorname{ann}(X) = \{r \in R : rx = 0 \text{ for all } x \in X\},$$

and

$$\mathcal{A}(X) = \{\operatorname{ann}(x) : x \in X \text{ and } x \neq 0\}.$$

Note that $\operatorname{ann}(x)$ and $\operatorname{ann}(X)$ are ideals ($\operatorname{ann}(0) = R$, and so $\mathcal{A}(X)$ is a family of proper ideals if $M \neq \{0\}$).

**Lemma C-5.162.** *Let $X$ be a nonempty scalar-closed subset of a nonzero finitely generated $R$-module $M$.*

(i) *An ideal $I$ maximal among the ideals in $\mathcal{A}(X)$ is a prime ideal.*

(ii) *There is a descending chain*

$$M = M_0 \supseteq M_1 \supseteq M_2 \supseteq \cdots \supseteq M_n = \{0\}$$

*whose factor modules $M_i/M_{i+1} \cong R/\mathfrak{p}_i$ for prime ideals $\mathfrak{p}_i$.*

**Proof.**

(i) Since $R$ is noetherian, the nonempty family $\mathcal{A}(X)$ contains a maximal element, by Proposition B-1.10 in Part 1; call it $I = \operatorname{ann}(x)$. Suppose that $a, b$ are elements in $R$ with $ab \in I$ and $b \notin I$; that is, $abx = 0$ but $bx \neq 0$. Thus, $\operatorname{ann}(bx) \supseteq I + Ra \supseteq I$. If $a \notin I$, then $\operatorname{ann}(bx) \supseteq I + Ra \supsetneq I$. But $bx \in X$ because $X$ is scalar-closed, so that $\operatorname{ann}(bx) \in \mathcal{A}(X)$, which contradicts the maximality of $I = \operatorname{ann}(x)$. Therefore, $a \in I$ and $I$ is a prime ideal.

(ii) Since $R$ is noetherian and $M$ is finitely generated, $M$ has the maximum condition on ideals. Thus, the nonempty family $\mathcal{A}(M)$ has a maximal element, say, $\mathfrak{p}_1 = \operatorname{ann}(x_1)$, which is prime, by part (i). Define $M_1 = \langle x_1 \rangle$, and note that $M_1 \cong R/\operatorname{ann}(x_1) = R/\mathfrak{p}_1$. Now repeat this procedure. Let $\mathfrak{p}_2 = \operatorname{ann}(x_2 + M_1)$ be a maximal element of $\mathcal{A}(M/M_1)$, so that $\mathfrak{p}_2$ is prime, and define $M_2 = \langle x_2, x_1 \rangle$. Note that $\{0\} \subseteq M_1 \subseteq M_2$ and that $M_2/M_1 \cong R/\operatorname{ann}(x_2 + M_1) = R/\mathfrak{p}_2$. The module $M$ has ACC, and so this process

terminates, say, with $M^* \subseteq M$. We must have $M^* = M$, however, lest the process continue for another step. Now reindex the subscripts to get the desired statement. •

**Lemma C-5.163 (Prime Avoidance).** *Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ be prime ideals in a ring $R$. If $J$ is an ideal with $J \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$, then $J$ is contained in some $\mathfrak{p}_i$.*

**Proof.** The proof is by induction on $n \geq 1$, and the base step is trivially true. For the inductive step, let $J \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_{n+1}$, and define

$$D_i = \mathfrak{p}_1 \cup \cdots \cup \widehat{\mathfrak{p}}_i \cup \cdots \cup \mathfrak{p}_{n+1}.$$

We may assume that $J \subsetneq D_i$ for all $i$, since otherwise the inductive hypothesis can be invoked to complete the proof. Hence, for each $i$, there exists $a_i \in J$ with $a_i \notin D_i$; since $J \subseteq D_i \cup \mathfrak{p}_i$, we must have $a_i \in \mathfrak{p}_i$. Consider the element

$$b = a_1 + a_2 \cdots a_{n+1}.$$

Now $b \in J$ because all the $a_i$ are. We claim that $b \notin \mathfrak{p}_1$. Otherwise, $a_2 \cdots a_{n+1} = b - a_1 \in \mathfrak{p}_1$; but $\mathfrak{p}_1$ is a prime ideal, and so $a_i \in \mathfrak{p}_1$ for some $i \geq 2$. This is a contradiction, for $a_i \in \mathfrak{p}_1 \subseteq D_i$ and $a_i \notin D_i$. Therefore, $b \notin \mathfrak{p}_i$ for any $i$, contradicting $J \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$. •

**Proposition C-5.164.** *There are finitely many prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ with*

$$\operatorname{Zer}(R) = \{r \in R : r = 0 \text{ or } r \text{ is a zero-divisor}\} \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n.$$

**Proof.** Since $\operatorname{Zer}(R)$ is scalar-closed, Lemma C-5.162(i) shows that any ideal $I$ that is a maximal member of the family $\mathcal{A}(X) = (\operatorname{ann}(x))_{x \in \operatorname{Zer}(R)}$ is prime (maximal members exist because $R$ is noetherian). Let $(\mathfrak{p}_\alpha)_{\alpha \in A}$ be the family of all such maximal members. If $x$ is a zero-divisor, then there is a nonzero $r \in R$ with $rx = 0$; that is, $x \in \operatorname{ann}(r)$ (of course, $r \in \operatorname{Zer}(R)$). It follows that every zero-divisor $x$ lies in some $\mathfrak{p}_\alpha$, and so $\operatorname{Zer}(R) \subseteq \bigcup_{\alpha \in A} \mathfrak{p}_\alpha$. It remains to prove that we may choose the index set $A$ to be finite.

Each $\mathfrak{p}_\alpha = \operatorname{ann}(x_\alpha)$ for some $x_\alpha \in X$; let $S$ be the submodule of $M$ generated by all the $x_\alpha$. Since $R$ is noetherian and $M$ is finitely generated, the submodule $S$ is generated by finitely many of the $x_\alpha$; say, $S = \langle x_1, \ldots, x_n \rangle$. We claim that $\operatorname{ann}(X) \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$, and it suffices to prove that $\mathfrak{p}_\alpha \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$ for all $\alpha$. Now $\mathfrak{p}_\alpha = \operatorname{ann}(x_\alpha)$, and $x_\alpha \in S$; hence,

$$x_\alpha = r_1 x_1 + \cdots + r_n x_n$$

for $r_i \in R$. If $a \in \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n = \operatorname{ann}(x_1) \cap \cdots \cap \operatorname{ann}(x_n)$, then $a x_i = 0$ for all $i$ and so $a x_\alpha = 0$. Therefore,

$$\mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n \subseteq \operatorname{ann}(x_\alpha) = \mathfrak{p}_\alpha.$$

But $\mathfrak{p}_\alpha$ is prime, so that $\mathfrak{p}_i \subseteq \mathfrak{p}_\alpha$ for some $i$,[9] and this contradicts the maximality of $\mathfrak{p}_i$. •

[9] Assume that $I_1 \cap \cdots \cap I_n \subseteq \mathfrak{p}$, where $\mathfrak{p}$ is prime. If $I_i \not\subseteq \mathfrak{p}$ for all $i$, then there are $u_i \in I_i$ with $u_i \notin \mathfrak{p}$. But $u_1 \cdots u_n \in I_1 \cap \cdots \cap I_n \subseteq \mathfrak{p}$; since $\mathfrak{p}$ is prime, some $u_i \in \mathfrak{p}$, a contradiction.

## Exercises

**C-5.64.** If $R = k[x_1, \dots, x_n]$, where $k$ is a field, and $\mathfrak{p} = (x_1, \dots, x_n)$, prove that $\operatorname{ht}(\mathfrak{p}) = n$.

**C-5.65.** Let $R$ be a commutative ring with FFR. Prove that every finitely generated projective $R$-module $P$ has a finitely generated free complement; that is, there is a finitely generated free $R$-module $F$ such that $P \oplus F$ is a free $R$-module. Compare with Exercise C-4.64 on page 417.

**C-5.66.** If $x_1, x_2, \dots, x_n$ is an $R$-sequence on an $R$-module $M$, prove that $(x_1) \subseteq (x_1, x_2) \subseteq \cdots \subseteq (x_1, x_2, \dots, x_n)$ is a strictly ascending chain.

$*$ **C-5.67.** Let $R = k[x, y, z]$, where $k$ is a field.

(i) Prove that $x, y(1-x), z(1-x)$ is an $R$-sequence.

(ii) Prove that $y(1-x), z(1-x), x$ is not an $R$-sequence.

$*$ **C-5.68.** Let $R$ be a commutative ring. Prove that a prime ideal $\mathfrak{p}$ in $R$ is minimal over an ideal $I$ if and only if $\operatorname{ht}(\mathfrak{p}/I) = 0$ in $R/I$.

$*$ **C-5.69.** If $k$ is a field, prove that $k[[x_1, \dots, x_n]]$ is noetherian.

**Hint.** Define the ***order*** $o(f)$ of a nonzero formal power series $f = (f_0, f_1, f_2, \dots)$ to be the smallest $n$ with $f_n \neq 0$. Find a proof similar to that of the Hilbert Basis Theorem (Zariski–Samuel [**234**], p. 138).

## C-5.7. Regular Local Rings

We are going to prove that local rings have finite global dimension if and only if they are *regular local* rings (such rings arise quite naturally in algebraic geometry), in which case they are UFDs. Let us begin with a localization result.

**All rings in this section are commutative and noetherian.**

**Proposition C-5.165.**

(i) *If $A$ is a finitely generated $R$-module, then*

$$\operatorname{pd}(A) = \sup_{\mathfrak{m}} \operatorname{pd}(A_{\mathfrak{m}}),$$

*where $\mathfrak{m}$ ranges over all the maximal ideals of $R$.*

(ii) *For every commutative ring $R$, we have*

$$\operatorname{D}(R) = \sup_{\mathfrak{m}} \operatorname{D}(R_{\mathfrak{m}}),$$

*where $\mathfrak{m}$ ranges over all the maximal ideals of $R$.*

**Proof.**

(i) We first prove that $\operatorname{pd}(A) \geq \operatorname{pd}(A_{\mathfrak{m}})$ for every maximal ideal $\mathfrak{m}$. There is nothing to prove if $\operatorname{pd}(A) = \infty$, and so we may assume that $\operatorname{pd}(A) = n < \infty$. Thus, there is a projective resolution

$$0 \to P_n \to P_{n-1} \to \cdots \to P_0 \to A \to 0.$$

Since $R_{\mathfrak{m}}$ is a flat $R$-module, by Theorem C-5.30,

$$0 \to R_{\mathfrak{m}} \otimes P_n \to R_{\mathfrak{m}} \otimes P_{n-1} \to \cdots \to R_{\mathfrak{m}} \otimes P_0 \to A_{\mathfrak{m}} \to 0$$

is exact; it is a projective resolution of $A_{\mathfrak{m}}$ because all $R_{\mathfrak{m}} \otimes P_i$ are projective $R_{\mathfrak{m}}$-modules. Therefore, $\operatorname{pd}(A_{\mathfrak{m}}) \leq n$. (Neither hypothesis $R$ noetherian nor $A$ finitely generated is needed for this implication.)

For the reverse inequality, we may assume that $\sup_{\mathfrak{m}} \operatorname{pd}(A_{\mathfrak{m}}) = n < \infty$. Since $R$ is noetherian, Theorem C-5.143(i) says that $\operatorname{pd}(A) = \operatorname{fd}(A)$. Now $\operatorname{pd}(A_{\mathfrak{m}}) \leq n$ if and only if $\operatorname{Tor}_{n+1}^{R_{\mathfrak{m}}}(A_{\mathfrak{m}}, B_{\mathfrak{m}}) = \{0\}$ for all $R_{\mathfrak{m}}$-modules $B_{\mathfrak{m}}$, by Lemma C-5.134. However, Proposition C-5.39 gives $\operatorname{Tor}_{n+1}^{R_{\mathfrak{m}}}(A_{\mathfrak{m}}, B_{\mathfrak{m}}) \cong \big(\operatorname{Tor}_{n+1}^{R}(A, B)\big)_{\mathfrak{m}}$. Therefore, $\operatorname{Tor}_{n+1}^{R}(A, B) = \{0\}$, by Proposition C-5.38(i). We conclude that $n \geq \operatorname{pd}(A)$.

(ii) This follows at once from part (i), for $\operatorname{D}(R) = \sup_A\{\operatorname{pd}(A)\}$, where $A$ ranges over all finitely generated (even cyclic) $R$-modules (Theorem C-5.130). •

**Definition.** If $R$ is a local ring with maximal ideal $\mathfrak{m}$, then $k = R/\mathfrak{m}$ is called its ***residue field***. We shall say that

$$(R, \mathfrak{m}, k)$$

is a local ring.

Theorem C-5.130 allows us to compute global dimension as the supremum of projective dimensions of cyclic modules. When $(R, \mathfrak{m}, k)$ is a local ring, there is a dramatic improvement: global dimension is determined by the projective dimension of one cyclic module—its residue field $k$.

**Lemma C-5.166.** *Let $(R, \mathfrak{m}, k)$ be a local ring. If $A$ is a finitely generated $R$-module, then*

$$\operatorname{pd}(A) \leq n \quad \textit{if and only if} \quad \operatorname{Tor}_{n+1}^{R}(A, k) = \{0\}.$$

**Proof.** Suppose $\operatorname{pd}(A) \leq n$. By Example C-5.131, we have $\operatorname{fd}(A) \leq \operatorname{pd}(A)$, so that $\operatorname{Tor}_{n+1}^{R}(A, B) = \{0\}$ for every $R$-module $B$. In particular, $\operatorname{Tor}_{n+1}^{R}(A, k) = \{0\}$.

We prove the converse by induction on $n \geq 0$. For the base step $n = 0$, we must prove that $\operatorname{Tor}_1^{R}(A, k) = \{0\}$ implies $\operatorname{pd}(A) = 0$; that is, $A$ is projective (hence free, since $R$ is local, by Proposition C-5.26). Let $\{a_1, \ldots, a_r\}$ be a ***minimal set of generators*** of $A$ (that is, no proper subset generates $A$),[10] let $F$ be the free $R$-module with basis $\{e_1, \ldots, e_r\}$, and let $\varphi\colon F \to A$ be the $R$-map with $\varphi(e_i) = a_i$. There is an exact sequence $0 \to N \xrightarrow{i} F \xrightarrow{\varphi} A \to 0$, where $N = \ker \varphi$ and $i$ is the

[10] A minimal generating set for $A = (2)$ in $\mathbb{Z}$ is $\{4, 6\}$; of course, there is a generating set of smaller cardinality.

inclusion. Since $R$ is noetherian, the submodule $N$ is finitely generated. For later use, we observe, as in the proof of Proposition C-5.26, that

$$N \subseteq \mathfrak{m}F.$$

Now the sequence $0 \to N \otimes_R k \xrightarrow{i\otimes 1} F \otimes_R k \xrightarrow{\varphi\otimes 1} A \otimes_R k \to 0$ is exact, because $\operatorname{Tor}_1^R(A,k) = \{0\}$. Hence, $i \otimes 1\colon N \otimes_R k \to F \otimes_R k$ is an injection.

Since $\square \otimes_R N$ is right exact, exactness of $0 \to \mathfrak{m} \xrightarrow{j} R \to k \to 0$ gives exactness of $\mathfrak{m}\otimes_R N \xrightarrow{j\otimes 1} R\otimes_R N \to k\otimes_R N \to 0$, so that $k\otimes_R N \cong R\otimes_R N/\operatorname{im}(j\otimes 1)$. Under the isomorphism $\lambda\colon R\otimes_R N \to N$, given by $r\otimes n \mapsto rn$, we have $\lambda(\operatorname{im}(j\otimes 1)) = \mathfrak{m}N$. Indeed, the map $\tau_N\colon N \otimes_R k \to N/\mathfrak{m}N$, given by $n \otimes b \mapsto n + \mathfrak{m}N$ (where $n \in N$ and $b \in k$), is a natural isomorphism. Thus, there is a commutative diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & N \otimes_R k & \xrightarrow{i\otimes 1} & F \otimes_R k \\ & & \downarrow{\scriptstyle \tau_N} & & \downarrow{\scriptstyle \tau_F} \\ & & N/\mathfrak{m}N & \xrightarrow[\overline{i}]{} & F/\mathfrak{m}F \end{array}$$

where $\overline{i}\colon n + \mathfrak{m}N \mapsto n + \mathfrak{m}F$. Since $i \otimes 1$ is an injection, so is $\overline{i}$. But $N \subseteq \mathfrak{m}F$ implies that the map $\overline{i}$ is the zero map. Therefore, $N/\mathfrak{m}N = \{0\}$ and $N = \mathfrak{m}N$. By Corollary C-2.8, Nakayama's Lemma, $N = \{0\}$, and so $\varphi\colon F \to A$ is an isomorphism; that is, $A$ is free.

For the inductive step, we must prove that if $\operatorname{Tor}_{n+2}^R(A,k) = \{0\}$, then $\operatorname{pd}(A) \leq n+1$. Take a projective resolution $\mathbf{P}$ of $A$, and let $\Omega_n(A,\mathbf{P})$ be its $n$th syzygy. Since $\mathbf{P}$ is also a flat resolution of $A$, we have $Y_n(A,\mathbf{P}) = \Omega_n(A,\mathbf{P})$. By Proposition C-5.132, $\operatorname{Tor}_{n+2}^R(A,k) \cong \operatorname{Tor}_1^R(Y_n(A,\mathbf{P}),k)$. The base step shows that $Y_n(A,\mathbf{P}) = \Omega_n(A,\mathbf{P})$ is free, and this gives $\operatorname{pd}(A) \leq n+1$, by Lemma C-5.119. •

**Corollary C-5.167.** *Let $(R,\mathfrak{m},k)$ be a local ring. If $A$ is a finitely generated $R$-module, then*

$$\operatorname{pd}(A) = \sup\{i : \operatorname{Tor}_i^R(A,k) \neq \{0\}\}.$$

**Proof.** If $n = \sup\{i : \operatorname{Tor}_i^R(A,k) \neq \{0\}\}$, then $\operatorname{pd}(A) \leq n$. But $\operatorname{pd}(A) \not< n$; that is, $\operatorname{pd}(A) = n$. •

**Theorem C-5.168.** *Let $(R,\mathfrak{m},k)$ be a local ring.*

(i) $\operatorname{D}(R) \leq n$ *if and only if* $\operatorname{Tor}_{n+1}^R(k,k) = \{0\}$.

(ii) $\operatorname{D}(R) = \operatorname{pd}(k)$.

**Proof.**

(i) If $\operatorname{D}(R) \leq n$, then Lemma C-5.166 gives $\operatorname{Tor}_{n+1}^R(k,k) = \{0\}$.

Conversely, if $\operatorname{Tor}_{n+1}^R(k,k) = \{0\}$, the same lemma gives $\operatorname{pd}(k) \leq n$. By Lemma C-5.134, we have $\operatorname{Tor}_{n+1}^R(A,k) = \{0\}$ for every $R$-module $A$. In particular, if $A$ is finitely generated, then Lemma C-5.166 gives $\operatorname{pd}(A) \leq n$.

Finally, Theorem C-5.130 shows that $\mathrm{D}(R) = \sup_A\{\mathrm{pd}(A)\}$, where $A$ ranges over all finitely generated (even cyclic) $R$-modules. Therefore, $\mathrm{D}(R) \leq n$.

(ii) Immediate from part (i). •

If $(R, \mathfrak{m}, k)$ is a local ring, then $\mathfrak{m}/\mathfrak{m}^2$ is an $(R/\mathfrak{m})$-module; that is, it is a vector space over $k$. Recall that a generating set $X$ of a module $M$ (over any ring) is *minimal* if no proper subset of $X$ generates $M$.

**Proposition C-5.169.** *Let $(R, \mathfrak{m}, k)$ be a local ring.*

(i) *Elements $x_1, \ldots, x_d$ form a minimal generating set for $\mathfrak{m}$ if and only if the cosets $x_i^* = x_i + \mathfrak{m}^2$ form a basis of $\mathfrak{m}/\mathfrak{m}^2$.*

(ii) *Any two minimal generating sets of $\mathfrak{m}$ have the same number of elements.*

**Proof.**

(i) If $x_1, \ldots, x_d$ is a minimal generating set for $\mathfrak{m}$, then $X^* = x_1^*, \ldots, x_d^*$ spans the vector space $\mathfrak{m}/\mathfrak{m}^2$. If $X^*$ is linearly dependent, then there is some $x_i^* = \sum_{j\neq i} r_j' x_J^*$, where $r_j' \in k$. Lifting this equation to $\mathfrak{m}$, we have $x_i \in \sum_{j\neq i} r_j x_j + \mathfrak{m}^2$. Thus, if $B = \langle x_j : j \neq i\rangle$, then $B + \mathfrak{m}^2 = \mathfrak{m}$. Hence,

$$\mathfrak{m}\,(\mathfrak{m}/B) = (B + \mathfrak{m}^2)/B = \mathfrak{m}/B.$$

By Nakayama's Lemma, $\mathfrak{m}/B = \{0\}$, and so $\mathfrak{m} = B$. This contradicts $x_1, \ldots, x_d$ being a minimal generating set. Therefore, $X^*$ is linearly independent, and hence it is a basis of $\mathfrak{m}/\mathfrak{m}^2$.

Conversely, assume that $x_1^*, \ldots, x_d^*$ is a basis of $\mathfrak{m}/\mathfrak{m}^2$, where $x_i^* = x_i + \mathfrak{m}^2$. If we define $A = \langle x_1, \ldots, x_d\rangle$, then $A \subseteq \mathfrak{m}$. If $y \in \mathfrak{m}$, then $y^* = \sum r_i' x_i^*$, where $r_i' \in k$, so that $y \in A + \mathfrak{m}^2$. Hence, $\mathfrak{m} = A + \mathfrak{m}^2$, and, as in the previous paragraph, Nakayama's Lemma gives $\mathfrak{m} = A$; that is, $x_1, \ldots, x_d$ generate $\mathfrak{m}$. If a proper subset of $x_1, \ldots, x_d$ generates $\mathfrak{m}$, then the vector space $\mathfrak{m}/\mathfrak{m}^2$ could be generated by fewer than $d$ elements, contradicting $\dim_k(\mathfrak{m}/\mathfrak{m}^2) = d$.

(ii) The number of elements in any minimal generating set is $\dim_k(\mathfrak{m}/\mathfrak{m}^2)$. •

**Definition.** If $(R, \mathfrak{m}, k)$ is a local ring, then $\mathfrak{m}/\mathfrak{m}^2$ is a finite-dimensional vector space over $k$. Write

$$V(R) = \dim_k(\mathfrak{m}/\mathfrak{m}^2).$$

Proposition C-5.169 shows that all minimal generating sets of $\mathfrak{m}$ have the same number of elements, namely, $V(R)$.

**Proposition C-5.170.** *Let $(R, \mathfrak{m}, k)$ be a local ring. If $x \in \mathfrak{m} - \mathfrak{m}^2$, define $R^* = R/(x)$ and $\mathfrak{m}^* = \mathfrak{m}/(x)$. Then $(R^*, \mathfrak{m}^*, k)$ is a local ring, and*

$$V(R) = V(R^*) + 1.$$

**Proof.** The reader may show that $\mathfrak{m}^*$ is the unique maximal ideal in $R^*$. Note that $R^*/\mathfrak{m}^* = [R/(x)]/[\mathfrak{m}/(x)] \cong R/\mathfrak{m} = k$.

Let $\{y_1^*, \ldots, y_t^*\}$ be a minimal generating set of $\mathfrak{m}^*$, and let $y_i^* = y_i + \mathfrak{m}$. It is clear that $\{x, y_1, \ldots, y_t\}$ generates $\mathfrak{m}$, and we now show that it is a minimal generating set; that is, their cosets mod $\mathfrak{m}$ form a basis of $\mathfrak{m}/\mathfrak{m}^2$.

If $rx + \sum_i r_i y_i \in \mathfrak{m}^2$, where $r, r_i \in R$, then we must show that each term lies in $\mathfrak{m}^2$; that is, all $r, r_i \in \mathfrak{m}$. Passing to $R^*$, we have $\sum_i r_i^* y_i^* \in \mathfrak{m}^2$, because $r^* x^* = 0$ (where $^*$ denotes coset mod $(x)$). But $\{y_1^*, \ldots, y_t^*\}$ is a basis of $\mathfrak{m}^*/(\mathfrak{m}^*)^2$, so that $r_i^* \in \mathfrak{m}^*$ and $r_i \in \mathfrak{m}$ for all $i$. Therefore, $rx \in \mathfrak{m}^2$. But $x \notin \mathfrak{m}^2$, and so $r \in \mathfrak{m}$, as desired. •

**Corollary C-5.171.** *If $(R, \mathfrak{m}, k)$ is a local ring, then* $\operatorname{ht}(\mathfrak{m}) \leq V(R)$, *and*

$$\dim(R) \leq V(R).$$

**Proof.** If $V(R) = d$, then $\mathfrak{m} = (x_1, \ldots, x_d)$. Since $\mathfrak{m}$ is obviously a minimal prime over itself, Theorem C-5.158, the Generalized Principal Ideal Theorem, gives $\operatorname{ht}(\mathfrak{m}) \leq d = V(R)$.

If $\mathfrak{p} \neq \mathfrak{m}$ is a prime ideal in $R$, then any prime chain, $\mathfrak{p} = \mathfrak{p}_0 \supsetneq \mathfrak{p}_1 \supsetneq \cdots \supsetneq \mathfrak{p}_h$, can be lengthened to a prime chain $\mathfrak{m} \supsetneq \mathfrak{p}_0 \supsetneq \mathfrak{p}_1 \supsetneq \cdots \supsetneq \mathfrak{p}_h$ of length $h + 1$. Therefore, $h < V(R)$, and so $\dim(R) = \operatorname{ht}(\mathfrak{m}) \leq V(R)$. •

**Definition.** A ***regular local ring*** is a local ring $(R, \mathfrak{m}, k)$ such that

$$\dim(R) = V(R).$$

It is clear that every field is a regular local ring of dimension 0, and every DVR is a regular local ring of dimension 1.

**Example C-5.172.** Regular local rings arise in algebraic geometry: varieties are viewed as algebraic analogs of differentiable manifolds, and geometric definitions are adapted to an algebraic setting. Recall some definitions from Section C-5.5. Let $I$ be an ideal in $k[X] = k[x_1, \ldots, x_n]$, where $k$ is an algebraically closed field; we regard each $f(X) \in k[X]$ as a $k$-valued function. Define the ***affine variety*** of $I$ by

$$V = \operatorname{Var}(I) = \{a \in k^n : f(a) = 0 \text{ for all } f \in I\},$$

and define its *coordinate ring* $k[V]$ by

$$k[V] = \{f|V : f(X) \in k[X]\}.$$

Now $k[V] \cong k[X]/\operatorname{Id}(V)$, where $\operatorname{Id}(V) = \{f(X) \in k[X] : f(v) = 0 \text{ for all } v \in V\}$ is an ideal in $k[X]$. Affine varieties $V$ are usually assumed to be *irreducible*; that is, $\operatorname{Id}(V)$ is a prime ideal (see Proposition B-6.42 in Part 1), so that the coordinate ring $k[V]$ is a domain.

If $a$ is a point in an affine variety $V$, then $\mathfrak{m}_a = \{f + \operatorname{Id}(V) \in k[V] : f(a) = 0\}$ is a maximal ideal in $k[V]$. Define $\mathcal{O}_a$ to be the localization

$$\mathcal{O}_a = k[V]_{\mathfrak{m}_a}.$$

Identify the maximal ideal $\mathfrak{m}_a$ with its localization in $k[V]_{\mathfrak{m}_a}$, and call $(\mathcal{O}_a, \mathfrak{m}_a, k)$ the ***local ring of*** $V$ ***at*** $a$. Now $V = \operatorname{Var}(I)$, where $I = \big(f_1(X), \ldots, f_t(X)\big)$. Each $f_i(X)$ has partial derivatives $\partial f_i/\partial x_j$ (which are defined formally, as derivatives of polynomials of a single variable are defined). The ***tangent space*** $T_a$ of $V$ at $a$ can be defined, and there is an isomorphism of the dual space $T_a^* \cong \mathfrak{m}_a/\mathfrak{m}_a^2$ (as vector

spaces over $k$); thus, $\dim_k(\mathfrak{m}_a/\mathfrak{m}_a^2) = \dim_k(T_a)$. The ***Jacobian*** of $f_1(X), \dots, f_t(X)$ is the $t \times n$ matrix over $k[X]$,

$$\operatorname{Jac}(f_1, \dots, f_t) = \big[\partial f_i / \partial x_j\big].$$

If $a \in V$, evaluating each entry in $\operatorname{Jac}(f_1, \dots, f_t)$ at $a$ gives a matrix $\operatorname{Jac}(f_1, \dots, f_t)_a$ over $k$. We say that $a$ is ***nonsingular*** if $\operatorname{rank}\big(\operatorname{Jac}(f_1, \dots, f_n)_a\big) = n - \dim(V)$. There are several equivalent ways of expressing nonsingularity of $a$, and all of them say that $a$ is nonsingular if and only if $(\mathcal{O}_a, \mathfrak{m}_a, k)$ is a regular local ring (Hartshorne [**95**], §1.5). (There is a notion of *projective variety*, which is essentially a compactification of an affine variety, and this discussion can be adapted to them as well.) ◀

**Proposition C-5.173.** *Let $(R, \mathfrak{m}, k)$ be a local ring. If $\mathfrak{m}$ can be generated by an $R$-sequence $x_1, \dots, x_d$, then $R$ is a regular local ring and*

$$d = \dim(R) = V(R).$$

**Remark.** We will soon prove the converse: in a regular local ring, the maximal ideal can be generated by an $R$-sequence. ◀

**Proof.** Consider the inequalities

$$d \le \operatorname{ht}(\mathfrak{m}) \le V(R) \le d.$$

The first inequality holds by Lemma C-5.161, the second by Corollary C-5.171, and the third by Proposition C-5.169. It follows that all the inequalities are, in fact, equalities, and the proposition follows because $\dim(R) = \operatorname{ht}(\mathfrak{m})$. •

**Example C-5.174.** Let $k$ be a field, and let $R = k[[x_1, \dots, x_r]]$ be the ring of ***formal power series*** in $r$ variables $x_1, \dots, x_r$. Recall that an element $f \in R$ is a sequence

$$f = (f_0, f_1, f_2, \dots, f_n, \dots),$$

where $f_n$ is a homogeneous polynomial of total degree $n$ in $k[x_1, \dots, x_r]$, and that multiplication is defined by

$$(f_0, f_1, f_2, \dots)(g_0, g_1, g_2, \dots) = (h_0, h_1, h_2, \dots),$$

where $h_n = \sum_{i+j=n} f_i g_j$. We claim that $R$ is a local ring with maximal ideal $\mathfrak{m} = (x_1, \dots, x_r)$ and residue field $k$. First, $R/\mathfrak{m} \cong k$, so that $\mathfrak{m}$ is a maximal ideal. To see that $\mathfrak{m}$ is the unique maximal ideal, it suffices to prove that if $f \in R$ and $f \notin \mathfrak{m}$, then $f$ is a unit. Now $f \notin \mathfrak{m}$ if and only if $f_0 \neq 0$, and we now show that $f$ is a unit if and only if $f_0 \neq 0$. If $fg = 1$, then $f_0 g_0 = 1$, and $f_0 \neq 0$; conversely, if $f_0 \neq 0$, we can solve $(f_0, f_1, f_2, \dots)(g_0, g_1, g_2, \dots) = 1$ recursively for $g_n$, and $fg = 1$, where $g = (g_0, g_1, g_2, \dots)$.

Exercise C-5.69 on page 510 shows that the ring $R$ is noetherian. But the quotient ring $R/(x_1, \dots, x_{i-1})$ is a domain, because it is isomorphic to $k[[x_i, \dots, x_r]]$, and so $x_i$ is a regular element on it. Hence, Proposition C-5.173 shows that $R = k[[x_1, \dots, x_r]]$ is a regular local ring, for $x_1, \dots, x_r$ is an $R$-sequence. ◀

The next lemma prepares us for induction.

**Lemma C-5.175.** *Let $(R, \mathfrak{m}, k)$ be a regular local ring. If $x \in \mathfrak{m} - \mathfrak{m}^2$, then $R/(x)$ is regular and* $\dim(R/(x)) = \dim(R) - 1$.

**Proof.** Since $R$ is regular, we have $\dim(R) = V(R)$. Let us note at the outset that $\dim(R) = \operatorname{ht}(\mathfrak{m})$. We must show that $\operatorname{ht}(\mathfrak{m}^*) = V(R/(x)) = \dim_k(\mathfrak{m}^*/\mathfrak{m}^{*2})$, where $\mathfrak{m}^* = \mathfrak{m}/(x)$. By Corollary C-5.160, $\operatorname{ht}(\mathfrak{m}) \leq \operatorname{ht}(\mathfrak{m}^*) + 1$. Hence,

$$\operatorname{ht}(\mathfrak{m}) - 1 \leq \operatorname{ht}(\mathfrak{m}^*) \leq V(R/(x)) = V(R) - 1 = \operatorname{ht}(\mathfrak{m}) - 1.$$

The next to last equation is Proposition C-5.170; the last equation holds because $R$ is regular. Therefore, $\dim(R/(x)) = \operatorname{ht}(\mathfrak{m}^*) = V(\mathfrak{m}^*)$, and so $R/(x)$ is regular with $\dim(R/(x)) = \dim(R) - 1$. •

We are now going to prove that regular local rings are domains, and we will then use this to prove the converse of Proposition C-5.173.

**Proposition C-5.176.** *Every regular local ring $(R, \mathfrak{m}, k)$ is a domain.*

**Proof.** The proof is by induction on $d = \dim(R)$. If $d = 0$, then $R$ is a field, by Exercise C-5.71 on page 524. If $d > 0$, let $\mathfrak{p}_1, \ldots, \mathfrak{p}_s$ be the minimal prime ideals in $R$ (there are only finitely many, by Corollary C-5.154). If $\mathfrak{m} - \mathfrak{m}^2 \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_s$, then Lemma C-5.163 would give $\mathfrak{m} \subseteq \mathfrak{p}_i$, which cannot occur because $d = \operatorname{ht}(\mathfrak{m}) > 0$. Therefore, there is $x \in \mathfrak{m} - \mathfrak{m}^2$ with $x \notin \mathfrak{p}_i$ for all $i$. By Lemma C-5.175, $R/(x)$ is regular of dimension $d - 1$. Now $R/(x)$ is a domain, by the inductive hypothesis, and so $(x)$ is a prime ideal. It follows that $(x)$ contains a minimal prime ideal; say, $\mathfrak{p}_i \subseteq (x)$.

If $\mathfrak{p}_i = (0)$, then $(0)$ is a prime ideal and $R$ is a domain. Hence, we may assume that $\mathfrak{p}_i \neq (0)$. For each nonzero $y \in \mathfrak{p}_i$, there exists $r \in R$ with $y = rx$. Since $x \notin \mathfrak{p}_i$, we have $r \in \mathfrak{p}_i$, so that $y \in x\mathfrak{p}_i$. Thus, $\mathfrak{p}_i \subseteq x\mathfrak{p}_i \subseteq \mathfrak{m}\mathfrak{p}_i$. As the reverse inclusion $\mathfrak{m}\mathfrak{p}_i \subseteq \mathfrak{p}_i$ is always true, we have $\mathfrak{p}_i = \mathfrak{m}\mathfrak{p}_i$. Nakayama's Lemma now applies, giving $\mathfrak{p}_i = (0)$, a contradiction. •

**Proposition C-5.177.** *A local ring $(R, \mathfrak{m}, k)$ is regular if and only if $\mathfrak{m}$ is generated by an $R$-sequence $x_1, \ldots, x_d$. Moreover, in this case,*

$$d = V(R).$$

**Proof.** We have already proved sufficiency, in Proposition C-5.173. If $R$ is regular, we prove the result by induction on $d \geq 1$, where $d = \dim(R)$. The base step holds, for $R$ is a domain and so $x$ is a regular element; that is, $x$ is not a zero-divisor. For the inductive step, the ring $R/(x)$ is regular of dimension $d - 1$, by Lemma C-5.175. Therefore, its maximal ideal is generated by an $(R/(x))$-sequence $x_1^*, \ldots, x_{d-1}^*$. In view of Proposition C-5.170, a minimal generating set for $\mathfrak{m}$ is $x, x_1, \ldots, x_{d-1}$. Finally, this is an $R$-sequence, by Exercise C-5.72(i) on page 524, because $x$ is not a zero-divisor. •

We are now going to characterize regular local rings by their global dimension.

**Lemma C-5.178.** *Let $(R, \mathfrak{m}, k)$ be a local ring, and let $A$ be an $R$-module with $\mathrm{pd}(A) = n$. If $x \in \mathfrak{m}$ and multiplication $\mu_x \colon A \to A$, given by $a \mapsto xa$, is injective, then $\mathrm{pd}(A/xA) = n + 1$.*

**Proof.** By hypothesis, there is an exact sequence

$$0 \to A \xrightarrow{\mu_x} A \to A/xA \to 0,$$

where $\mu_x \colon a \mapsto xa$. By Lemma C-5.144, we have $\mathrm{pd}(A/xA) \leq n + 1$.

Consider the portion of the long exact sequence arising from tensoring by $k$:

$$0 = \mathrm{Tor}^R_{n+1}(A, k) \to \mathrm{Tor}^R_{n+1}(A/xA, k) \xrightarrow{\partial} \mathrm{Tor}^R_n(A, k) \xrightarrow{(\mu_x)_*} \mathrm{Tor}^R_n(A, k).$$

The first term is $\{0\}$, for $\mathrm{pd}(A) \leq n$ if and only if $\mathrm{Tor}^R_{n+1}(A, k) = \{0\}$, by Lemma C-5.166. The induced map $(\mu_x)_*$ is multiplication by $x$. But if $\mu'_x \colon k \to k$ is multiplication by $x$, then $x \in \mathfrak{m}$ implies $\mu'_x = 0$; therefore, $(\mu_x)_* = (\mu'_x)_* = 0$. Exactness now implies that $\partial \colon \mathrm{Tor}^R_{n+1}(A/xA, k) \to \mathrm{Tor}^R_n(A, k)$ is an isomorphism. Since $\mathrm{pd}(A) = n$, we have $\mathrm{Tor}^R_n(A, k) \neq \{0\}$, so that $\mathrm{Tor}^R_{n+1}(A/xA, k) \neq \{0\}$. Therefore, $\mathrm{pd}(A/xA) \geq n + 1$, as desired. •

**Proposition C-5.179.** *If $(R, \mathfrak{m}, k)$ is a regular local ring, then*

$$\mathrm{D}(R) = V(R) = \dim(R).$$

**Proof.** Since $R$ is regular, Proposition C-5.177 says that $\mathfrak{m}$ can be generated by an $R$-sequence $x_1, \ldots, x_d$. Applying Lemma C-5.178 to the modules

$$R,\ R/(x_1),\ R/(x_1, x_2), \ldots, R/(x_1, \ldots, x_d) = R/\mathfrak{m} = k,$$

we see that $\mathrm{pd}(k) = d$. By Proposition C-5.173, $d = V(R) = \dim(R)$. On the other hand, Theorem C-5.168(ii) gives $d = \mathrm{pd}(k) = \mathrm{D}(R)$. •

The converse of Proposition C-5.179: a noetherian local ring of finite global dimension is regular, is more difficult to prove. The following proof is essentially that in Lam, [**135**], Chapter 2, §5F.

**Lemma C-5.180.** *Let $(R, \mathfrak{m}, k)$ be a local ring of finite global dimension. If $V(R) \leq \mathrm{D}(R)$ and $\mathrm{D}(R) \leq d$, where $d$ is the length of a longest $R$-sequence in $\mathfrak{m}$, then $R$ is a regular local ring.*

**Proof.** By Corollary C-5.171, $\dim(R) \leq V(R)$. By hypothesis, $V(R) \leq \mathrm{D}(R) \leq d$, while Lemma C-5.161 gives $d \leq \mathrm{ht}(\mathfrak{m}) = \dim(R)$. Therefore, $\dim(R) = V(R)$, and so $R$ is a regular local ring. •

In proving that $\mathrm{D}(R)$ finite implies $R$ regular, we cannot assume that $R$ is a domain (though this will turn out to be true); hence, we must deal with zero-divisors. Recall that $\mathrm{Zer}(R)$ is the set of all zero-divisors in $R$.

**Proposition C-5.181.** *Let $(R, \mathfrak{m}, k)$ be a local ring.*

(i) *If $\mathfrak{m} - \mathfrak{m}^2$ consists of zero-divisors, then there is a nonzero $a \in R$ with $a\mathfrak{m} = (0)$.*

(ii) *If $0 < \mathrm{D}(R) = n < \infty$, then there exists a nonzero-divisor $x \in \mathfrak{m} - \mathfrak{m}^2$.*

**Proof.**

(i) By Proposition C-5.164, there are prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ with

$$\mathfrak{m} - \mathfrak{m}^2 \subseteq \mathrm{Zer}(R) \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n.$$

If we can show that

$$\mathfrak{m} \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n, \tag{1}$$

then Prime Avoidance, Lemma C-5.163, gives $\mathfrak{m} \subseteq \mathfrak{p}_i$ for some $i$. But $\mathfrak{p}_i = \mathrm{ann}(a)$ for some $a \in \mathfrak{m}$, so that $a\mathfrak{m} = (0)$, as desired.

To verify (1), it suffices to prove $\mathfrak{m}^2 \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$. Now $\mathfrak{m} \neq \mathfrak{m}^2$, by Nakayama's Lemma (we may assume that $\mathfrak{m} \neq (0)$, for the result is trivially true otherwise), and so there exists $x \in \mathfrak{m} - \mathfrak{m}^2 \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$. Let $y \in \mathfrak{m}^2$. For every integer $s \geq 1$, we have $x + y^s \in \mathfrak{m} - \mathfrak{m}^2 \subseteq \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$; that is, $x + y^s \in \mathfrak{p}_j$ for some $j = j(s)$. By the Pigeonhole Principle, there is an integer $j$ and integers $s < t$ with $x + y^s, x + y^t \in \mathfrak{p}_j$. Subtracting, $y^s(1 - y^{t-s}) \in \mathfrak{p}_j$. But $1 - y^{t-s}$ is a unit (if $u \in \mathfrak{m}$, then $1 - u$ is a unit; otherwise, $(1 - u)$ is a proper ideal, $(1 - u) \subseteq \mathfrak{m}$, $1 - u \in \mathfrak{m}$, and $1 \in \mathfrak{m}$). Since $\mathfrak{p}_j$ is a prime ideal, $y \in \mathfrak{p}_j$.

(ii) **(Griffith)** In light of (i), it suffices to show that there is no nonzero $a \in \mathfrak{m}$ with $a\mathfrak{m} = (0)$. If, on the contrary, such an $a$ exists and $\mu\colon R \to R$ is given by $\mu\colon r \mapsto ar$, then $\mathfrak{m} \subseteq \ker \mu$. This inclusion cannot be strict; if $b \in \ker \mu$, then $ab = 0$, but if $b \notin \mathfrak{m}$, then $b$ is a unit (for $Rb$ is not contained in the maximal ideal), and so $ab \neq 0$. Hence, $\mathfrak{m} = \ker \mu$. Thus, $k = R/\mathfrak{m} = R/\ker \mu \cong \mathrm{im}\, \mu = Ra$; that is, $k \cong Ra$. Consider the exact sequence $0 \to Ra \to R \to R/Ra \to 0$; by Exercise C-5.63 on page 496, either $\mathrm{pd}(R/Ra)\,\mathrm{pd}(Ra) + 1 = \mathrm{pd}(k) + 1$ or $\mathrm{pd}(R/Ra) = 0$. In the first case, $\mathrm{pd}(R/Ra) = \mathrm{pd}(k) + 1 > \mathrm{pd}(k)$, contradicting Theorem C-5.168(ii) (which says that $\mathrm{pd}(k) = \mathrm{D}(R)$). In the second case, $0 = \mathrm{pd}(Ra) = \mathrm{pd}(k)$, contradicting $\mathrm{pd}(k) = \mathrm{D}(R) > 0$. •

**Theorem C-5.182 (Serre–Auslander–Buchsbaum).** *A local ring $(R, \mathfrak{m}, k)$ is regular if and only if $\mathrm{D}(R)$ is finite; in fact,*

$$\mathrm{D}(R) = V(R) = \dim(R).$$

**Proof.** Necessity is Proposition C-5.179. We prove the converse by induction on $\mathrm{D}(R) = n \geq 0$. If $n = 0$, then $R$ is semisimple. Since $R$ is commutative, it is the direct product of finitely many fields; since $R$ is local, it is a field, and hence it is regular.

If $n \geq 1$, then Proposition C-5.181(ii) says that $\mathfrak{m} - \mathfrak{m}^2$ contains a nonzero-divisor $x$. Now $(R^*, \mathfrak{m}^*, k)$ is a local ring, where $R^* = R/(x)$ and $\mathfrak{m}^* = \mathfrak{m}/(x)$. Since $x$ is not a zero-divisor, it is regular on $\mathfrak{m}$; since $\mathrm{pd}_R(\mathfrak{m}) < \infty$, Corollary C-5.151 gives $\mathrm{pd}_{R^*}(\mathfrak{m}/x\mathfrak{m}) < \infty$.

Consider a short exact sequence of $R$-modules

$$0 \to k \xrightarrow{\alpha} B \to C \to 0 \tag{2}$$

in which $\alpha(1) = e$, where $e \in B - \mathfrak{m}B$. Now the coset $e + \mathfrak{m}B$ is part of a basis of the $k$-vector space $B/\mathfrak{m}B$, and so there is a $k$-map $\beta\colon B/\mathfrak{m}B \to k$ with $\beta(e + \mathfrak{m}B) = 1$. We can use the composite $\pi\colon B \xrightarrow{\text{nat}} B/\mathfrak{m}B \xrightarrow{\beta} k$ to show that the exact sequence (2) splits, for $\pi\alpha = 1_k$. In particular, this applies when $B = \mathfrak{m}/x\mathfrak{m}$ and $k$ is the cyclic submodule generated by $e + x\mathfrak{m}$. Thus, $k$ is a direct summand of $\mathfrak{m}/x\mathfrak{m}$. It follows that $\operatorname{pd}_{R^*}(k) < \infty$, so that Proposition C-5.152 gives $\operatorname{pd}_{R^*}(k) = \operatorname{pd}_R(k) - 1$. Theorem C-5.168 now gives

$$\operatorname{D}(R^*) = \operatorname{pd}_{R^*}(k) = n - 1.$$

By induction, $R^*$ is a regular local ring and $\dim(R^*) = n - 1$. Hence, there is a prime chain of length $n - 1$ in $R^*$,

$$\mathfrak{p}_0^* \supsetneq \mathfrak{p}_1^* \supsetneq \cdots \supsetneq \mathfrak{p}_{n-1}^* = (0)$$

(we have $\mathfrak{p}_{n-1}^* = (0)$ because $R^*$ is a domain (being regular) so that $(0)$ is a prime ideal). Taking inverse images gives a prime chain in $R$:

$$\mathfrak{p}_0 \supsetneq \mathfrak{p}_1 \supsetneq \cdots \supsetneq \mathfrak{p}_{n-1} = (x).$$

Were $(x)$ a minimal prime ideal, then every element in it would be nilpotent, by Proposition C-5.24. Since $x$ is not a zero-divisor, it is not nilpotent, and so there is a prime ideal $\mathfrak{q} \subsetneq (x)$. Hence, $\dim(R) \geq n$.

Since $x \in \mathfrak{m} - \mathfrak{m}^2$, Proposition C-5.170 says that $V(R) = 1 + V(R^*) = 1 + \dim(R^*) = n$. Therefore,

$$n = V(R) \geq \dim(R) \geq n$$

(the inequality $V(R) \geq \dim(R)$ always being true), and $\dim(R) = V(R)$; that is, $R$ is a regular local ring of dimension $n$. •

**Corollary C-5.183.** *If $S \subseteq R$ is multiplicative, where $(R, \mathfrak{m}, k)$ is a regular local ring, then $S^{-1}R$ is also a regular local ring. In particular, if $\mathfrak{p}$ is a prime ideal in $R$, then $R_{\mathfrak{p}}$ is regular.*

**Proof.** Theorem C-5.143(ii) says that $\operatorname{D}(R) = \operatorname{wD}(R)$ in this case. But Proposition C-5.142 says that $\operatorname{wD}(S^{-1}R) \leq \operatorname{wD}(R)$. Therefore,

$$\operatorname{D}(S^{-1}R) = \operatorname{wD}(S^{-1}R) \leq \operatorname{wD}(R) = \operatorname{D}(R) < \infty.$$

It follows from Corollary C-5.20 that $S^{-1}R$ is a local ring; therefore, $S^{-1}R$ is regular, by Theorem C-5.182. The second statement follows: if $S = R - \mathfrak{p}$, then $R_{\mathfrak{p}}$ is a local ring, and so the Serre–Auslander–Buchsbaum Theorem says that $R_{\mathfrak{p}}$ is regular. •

There are several proofs that regular local rings are unique factorization domains; most use the notion of *depth* that was used in the original proof of Auslander and Buchsbaum. If $M$ is a finitely generated $R$-module, we generalize the notion of $R$-sequence: the ***depth*** of $M$ is the maximal length of a sequence $r_1, \ldots, r_n$ in $R$ such that $x_1$ is regular on $M$, $x_2$ is regular on $M/(x_1)M$, $\ldots$, $x_n$ is regular on

$M/(x_1, \ldots, x_{n-1})M$. The depth of $M$ was originally called its *codimension* because of the following result of Auslander and Buchsbaum.

**Theorem C-5.184 (Codimension Theorem).** *Let $(R, \mathfrak{m}, k)$ be a local ring, and let $M$ be a finitely generated $R$-module with $\operatorname{pd}(M) < \infty$. Then*

$$\operatorname{pd}(M) + \operatorname{depth}(M) = \operatorname{depth}(R).$$

*In particular, if $R$ is regular, then $\operatorname{depth}(R) = \operatorname{D}(R)$ and*

$$\operatorname{pd}(M) + \operatorname{depth}(M) = \operatorname{D}(R).$$

**Proof.** Eisenbud [**60**], p. 479. •

Nagata [**164**] proved, using the Serre–Auslander–Buchsbaum Theorem, that if one knew that every regular local ring $R$ with $\operatorname{D}(R) = 3$ is a UFD, then it would follow that every regular local ring is a UFD.

Here is a sketch of the original proof of Auslander–Buchsbaum that every regular local ring $R$ is a unique factorization domain. They began with a standard result of commutative algebra.

**Lemma C-5.185.** *If $R$ is a noetherian domain, then $R$ is a* UFD *if and only if every prime ideal of height* 1 *is principal.*

**Proof.** Let $R$ be a UFD, and let $\mathfrak{p}$ be a prime ideal of height 1. If $a \in \mathfrak{p}$ is nonzero, then $a = \pi_1^{e_1} \cdots \pi_n^{e_n}$, where the $\pi_i$ are irreducible and $e_i \geq 1$. Since $\mathfrak{p}$ is prime, one of the factors, say, $\pi_j \in \mathfrak{p}$. Of course, $R\pi_j \subseteq \mathfrak{p}$. But $R\pi_j$ is a prime ideal, by Proposition A-3.124 in Part 1, so that $R\pi_j = \mathfrak{p}$, because $\operatorname{ht}(\mathfrak{p}) = 1$.

Conversely, since $R$ is noetherian, Lemma A-3.125 in Part 1 shows that every nonzero nonunit in $R$ is a product of irreducibles, and so Proposition A-3.124 in Part 1 says that it suffices to prove, for every irreducible $\pi \in R$, that $R\pi$ is a prime ideal. Choose a prime ideal $\mathfrak{p}$ that is minimal over $R\pi$. By the Principal Ideal Theorem, Theorem C-5.157, we have $\operatorname{ht}(\mathfrak{p}) = 1$, and so the hypothesis gives $\mathfrak{p} = Ra$ for some $a \in R$. Therefore, $\pi = ua$ for some $u \in R$. Since $\pi$ is irreducible, we must have $u$ a unit, and so $R\pi = Ra = \mathfrak{p}$, as desired. •

In light of Nagata's result, Auslander and Buchsbaum could focus on the three-dimensional case.

**Lemma C-5.186.** *If $(R, \mathfrak{m}, k)$ is a local ring with $\operatorname{D}(R) = 3$ and $\mathfrak{p} \neq \mathfrak{m}$ is a prime ideal in $R$, then*

$$\operatorname{pd}(\mathfrak{p}) \leq 1.$$

**Proof.** By hypothesis, there exists $x \in \mathfrak{m} - \mathfrak{p}$. Now $x$ is regular on $R/\mathfrak{p}$: if $x(r+\mathfrak{p}) = \mathfrak{p}$, then $xr \in \mathfrak{p}$ and $r \in \mathfrak{p}$, for $x \notin \mathfrak{p}$ and $\mathfrak{p}$ is prime. Therefore, $\operatorname{depth}(R/\mathfrak{p}) \geq 1$ and so $\operatorname{pd}(R/\mathfrak{p}) \leq 2$, by the Codimension Theorem. But there is an exact sequence $0 \to \mathfrak{p} \to R \to R/\mathfrak{p} \to 0$, which shows that $\operatorname{pd}(\mathfrak{p}) \leq 1$. •

Auslander and Buchsbaum showed that $\operatorname{pd}(\mathfrak{p}) = 1$ gives a contradiction, so that $\operatorname{pd}(\mathfrak{p}) = 0$; that is, $\mathfrak{p}$ is a projective, hence free, $R$-module. But any ideal in a domain $R$ that is free as an $R$-module must be principal. This completed the proof.

A second proof, not using Nagata's difficult proof, is based on the following criterion.

**Proposition C-5.187.** *If $R$ is a noetherian domain for which every finitely generated $R$-module has* FFR*, then $R$ is a unique factorization domain.*

**Proof.** This is Theorem 184 in Kaplansky [**118**]. The proof uses a criterion for a domain to be a unique factorization domain (his Theorem 179), which involves showing that if a commutative ring $R$ has the property that every finitely generated $R$-module has FFR, then so does $R[x]$. •

Unique factorization for regular local rings follows easily. If $\mathrm{D}(R) = n$, then every finitely generated $R$-module $M$ has a projective resolution $0 \to P_n \to \cdots \to P_0 \to M \to 0$ in which every $P_i$ is finitely generated (Lemma C-5.41). But (finitely generated) projective $R$-modules are free (Proposition C-5.26), and so every finitely generated $R$-module has FFR.

We now give a complete (third) proof that every regular local ring is a UFD; we begin with two elementary lemmas.

**Lemma C-5.188.** *Let $R$ be a noetherian domain, let $x \in R$ be a nonzero element with $Rx$ a prime ideal, and denote $S^{-1}R$ by $R_x$, where $S = \{x^n : n \geq 0\}$. Then $R$ is a* UFD *if and only if $R_x$ is a* UFD *for all such $x$.*

**Proof.** We leave necessity as a routine exercise for the reader. For sufficiency, assume that $R_x$ is a UFD. Let $\mathfrak{p}$ be a prime ideal in $R$ of height 1. If $x \in \mathfrak{p}$, then $Rx \subseteq \mathfrak{p}$ and, since $\mathrm{ht}(\mathfrak{p}) = 1$, we have $Rx = \mathfrak{p}$ (for $Rx$ is prime), and so $\mathfrak{p}$ is principal in this case. We may now assume that $x \notin \mathfrak{p}$; that is, $S \cap \mathfrak{p} = \varnothing$. It follows that $\mathfrak{p}R_x$ is a prime ideal in $R_x$ of height 1, and so it is principal, by hypothesis. There is some $a \in \mathfrak{p}$ and $n \geq 0$ with $\mathfrak{p}R_x = R_x(a/x^n) = R_x a$, for $x$ is a unit in $R_x$. We may assume that $a \notin Rx$. If $a = a_1 x$ and $a_1 \notin Rx$, then replace $a$ by $a_1$, for $R_x a = R_x a_1$. If $a_1 = a_2 x$ and $a_2 \notin Rx$, then replace $a_1$ by $a_2$, for $R_x a_1 = R_x a_2$. If this process does not stop, there are equations $a_m = a_{m+1}x$ for all $m \geq 1$, which give rise to an ascending sequence $Ra_1 \subseteq Ra_2 \subseteq \cdots$. Since $R$ is noetherian, $Ra_m = Ra_{m+1}$ for some $m$. Hence, $a_{m+1} = ra_m$ for some $r \in R$, and $a_m = a_{m+1}x = ra_m x$. Since $R$ is a domain, $1 = rx$; thus, $x$ is a unit, contradicting $Rx$ being a prime (hence, proper) ideal. Clearly, $Ra \subseteq \mathfrak{p}$; we claim that $Ra = \mathfrak{p}$. If $b \in \mathfrak{p}$, then $b = (r/x^m)a$ in $R_x$, where $r \in R$ and $m \geq 0$. Hence, $x^m b = ra$ in $R$. Choose $m$ minimal. If $m > 0$, then $ra = x^m b \in Rx$; since $Rx$ is prime, either $r \in Rx$ or $a \in Rx$. But $a \notin Rx$ since $S \cap \mathfrak{p} = \varnothing$, so that $r = xr'$. As $R$ is a domain, this gives $r'a = x^{m-1}b$, contradicting the minimality of $m$. We conclude that $m = 0$, and so $\mathfrak{p} = Ra$ is principal. Lemma C-5.185 now shows that $R$ is a UFD. •

The following lemma is true when the localizing ideal is prime; however, we will use it only in the case the ideal is maximal.

**Lemma C-5.189.** *Let $R$ be a domain, and let $I$ be a nonzero projective ideal in $R$. If $\mathfrak{m}$ is a maximal ideal in $R$, then*

$$I_{\mathfrak{m}} \cong R_{\mathfrak{m}}.$$

**Proof.** Since $I$ is a projective $R$-module, $I_{\mathfrak{m}}$ is a projective $R_{\mathfrak{m}}$-module. As $R_{\mathfrak{m}}$ is a local ring, however, $I_{\mathfrak{m}}$ is a free $R_{\mathfrak{m}}$-module. But $I_{\mathfrak{m}}$ is an ideal in a domain $R_{\mathfrak{m}}$, and so it must be principal; that is, $I_{\mathfrak{m}} \cong R_{\mathfrak{m}}$. •

**Theorem C-5.190 (Auslander–Buchsbaum).** *Every regular local ring $(R, \mathfrak{m}, k)$ is a unique factorization domain.*

**Proof (Kaplansky).** The proof is by induction on the Krull dimension $\dim(R)$, the cases $n = 0$ ($R$ is a field) and $n = 1$ ($R$ is a DVR) being obvious (Exercise C-5.71 on page 524). For the inductive step, choose $x \in \mathfrak{m} - \mathfrak{m}^2$. By Lemma C-5.175, $R/Rx$ is a regular local ring with $\dim(R/Rx) < \dim(R)$; by Proposition C-5.176, $R/Rx$ is a domain, and so $Rx$ is a prime ideal. It suffices, by Lemma C-5.188, to prove that $R_x$ is a UFD (where $R_x = S^{-1}R$ for $S = \{x^n : n \geq 0\}$). Let $\mathfrak{P}$ be a prime ideal of height 1 in $R_x$; we must show that $\mathfrak{P}$ is principal. Define $\mathfrak{p} = \mathfrak{P} \cap R$ (since $R$ is a domain, $R_x \subseteq \mathrm{Frac}(R)$, so that the intersection makes sense). Since $R$ is a regular local ring, $\mathrm{D}(R) < \infty$, and so the $R$-module $\mathfrak{p}$ has a free resolution of finite length:

$$0 \to F_n \to F_{n-1} \to \cdots \to F_0 \to \mathfrak{p} \to 0.$$

Tensoring by $R_x$, which is a flat $R$-module (Theorem C-5.30), gives a free $R_x$-resolution of $\mathfrak{P}$ (for $\mathfrak{P} = R_x\mathfrak{p}$):

$$0 \to F'_n \to F'_{n-1} \to \cdots \to F'_0 \to \mathfrak{P} \to 0, \tag{3}$$

where $F'_i = R_x \otimes_R F_i$.

We claim that $\mathfrak{P}$ is projective. By Proposition C-5.165, it suffices to show that every localization $\mathfrak{P}_{\mathfrak{M}}$ is projective, where $\mathfrak{M}$ is a maximal ideal in $R_x$. Now $(R_x)_{\mathfrak{M}}$ is a localization of $R$, and so it is a regular local ring, by Corollary C-5.186; its dimension is smaller than $D(R)$, and so it is a UFD, by induction. Now $\mathfrak{P}_{\mathfrak{M}}$, being a height 1 prime ideal in the UFD $(R_x)_{\mathfrak{M}}$, is principal. But principal ideals in a domain are free, hence projective, and so $\mathfrak{P}_{\mathfrak{M}}$ is projective. Therefore, $\mathfrak{P}$ is projective.

The exact sequence (3) "factors" into split short exact sequences. Since $\mathfrak{P}$ is projective, we have $F'_0 \cong \mathfrak{P} \oplus \Omega_0$, where $\Omega_0 = \ker(F'_0 \to \mathfrak{P})$. Thus, $\Omega_0$ is projective, being a summand of a free module, and so $F'_1 \cong \Omega_1 \oplus \Omega_0$, where $\Omega_1 = \ker(F'_1 \to F'_0)$. More generally, $F'_i \cong \Omega_i \oplus \Omega_{i-1}$ for all $i \geq 1$. Hence,

$$F'_0 \oplus F'_1 \oplus \cdots \oplus F'_n \cong (\mathfrak{P} \oplus \Omega_0) \oplus (\Omega_1 \oplus \Omega_0) \oplus \cdots .$$

Since projective modules over a local ring are free, we see that there are finitely generated free $R_x$-modules $Q$ and $Q'$ with

$$Q \cong \mathfrak{P} \oplus Q'.$$

Recall that $\mathrm{rank}(Q) = \dim_K(K \otimes_{R_x} Q)$, where $K = \mathrm{Frac}(R_x)$; now $\mathrm{rank}(\mathfrak{P}) = 1$ and $\mathrm{rank}(Q') = r$, say, so that $\mathrm{rank}(Q) = r + 1$.

We must still show that $\mathfrak{P}$ is principal. Now

$$\bigwedge^{r+1} Q \cong \bigwedge^{r+1} (\mathfrak{P} \oplus Q').$$

Since $Q$ is free of rank $r+1$, the Binomial Theorem (Theorem B-5.32 in Part 1) gives $\bigwedge^{r+1} Q \cong R_x$. On the other hand, Theorem B-5.35 in Part 1 gives

$$\bigwedge^{r+1}(\mathfrak{P} \oplus Q') \cong \sum_{i=0}^{r+1} \Big( \bigwedge^{i} \mathfrak{P} \otimes_{R_x} \bigwedge^{r+1-i} Q' \Big). \tag{4}$$

We claim that $\bigwedge^{i} \mathfrak{P} = \{0\}$ for all $i > 1$. By Lemma C-5.189, we have $\mathfrak{P}_{\mathfrak{M}} \cong (R_x)_{\mathfrak{M}}$ for every maximal ideal $\mathfrak{M}$ in $R_x$. Now Exercise C-5.25 on page 445 gives

$$\Big( \bigwedge^{i} \mathfrak{P} \Big)_{\mathfrak{M}} \cong \bigwedge^{i} (\mathfrak{P}_{\mathfrak{M}}) \cong \bigwedge^{i} (R_x)_{\mathfrak{M}}$$

for all maximal ideals $\mathfrak{M}$ and all $i$. But $\bigwedge^{i}(R_x)_{\mathfrak{M}} = \{0\}$ for all $i > 1$ (by the Binomial Theorem or by the simpler Corollary B-5.30 in Part 1), so that Proposition C-5.38 gives $\bigwedge^{i} \mathfrak{P} = \{0\}$ for all $i > 1$.

We have just seen that most of the terms in (4) are $\{0\}$; what survives is

$$\bigwedge^{r+1}(\mathfrak{P} \oplus Q') \cong \Big( \bigwedge^{0} \mathfrak{P} \otimes_{R_x} \bigwedge^{r+1} Q' \Big) \oplus \Big( \bigwedge^{1} \mathfrak{P} \otimes_{R_x} \bigwedge^{r} Q' \Big).$$

But $\bigwedge^{r+1} Q' = \{0\}$ and $\bigwedge^{r} Q' \cong R_x$, because $Q'$ is free of rank $r$. Therefore, $\bigwedge^{r+1}(\mathfrak{P} \oplus Q') \cong \mathfrak{P}$. Since $\mathfrak{P} \cong \bigwedge^{r+1}(\mathfrak{P} \oplus Q') \cong \bigwedge^{r+1} Q \cong R_x$, we have $\mathfrak{P} \cong R_x$ principal. Thus, $R_x$, and hence $R$, is a UFD. •

Having studied localization, we turn, briefly, to globalization, merely describing its setting. To a commutative noetherian ring $R$, we have associated a family of local rings $R_{\mathfrak{p}}$, one for each prime ideal $\mathfrak{p}$, and an $R$-module $M$ has localizations $M_{\mathfrak{p}} = R_{\mathfrak{p}} \otimes_R M$. These data are assembled into a ***sheaf*** over $\operatorname{Spec}(R)$, where $\operatorname{Spec}(R)$ is equipped with the Zariski topology. There are two ways to view a sheaf. One way resembles covering spaces in algebraic topology. Equip the disjoint union

$$E(M) = \bigcup_{\mathfrak{p} \in \operatorname{Spec}(R)} M_{\mathfrak{p}}$$

with a topology so that the ***projection*** $\pi\colon E(M) \to \operatorname{Spec}(R)$, defined by $\pi(e) = \mathfrak{p}$ if $e \in M_{\mathfrak{p}}$, is a local homeomorphism. All such sheaves form a category $\mathbf{Sh}(R)$, where a morphism $\varphi\colon E(M) \to E(M')$ is a continuous map with $\varphi | M_{\mathfrak{p}}\colon M_{\mathfrak{p}} \to M'_{\mathfrak{p}}$ an $R_{\mathfrak{p}}$-map for all $\mathfrak{p}$. Now $\mathbf{Sh}(R)$ is an abelian category (Rotman [**187**], p. 309), which has enough injectives ([**187**], p. 314). A ***section*** of $E(M)$ over an open set $U \subseteq \operatorname{Spec}(R)$ is a continuous function $s\colon U \to E(M)$ with $\pi s = 1_U$. The family $\Gamma(U, E(M))$ of all sections over $U$ is an abelian group. Sections give rise to the second way of viewing a sheaf, for $\Gamma(\square, E(M))$ is a presheaf of abelian groups. We call $\Gamma(\operatorname{Spec}(R), E(M))$ the ***global sections*** of $E(M)$. Now global sections $\Gamma(\operatorname{Spec}(R), \square)\colon \mathbf{Sh}(R) \to \mathbf{Ab}$ is a left exact additive functor ([**187**], p. 378), and its derived functors $H^n(\operatorname{Spec}(R), \square)$ are called ***sheaf cohomology***. These cohomology groups are the most important tool in globalizing. We strongly recommend the article by Serre [**198**] for a lucid discussion.

## Exercises

**C-5.70.** If $(R, \mathfrak{m}, k)$ is a noetherian local ring and $B$ is a finitely generated $R$-module, prove that

$$\operatorname{depth}(B) = \min\{i : \operatorname{Ext}^i_R(k, B) \neq \{0\}\}.$$

∗ **C-5.71.** Let $R$ be a regular local ring.

(i) Prove that $R$ is a field if and only if $\dim(R) = 0$.

(ii) Prove that $R$ is a DVR if and only if $\dim(R) = 1$.

∗ **C-5.72.** (i) Let $(R, \mathfrak{m}, k)$ be a noetherian local ring, and let $x \in R$ be a regular element; i.e., $x$ is not a zero-divisor. If $x_1 + (x), \dots, x_s + (x)$ is an $(R/(x))$-sequence, prove that $x, x_1, \dots, x_s$ is an $R$-sequence.

(ii) Let $R$ be a commutative ring. If $x_1, \dots, x_d$ is an $R$-sequence, prove that the cosets $x_2 + (x_1), \dots, x_d + (x_1)$ form an $(R/(x_1))$-sequence.

**C-5.73.** Let $R$ be a noetherian (commutative) ring with Jacobson radical $J = J(R)$. If $B$ is a finitely generated $R$-module, prove that

$$\bigcap_{n\geq 1} J^n B = \{0\}.$$

Conclude that if $(R, \mathfrak{m}, k)$ is a noetherian local ring, then $\bigcap_{n\geq 1} \mathfrak{m}^n B = \{0\}$.

**Hint.** Let $D = \bigcap_{n\geq 1} J^n B$, observe that $JD = D$, and use Nakayama's Lemma.

**C-5.74.** Use the Rees Lemma to prove a weaker version of Proposition C-5.179: if $(R, \mathfrak{m}, k)$ is a regular local ring, then $D(R) \geq V(R) = \dim(R)$.

**Hint.** If $\operatorname{Ext}^d_R(k, R) \neq \{0\}$, then $D(R) > d - 1$; that is, $D(R) \geq d$. Let $\mathfrak{m} = (x_1, \dots, x_d)$, where $x_1, \dots, x_d$ is an $R$-sequence. Then

$$\begin{aligned}\operatorname{Ext}^d_R(k, R) &\cong \operatorname{Ext}^{d-1}_{R/(x_1)}(k, R/(x_1))\\ &\cong \operatorname{Ext}^{d-2}_{R/(x_1,x_2)}(k, R/(x_1, x_2))\\ &\cong \cdots\\ &\cong \operatorname{Ext}^0_k(k, k) \cong \operatorname{Hom}_k(k, k) \cong k \neq \{0\}.\end{aligned}$$

**C-5.75.** If $k$ is a field, prove that the ring of formal power series $k[[x_1, \dots, x_n]]$ is a UFD.

---

**Example C-5.191.** Note: $\mathbf{ComRings}^{\mathrm{op}} \cong$ category of affine schemes.

If $R$ is a commutative ring, then $X = \operatorname{Spec}(R)$ is the set of all of its prime ideals. The ***Zariski topology*** has as *closed sets* those subsets of the form

$$V(S) = \{\mathfrak{p} \in \operatorname{Spec}(R) : S \subseteq \mathfrak{p}\},$$

where $S$ is any subset of $R$. Of course, open sets are complements of closed sets. A base[11] of the Zariski topology turns out to be all $D(s) = X - V(\{s\})$, where $s \in R$ is nonzero. Thus,

$$D(s) = \{\mathfrak{p} \in \operatorname{Spec}(R) : s \notin \mathfrak{p}\}.$$

[11] Recall that a ***base*** of a topology is a family of open subsets $\mathcal{B}$ such that every open set is a union of sets in $\mathcal{B}$.

Exercise C-4.37 on page 377 shows that we can define a presheaf on a space $X$ by giving its values on basic open sets. If $D(t) \subseteq D(s)$, then $t \in \sqrt{Rs}$, by Hilbert's Nullstellensatz, and so $t^n = rs$ for some $r \in R$ and $n \geq 0$. The ***structure sheaf of*** $R$ is the presheaf $\mathcal{O}$ over $X = \operatorname{Spec}(R)$ of commutative rings having sections $\mathcal{O}(D(s)) = s^{-1}R$ and restriction maps $\rho^{D(s)}_{D(t)} \colon s^{-1}R \to t^{-1}R$ defined by $u/s^m \mapsto ur^m/t^{nm}$ (recall that $t^n = rs$). The structure sheaf $\mathcal{O}$ is a sheaf of commutative rings, and the stalk $\mathcal{O}_{\mathfrak{p}}$ is the localization $R_{\mathfrak{p}}$. (See Hartshorne [**95**], p. 71.) ◀

**Example C-5.192.** Serre [**198**] developed the theory of sheaves over spaces $X$ that need not be Hausdorff, enabling him to apply sheaves in algebraic geometry. For example, the structure sheaf $\mathcal{O}$ of a commutative ring $R$ is a sheaf of commutative rings over $X = \operatorname{Spec}(R)$, and $\operatorname{Spec}(R)$ is rarely Hausdorff. Because of the importance of Serre's paper, it has acquired a nickname; it is usually referred to as FAC.

**Definition.** An ***$\mathcal{O}$-Module*** (note the capital M), where $\mathcal{O}$ is a sheaf of commutative rings over a space $X$, is a sheaf $\mathcal{F}$ of abelian groups over $X$ such that

(i) $\mathcal{F}(U)$ is an $\mathcal{O}(U)$-Module for every open $U \subseteq X$,

(ii) if $U \subseteq V$, then $\mathcal{F}(U)$ is also an $\mathcal{O}(V)$-Module, and the restriction $\rho^V_U \colon \mathcal{F}(V) \to \mathcal{F}(U)$ is an $\mathcal{O}(V)$-Module homomorphism.

If $\mathcal{F}$ and $\mathcal{G}$ are $\mathcal{O}$-Modules, then an ***$\mathcal{O}$-morphism*** $\tau \colon \mathcal{F} \to \mathcal{G}$ is a sheaf map such that $\tau_U \colon \mathcal{F}(U) \to \mathcal{G}(U)$ is an $\mathcal{O}(U)$-map for every open set $U$.

For example, if $\mathcal{O}$ is the structure sheaf of a commutative ring $R$, then every $R$-module $M$ gives rise to an $\mathcal{O}$-Module $\widetilde{M}$ over $\operatorname{Spec}(R)$ whose stalk over $\mathfrak{p} \in \operatorname{Spec}(R)$ is $M_{\mathfrak{p}} = R_{\mathfrak{p}} \otimes_R M$.

All $\mathcal{O}$-Modules and $\mathcal{O}$-morphisms form an abelian category ${}_{\mathcal{O}}\mathbf{Mod}$ which has a version of tensor product. If $\mathcal{F}$ and $\mathcal{G}$ are $\mathcal{O}$-Modules, then $U \mapsto \mathcal{F}(U) \otimes_{\mathcal{O}(U)} \mathcal{G}(U)$ is a presheaf, and the ***tensor product*** $\mathcal{F} \otimes_{\mathcal{O}} \mathcal{G}$ is defined to be its sheafification. There is a faithful exact functor ${}_R\mathbf{Mod} \to {}_{\mathcal{O}}\mathbf{Mod}$ with $M \mapsto \widetilde{M}$, and ${}_R\mathbf{Mod}$ is isomorphic to the full subcategory of ${}_{\mathcal{O}}\mathbf{Mod}$ generated by all $\widetilde{M}$ (see Hartshorne [**95**], II §5).

**Definition.** If $\mathcal{O}$ is a sheaf of commutative rings over a space $X$, then an $\mathcal{O}$-Module $\mathcal{F}$ is ***coherent***[12] if there is an exact sequence

$$\mathcal{O}^s \to \mathcal{O}^r \to \mathcal{F} \to 0,$$

where $r, s$ are natural numbers and $\mathcal{O}^r$ is the direct sum of $r$ copies of $\mathcal{O}$. (We remark that $\mathcal{O}^r$ is *not* a projective object in the category of $\mathcal{O}$-Modules.)

If $\mathcal{F}$ is an $\mathcal{O}$-Module over $X$, then an ***$r$-chart*** is an ordered pair $(U, \varphi)$, where $\varphi \colon \mathcal{F}|U \to \mathcal{O}^r|U$ is an $\mathcal{O}$-isomorphism of $\mathcal{O}$-Modules; we call $U$ the ***coordinate neighborhood*** of the chart. An $\mathcal{O}$-Module $\mathcal{F}$ is ***locally free of rank*** $r$ if there is a family $(U_i, \varphi_i)_{i \in I}$ of $r$-charts, called an ***atlas***, whose coordinate neighborhoods

[12] A coherent $\mathcal{F}$-Module is an analog of a finitely presented module, and coherent rings are so called because their finitely generated modules are analogous to coherent $\mathcal{O}$-Modules.

form an open cover of $X$. An ***invertible sheaf***[13] is a locally free $\mathcal{O}$-Module of rank 1.

Let $\mathcal{F}$ be a locally free $\mathcal{O}$-Module over a space $X$, and let $(U_i, \varphi_i)_{i\in I}$ be an atlas. Whenever an intersection $U_{ij} = U_i \cap U_j$ is nonempty, we can define $\mathcal{O}|U_{ij}$-isomorphisms $\varphi_i\colon (\mathcal{F}|U_i)|U_{ij} \to \mathcal{O}^r|U_{ij}$ and $\varphi_j\colon (\mathcal{F}|U_j)|U_{ij} \to \mathcal{O}^r|U_{ij}$ (these isomorphisms are really restrictions of $\varphi_i$ and $\varphi_j$). Now define $\mathcal{O}|U_{ij}$-automorphisms of $\mathcal{O}^r|U_{ij}$

$$g_{ij} = \varphi_i\varphi_j^{-1},$$

called ***transition functions***. Transition functions satisfy the ***cocycle conditions***:

(i) $g_{ij}g_{jk}g_{ki} = 1_{\mathcal{O}^r|U_{ij}}$ for all $i, j \in I$ $\qquad (\varphi_i\varphi_j^{-1}\varphi_j\varphi_k^{-1}\varphi_k\varphi_i^{-1} = 1)$;

(ii) $g_{ii} = 1_{\mathcal{O}^r|U_i}$ for all $i \in I$.

Of course, transition functions depend on the choice of atlas $(U_i, \varphi_i)_{i\in I}$. Consider new transition functions arising from a new atlas $(U_i, \widetilde{\varphi}_i)_{i\in I}$ in which we vary only the $\mathcal{O}|U_i$-isomorphisms, keeping the same coordinate neighborhoods. If we define $h_i$ by $\widetilde{\varphi}_i = h_i\varphi_i^{-1}$, then the new transition functions are

$$\widetilde{g}_{ij} = \widetilde{\varphi}_i\widetilde{\varphi}_j^{-1} = h_i\varphi_i\varphi_j^{-1}h_j^{-1} = h_ig_{ij}h_j^{-1}.$$

Let $(g_{ij})$, $(\widetilde{g}_{ij})$, where $g_{ij}, \widetilde{g}_{ij} \in \mathrm{Aut}(\mathcal{O}^r|U_{ij})$, be two families that may not have arisen as transition functions of a locally free $\mathcal{O}$-Module of rank $r$. Call $(g_{ij})$, $(\widetilde{g}_{ij})$ ***equivalent*** if there are $\mathcal{O}(U_{ij})$-isomorphisms $h_i$ such that

$$\widetilde{g}_{ij} = h_ig_{ij}h_j^{-1}.$$

**Definition.** Locally free $\mathcal{O}$-Modules $\mathcal{F}$ and $\mathcal{G}$ of rank $r$ are ***isomorphic*** if their transition functions $(g_{ij})$ and $(\widetilde{g}_{ij})$ are equivalent.

Given a family $(g_{ij})$, where $g_{ij} \in \mathrm{Aut}(\mathcal{O}^r|U_{ij})$, that satisfies the cocycle conditions, it is not hard to see that there is a unique (up to isomorphism) locally free $\mathcal{O}$-Module $\mathcal{F}$ whose transition functions are the given family. In particular, if $\mathcal{F}$ is the constant sheaf with $\mathcal{F}(U)$ of rank $r$, then there is an open cover $\mathcal{U}$ giving transition functions $g_{ij} = h_ih_j^{-1}$.

A locally free $\mathcal{O}$-Module of rank $r$ is almost classified by an equivalence class of cocycles $(g_{ij})$; we must still investigate transition functions that arise from an atlas having different families of coordinate neighborhoods. It turns out that transition functions are elements of a certain cohomology *set* of a sheaf with coefficients in the general linear group $\mathrm{GL}(r, k)$ (cohomology need not be a group when coefficients lie in a nonabelian group). ◀

[13] If $\mathcal{F}$ is an invertible sheaf, then there exists an (invertible) sheaf $\mathcal{G}$ with $\mathcal{F} \otimes_{\mathcal{O}} \mathcal{G} \cong \mathcal{O}$.

# Bibliography


1. Adem, A., and Milgram, R. J., *Cohomology of Finite Groups*, Springer-Verlag, Berlin, 1994.
2. Adian, S. I., *The Burnside Problem and Identities in Groups*, translated from the Russian by John Lennox and James Wiegold, Ergebnisse der Mathematik und ihrer Grenzgebiete 95, Springer-Verlag, New York, 1979.
3. Albert, A. A., editor, *Studies in Modern Algebra*, MAA Studies in Mathematics, Vol. 2, Mathematical Association of America, Washington, 1963.
4. Amitsur, S. A., Finite subgroups of division rings, *Trans. Amer. Math. Soc.* 80 (1955), 361–386.
5. Anderson, F. W., and Fuller, K. R., *Rings and Categories of Modules*, 2nd ed., Springer-Verlag, Berlin, 1992.
6. Arnold, D. M., A duality for torsion-free modules of finite rank over a discrete valuation ring, *Proc. London Math. Soc.* (3) 24 (1972), 204–216.
7. Artin, E., *Galois Theory*, 2nd ed., Notre Dame, 1955; Dover reprint, Mineola, 1998.
8. ———, *Geometric Algebra*, Interscience Publishers, New York, 1957.
9. Artin, E., Nesbitt, C. J., and Thrall, R. M., *Rings with Minimum Condition*, University of Michigan Press, Ann Arbor, 1968.
10. Aschbacher, M., *Finite Group Theory*, Cambridge University Press, Cambridge, 1986.
11. Aschbacher, M., Lyons, R., Smith, S. D., Solomon, R., *The Classification of Finite Simple Groups. Groups of Characteristic 2 Type*, Mathematical Surveys and Monographs, 172. American Mathematical Society, Providence, 2011.
12. Atiyah, M., and Macdonald, I. G., *Introduction to Commutative Algebra*, Addison–Wesley, Reading, 1969.
13. Atiyah, M. F., and Hirzebruch, F., Vector bundles and homogeneous spaces, *Proc. Sympos. Pure Math.*, 1961, Vol. III, pp. 7–38, American Mathematical Society, Providence, R.I.
14. Babakhanian, A., *Cohomological Methods in Group Theory*, Marcel Dekker, New York, 1972.
15. Baker, A., *Transcendental Number Theory*, Cambridge University Press, Cambridge, 1975.
16. Barr, M., The existence of free groups, *Amer. Math. Monthly* 79 (1972), 364–367.
17. Bass, H., *Algebraic K-Theory*, W. A. Benjamin, New York, 1968.
18. Becker, T., and Weispfenning, V., *Gröbner Bases*: *A Computational Approach to Commutative Algebra*, Springer-Verlag, New York, 1993.

19. Besche, H. U., Eick, B., and O'Brien, E. A., The groups of order at most 2000, *Electron. Res. Announc. AMS* 7 (2001), 1–4.
20. Biggs, N. L., *Discrete Mathematics*, 2nd ed., Oxford University Press, Oxford, 2002.
21. Birkhoff, G., and Mac Lane, S., *A Survey of Modern Algebra*, 4th ed., Macmillan, New York, 1977.
22. Blyth, T. S., *Module Theory; An Approach to Linear Algebra*, Oxford University Press, Oxford, 1990.
23. Borevich, Z. I., and Shafarevich, I. R., *Number Theory*, Academic Press, Orlando, 1966.
24. Bott, R., and Tu, L. W., *Differential Forms in Algebraic Topology*, Springer-Verlag, New York, 1982.
25. Bourbaki, N., *Elements of Mathematics*; *Algebra* I; *Chapters* 1–3, Springer-Verlag, New York, 1989.
26. ———, *Elements of Mathematics*; *Commutative Algebra*, Addison–Wesley, Reading, 1972.
27. Bridson, M. R., and Haefliger, A., *Metric Spaces of Non-positive Curvature*, Grundlehren der Mathematischen Wissenschaften 319, Springer-Verlag, Berlin, 1999.
28. Brown, K. S., *Cohomology of Groups*, Springer-Verlag, Berlin, 1982.
29. Bruns, W., and Herzog, J., *Cohen–Macaulay Rings*, Cambridge University Press, Cambridge, 1993.
30. Buchberger, B., and Winkler, F., editors, *Gröbner Bases and Applications*, LMS Lecture Note Series 251, Cambridge University Press, Cambridge, 1998.
31. Burnside, W., *The Theory of Groups of Finite Order*, 2nd ed., Cambridge University Press, Cambridge, 1911; Dover reprint, Mineola, 1955.
32. Caenepeel, S., *Brauer Groups, Hopf Algebras, and Galois Theory*, Kluwer, Dordrecht, 1998.
33. Cajori, F., *A History of Mathematical Notation*, Open Court, 1928; Dover reprint, Mineola, 1993.
34. Cameron, P. J., Finite permutation groups and finite simple groups, *Bull. London Math. Soc.* 13 (1981), no. 1, 1–22.
35. Carmichael, R., *An Introduction to the Theory of Groups*, Ginn, New York, 1937.
36. Cartan, H., and Eilenberg, S., *Homological Algebra*, Princeton University Press, Princeton, 1956.
37. Carter, R., *Simple Groups of Lie Type*, Cambridge University Press, Cambridge, 1972.
38. Cassels, J. W. S., and Fröhlich, A., *Algebraic Number Theory*, Thompson Book Co., Washington, D.C., 1967.
39. Chase, S. U., Harrison, D. K., and Rosenberg, A., *Galois Theory and Cohomology of Commutative Rings*, Mem. Amer. Math. Soc., No. 52, Providence, 1965, pp. 15–33.
40. Claborn, L., Every Abelian group is a class group, *Pacific J. Math* 18 (1966), 219–222.
41. Cohen, D. E., *Groups of Cohomological Dimension* 1, Lecture Notes in Mathematics, Vol. 245, Springer-Verlag, New York, 1972.
42. Cohn, P. M., *Free Rings and Their Relations*, Academic Press, New York, 1971.
43. Collins, D. J., Grigorchuk, R. I., Kurchanov, P. F., and Zieschang, H., *Combinatorial Group Theory and Applications to Geometry*, 2nd ed., Springer-Verlag, New York, 1998.
44. Conway, J. H., Curtis, R. T., Norton, S. P., Parker, R. A., and Wilson, R. A., *ATLAS of Finite Groups*, Oxford University Press, Oxford, 1985.
45. Cox, D., Little, J., and O'Shea, D., *Ideals, Varieties, and Algorithms*, 2nd ed., Springer-Verlag, New York, 1997.
46. Coxeter, H. S. M., and Moser, W. O. J., *Generators and Relations for Discrete Groups*, Springer-Verlag, New York, 1972.
47. Cuoco, A. A., and Rotman, J. J., *Learning Modern Algebra from Early Attempts to Prove Fermat's Last Theorem*, MAA Textbooks, Mathematical Association of America, Washington, DC, 2013.

48. Curtis, C. W., and Reiner, I., *Representation Theory of Finite Groups and Associative Algebras*, Interscience, New York, 1962.
49. Dade, E. C., Localization of injective modules, *J. Algebra* 69 (1981), 415–425.
50. Dauns, J., *Modules and Rings*, Cambridge University Press, Cambridge, 1994.
51. De Meyer, F., and Ingraham, E., *Separable Algebras over Commutative Rings*, Lecture Notes in Mathematics, Vol. 181, Springer-Verlag, New York, 1971.
52. Dieudonné, J., *La Géometrie des Groupes Classiques*, Springer-Verlag, Berlin, 1971.
53. Dixon, J. D., du Sautoy, M. P. F., Mann, A., and Segal, D., *Analytic Pro-p Groups*, 2nd ed., Cambridge University Press, Cambridge, 1999.
54. Dlab, V., and Ringel, C. M., *Indecomposable representations of graphs and algebras*, Mem. Amer. Math. Soc. 6 (1976), no. 173.
55. Doerk, K., and Hawkes, T., *Finite Soluble Groups*, de Gruyter Expositions in Mathematics 4, Walter de Gruyter, Berlin, 1992.
56. Dornhoff, L., *Group Representation Theory, Part* A, *Ordinary Representation Theory*, Marcel Dekker, New York, 1971.
57. Drozd, Yu. A., and Kirichenko, V. V., *Finite-Dimensional Algebras*, Springer-Verlag, New York, 1994.
58. Dummit, D. S., and Foote, R. M., *Abstract Algebra*, 2nd ed., Prentice Hall, Upper Saddle River, 1999.
59. Dye, R. L., On the Arf Invariant, *Journal of Algebra* 53 (1978), pp. 36–39.
60. Eisenbud, D., *Commutative Algebra with a View Toward Algebraic Geometry*, Springer-Verlag, New York, 1995.
61. Eklof, P. C., Whitehead's problem is undecideable, *American Mathematical Monthly* 83 (1976), 775–788.
62. Evens, L., *The Cohomology of Groups*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 1991.
63. Farb, B., and Dennis, R. K., *Noncommutative Algebra*, Springer-Verlag, New York, 1993.
64. Farb, B., and Margalit, D., *A Primer on Mapping Class Groups*, Princeton Mathematical Series 49, Princeton University Press, Princeton, 2012.
65. Feit, W., *Characters of Finite Groups*, W. A. Benjamin, New York, 1967.
66. Finney Jr., R. L., and Rotman, J. J., Paracompactness of locally compact Hausdorff spaces, *Michigan Math. J.* 17 (1970), 359–361.
67. Fitchas, N., and Galligo, A., Nullstellensatz effectif et conjecture de Serre (théorème de Quillen–Suslin) pour le calcul formel, *Math. Nachr.* 149 (1990), 231–253.
68. Formanek, E., Central polynomials for matrix rings, *J. Algebra* 23 (1972), 129–132.
69. Freyd, P., *Abelian Categories*, Harper & Row, New York, 1964.
70. Fröhlich, A., and Taylor, M. J., *Algebraic Number Theory*, Cambridge Studies in Advanced Mathematics 27, Cambridge University Press, Cambridge, 1991.
71. Fuchs, L., *Abelian Groups*, Publishing House of the Hungarian Academy of Science, Budapest, 1958.
72. Fuchs, L., *Infinite Abelian Groups* I, Academic Press, Orlando, 1970.
73. _____, *Infinite Abelian Groups* II, Academic Press, Orlando, 1973.
74. Fulton, W., *Algebraic Curves*, Benjamin, New York, 1969.
75. Fulton, W., and Harris, J., *Representation Theory. A First Course*, Graduate Texts in Mathematics, 129, Springer-Verlag, New York, 2004.
76. _____, *Algebraic Topology*; *A First Course*, Springer-Verlag, New York, 1995.
77. Gaal, L., *Classical Galois Theory with Examples*, 4th ed., Chelsea, American Mathematical Society, Providence, 1998.

78. Gabriel, P., Unzerlegbare Darstellungen. I, *Manuscripta Math.* 6 (1972), 71-103; correction, ibid. 6 (1972), 309.

79. Gelfand, S. I., and Manin, Y. I., *Methods of Homological Algebra*, 2nd ed., Springer-Verlag, New York, 2003.

80. Godement, R., *Topologie Algébrique et Théorie des Faisceaux*, Actualités Scientifiques et Industrielles 1252, Hermann, Paris, 1964.

81. Gorenstein, D., Lyons, R., and Solomon, R., *The Classification of the Finite Simple Groups*, Math. Surveys and Monographs, Vol. 40, American Mathematical Society, Providence, 1994.

82. Greub, W. H., *Multilinear Algebra*, Springer-Verlag, New York, 1967.

83. Gromov, M. L., *Groups of Polynomial Growth and Expanding Maps*, Inst. Hautes Études Sci. Publ. Math., No. 53, 1981, 53–73.

84. Gromov, M. L., *Metric Structures for Riemannian and non-Riemannian Spaces*. (English summary) based on the 1981 *Structures métriques pour les variétés riemanniennes*), Progress in Mathematics, 152, Birkhäuser, Boston, 1999.

85. Grothendieck, A., *Sur quelques points d'algébre homologique*, Tôhoku Math. J. (2) 9, 1957, 119–221.

86. Gruenberg, K. W., *Cohomological Topics in Group Theory*, Lecture Notes in Mathematics, 143, Springer-Verlag, New York, 1970.

87. Gunning, R. C., *Lectures on Riemann Surfaces*, Princeton Mathematical Notes, Princeton, 1966.

88. Hadlock, C., *Field Theory and Its Classical Problems*, Carus Mathematical Monographs, No. 19, Mathematical Association of America, Washington, 1978.

89. Hahn, A. J., *Quadratic Algebras, Clifford Algebras, and Arithmetic Witt Groups*, Universitext, Springer-Verlag, New York, 1994.

90. Hall, M., Jr., *The Theory of Groups*, Macmillan, New York, 1959.

91. Hall, P., *The Edmonton notes on nilpotent groups*, Queen Mary College Mathematics Notes. Mathematics Department, Queen Mary College, London, 1969.

92. ———, *The Collected Works of Philip Hall*, compiled and with a preface by K. W. Gruenberg and J. E. Roseblade, with an obituary by Roseblade, Oxford Science Publications, The Clarendon Press, Oxford University Press, New York, 1988.

93. Hardy, G. H., and Wright, E. M., *An Introduction to the Theory of Numbers*, 4th ed., Oxford University Press, Oxford, 1960.

94. Harris, J., *Algebraic Geometry*, Springer-Verlag, New York, 1992.

95. Hartshorne, R., *Algebraic Geometry*, Springer-Verlag, New York, 1977.

96. Herrlich, H., and Strecker, G. E., *Category Theory. An Introduction*, Allyn & Bacon, Boston, 1973.

97. Herstein, I. N., *Topics in Algebra*, 2nd ed., Wiley, New York, 1975.

98. ———, *Noncommutative Rings*, Carus Mathematical Monographs, No. 15, Mathematical Association of America, Washington, 1968.

99. Higgins, P. J., *Notes on Categories and Groupoids*, van Nostrand–Reinhold, London, 1971.

100. Hughes, D. R., and Singhi, N. M., Partitions in matrices and graphs, *European J. Combin.* 12 (1991), no. 3, 223–235.

101. Humphreys, J. E., *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag, New York, 1972.

102. Huppert, B., *Character Theory of Finite Groups*, de Gruyter, Berlin, 1998.

103. ———, *Endliche Gruppen* I, Springer-Verlag, New York, 1967.

104. Hurewicz, W., and Wallman, H., *Dimension Theory*, Princeton University Press, Princeton, 1948.

105. Isaacs, I. M., *Character Theory of Finite Groups*, AMS Chelsea Publishing, Providence, 2006.

106. _____, *Finite Group Theory*, Graduate Studies in Mathematics, Vol. 92, American Mathematical Society, Providence, 2008.

107. _____, *Algebra, A Graduate Course*, Graduate Studies in Mathematics, Vol. 100, American Mathematical Society, Providence, 2009.

108. _____, Roots of polynomials in algebraic extensions of fields, *American Mathematical Monthly* 87 (1980), 543–544.

109. Ivan, S. V., The free Burnside groups of sufficiently large exponents, *Internat. J. Algebra Comput.* 4 (1994).

110. Jacobson, N., *Basic Algebra* I, Freeman, San Francisco, 1974.

111. _____, *Basic Algebra* II, Freeman, San Francisco, 1980.

112. _____, *Finite-Dimensional Division Algebras over Fields*, Springer-Verlag, New York, 1996.

113. _____, *Lie Algebras*, Interscience Tracts, Number 10, Wiley, New York, 1962.

114. _____, *Structure of Rings*, Colloquium Publications, 37, American Mathematical Society, Providence, 1956.

115. Jategaonkar, A. V., A counterexample in ring theory and homological algebra, *J. Algebra* 12 (1966), 97–105.

116. Johnson, D. L., *Topics in the Theory of Group Presentations*, Cambridge University Press, Cambridge, 1980.

117. Jordan, C., *Traité des Substitutions*, Gauthier-Villars, Paris, 1870.

118. Kaplansky, I., *Commutative Rings*, University of Chicago Press, Chicago, 1974.

119. _____, *Fields and Rings*, 2nd ed., University of Chicago Press, Chicago, 1972.

120. _____, *Infinite Abelian Groups*, 2nd ed., University of Michigan Press, Ann Arbor, 1969.

121. _____, *Linear Algebra and Geometry*: *A Second Course*, Allyn & Bacon, Boston, 1969.

122. _____, *Set Theory and Metric Spaces*, Chelsea, American Mathematical Society, Providence, 1977.

123. _____, Projective modules, *Annals Math.* 68 (1958), 372–377.

124. Kharazishvili, *Nonmeasurable Sets and Functions*, North Holland Mathematics Studies 195, Elsevier, Amsterdam, 2004.

125. King, R. B., *Beyond the Quartic Equation*, Birkhäuser, Boston, 1996.

126. Knapp, A. W., *Basic Algebra*, Birkhäuser, Boston, 1996.

127. Knapp, A. W., *Advanced Algebra*, Birkhäuser, Boston, 2007.

128. Kostrikin, A. I., and Shafarevich, I. R. (editors), *Algebra* IX. *Finite Groups of Lie Type*; *Finite-Dimensional Division Algebras*, Encyclopaedia of Mathematical Sciences, 77, Springer-Verlag, New York, 1996.

129. Kurosh, A. G., *The Theory of Groups*, Volume One, Chelsea, New York, 1955.

130. Kurosh, A. G., *The Theory of Groups*, Volume Two, Chelsea, New York, 1956.

131. Lady, E. L., Nearly isomorphic torsion-free abelian groups, *J. Algebra* 35 (1975), 235–238.

132. Lam, T. Y., *The Algebraic Theory of Quadratic Forms*, Benjamin, Reading, 1973, 2nd revised printing, 1980.

133. _____, *A First Course in Noncommutative Rings*, Springer-Verlag, New York, 1991.

134. _____, Representations of finite groups: A hundred years, Part I, *Notices Amer. Math. Soc.* 45 (1998), no. 4, 465–474.

135. _____, *Lectures on Modules and Rings*, Springer-Verlag, New York, 1999.

136. _____, Lam, T. Y., *Serre's problem on projective modules*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2006.

137. Lam, T. Y., and Siu, M. K., $K_0$ and $K_1$—an introduction to algebraic $K$-theory, *Amer. Math. Monthly* 82 (1975), 329–364.

138. Lang, S., *Algebra*, Addison–Wesley, Reading, 1965.

139. Ledermann, W., *Introduction to Group Characters*, 2nd ed., Cambridge University Press, Cambridge, 1987.

140. Lennox, J. C., and Robinson, D. J. S., *The Theory of Infinite Soluble Groups*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, Oxford, 2004.

141. Leon, S. J., *Linear Algebra with Applications*, 8th ed., Prentice Hall, Upper Saddle River, 2010.

142. Lyndon, R. C., and Schupp, P. E., *Combinatorial Group Theory*, Springer-Verlag, New York, 1977.

143. Macdonald, I. G., *Algebraic Geometry*; *Introduction to Schemes*, Benjamin, New York, 1968.

144. Mac Lane, S., *Categories for the Working Mathematician*, Springer-Verlag, New York, 1971.

145. _____, *Homology*, Springer-Verlag, New York, 3rd corrected printing, 1975.

146. Mac Lane, S., and Birkhoff, G., *Algebra*, MacMillan, New York, 1967.

147. Malle, G., and Matzat, B., *Inverse Galois Theory*, Springer-Verlag, New York, 1999.

148. Mann, A., *How Groups Grow*, Cambridge University Press, Cambridge, 2012.

149. Massey, W. S., *Algebraic Topology: An Introduction*, Harcourt, Brace & World, New York, 1967.

150. Matsumura, H., *Commutative Ring Theory*, Cambridge University Press, Cambridge, 1986.

151. May, J. P., Munshi's proof of the Nullstellensatz, *Amer. Math. Monthly* 110 (2003), 133–140.

152. McCleary, J., *User's Guide to Spectral Sequences*, Publish or Perish, Wilmington, 1985.

153. McConnell, J. C., and Robson, J. C., *Noncommutative Noetherian Rings*, Wiley, New York, 1987.

154. McCoy, N. H., and Janusz, G. J., *Introduction to Modern Algebra*, 5th ed., W. C. Brown Publishers, Dubuque, Iowa, 1992.

155. Miller, W., The maximal order of an element of a finite symmetric group, *Amer. Math. Monthly* 94 (1987), 315–322.

156. Milnor, J., *Introduction to Algebraic K-Theory*, Annals of Mathematical Studies, No. 72, Princeton University Press, Princeton, 1971.

157. Mitchell, B., *Theory of Categories*, Academic Press, New York, 1965.

158. Montgomery, S., and Ralston, E. W., *Selected Papers on Algebra*, Raymond W. Brink Selected Mathematical Papers, Vol. 3, Mathematical Association of America, Washington, 1977.

159. Morita, K., Duality for modules and its application to the theory of rings with minimum condition, *Sci. Rep. Tokyo Kyoiku Daigaku* 6 (1958), 83–142.

160. Mumford, D., *The Red Book of Varieties and Schemes*, Lecture Notes in Mathematics, 1358, Springer-Verlag, New York, 1988.

161. Mumford, D., and Oda, T., *Algebraic Geometry* II, Texts and Readings in Mathematics, 73, Hindustan Book Agency, New Delhi, 2015.

162. Munkres, J. R., *Topology, A First Course*, Prentice Hall, Upper Saddle River, 1975.

163. _____, *Elements of Algebraic Topology*, Addison–Wesley, Reading, 1984.

164. Nagata, M., A general theory of algebraic geometry over Dedekind rings II, *Amer. J. Math.* 80 (1958), 382–420.

165. Navarro, G., On the fundamental theorem of finite abelian groups, *Amer. Math. Monthly* 110 (2003), pp. 153–154.

166. Neukirch, J., Schmidt, A., and Wingberg, K., *Cohomology of Number Fields*, Grundlehren der mathematischen Wissenschaften, Vol. 323, Springer-Verlag, New York, 2000.

167. Neumann, P., Stoy, G. A., and Thompson, E. C., *Groups and Geometry*, Oxford University Press, Oxford, 1994.

168. Niven, I., and Zuckerman, H. S., *An Introduction to the Theory of Numbers*, Wiley, New York, 1972.

169. Northcott, D. G., *Ideal Theory*, Cambridge University Press, Cambridge, 1953.

170. Ojanguren, M., and Sridharan, R., Cancellation of Azumaya algebras, *J. Algebra* 18 (1971), 501–505.

171. Ol'shanskii, A. Y., *Geometry of Defining Relations in Groups*, Kluwer Academic Publishers, Dordrecht, 1991.

172. O'Meara, O. T., *Introduction to Quadratic Forms*, Springer-Verlag, New York, 1971.

173. Orzech, M., and Small, C., *The Brauer Group of Commutative Rings*, Lecture Notes in Pure and Applied Mathematics, Marcel Dekker, New York, 1975.

174. Osborne, M. S., *Basic Homological Algebra*, Springer-Verlag, New York, 2000.

175. Pollard, H., *The Theory of Algebraic Numbers*, Carus Mathematical Monographs, No. 9, Mathematical Association of America, Washington, 1950.

176. Procesi, C., *Rings with Polynomial Identities*, Marcel Dekker, New York, 1973.

177. Razmyslov, Ju. P., A certain problem of Kaplansky (Russian), *Izv. Akad. Nauk SSSR* Ser. Mat. 37 (1973), 483–501.

178. Rado, R., A proof of the basis theorem for finitely generated Abelian groups, *J. London Math. Soc.* 26 (1951), pp. 75–76, erratum, 160.

179. Reid, M., and Szendrői, B., *Geometry and Topology*, Cambridge University Press, Cambridge, 2005.

180. Reiner, I., *Maximal Orders*, Academic Press, London, 1975; Oxford University Press, Oxford, 2003.

181. Robinson, D. J. S., *A Course in the Theory of Groups*, 2nd ed., Springer-Verlag, New York, 1996.

182. Roman, S., *Field Theory*, 2nd ed., Graduate Texts in Mathematics, 158, Springer, New York, 2006.

183. Rosenberg, J., *Algebraic K-Theory and Its Applications*, Springer-Verlag, New York, 1994.

184. Rosset, S., A new proof of the Amitsur-Levitski identity, *Israel Journal of Mathematics* 23 (1976), 187–188.

185. Rotman, J. J., *A First Course in Abstract Algebra*, 3rd ed., Prentice Hall, Upper Saddle River, NJ, 2006.

186. ———, *Galois Theory*, 2nd ed., Springer-Verlag, New York, 1998.

187. ———, *An Introduction to Homological Algebra*, 2nd ed., Springer, New York, 2009.

188. ———, *An Introduction to the Theory of Groups*, Graduate Texts in Mathematics, 148, 4th ed., Springer-Verlag, New York, 1995.

189. ———, The Grothendieck group of torsion-free abelian groups of finite rank, *Proc. London Math. Soc.* 13 (1963), 724–732.

190. ———, Covering complexes with applications to algebra, *Rocky Mountain J. of Math.* 3 (1973), 641–674.

191. ———, *An Introduction to Algebraic Topology*, Graduate Texts in Mathematics, 119, Springer-Verlag, New York, 1988.

192. ———, *Journey into Mathematics*, Prentice Hall, Upper Saddle River, 1998, Dover reprint, Mineola, 2007.

193. Rowen, L. H., *Polynomial Identities in Ring Theory*, Academic Press, New York, 1980.

194. ———, *Ring Theory*, Vols. I, II, Academic Press, Boston, 1988.

195. Ryser, H. J., *Combinatorial Mathematics*, The Carus Mathematical Monographs, No. 14, The Mathematical Association of America; distributed by John Wiley and Sons, Inc., New York 1963.

196. Samuel, P. *Algebraic Theory of Numbers*, Houghton Mifflin, Boston, 1970.

197. Sąsiada, E., Proof that every countable and reduced torsion-free abelian group is slender, *Bull. Acad. Polon. Sci.* 7 (1959), 143–144.

198. Serre, J.-P., Faisceaux algébriques cohérents, *Annals Math.* 61 (1955), 197–278.

199. ———, *Corps Locaux*, Hermann, Paris, 1968; English transl., *Local Fields*, Graduate Texts in Mathematics, 67, Springer-Verlag, 1979.

200. ———, *Algébre Locale: Multiplicités*, Lecture Notes in Mathematics, 11, 3rd ed., Springer-Verlag, New York, 1975; English transl., *Local Algebra*, Springer Monographs in Mathematics, Springer-Verlag, 2000.

201. ———, *Linear Representations of Finite Groups*, Graduate Texts in Mathematics, 42, Springer-Verlag, New York, 1977.

202. ———, *Trees*, Springer-Verlag, New York, 1980.

203. ———, *Topics in Galois Theory*, Jones and Bartlett, Boston, 1992.

204. Shafarevich, I. R., *Algebra* I. *Basic Notions of Algebra*, Encyclopaedia of Mathematical Sciences, 11, Springer-Verlag, Berlin, 1990.

205. Sharpe, R. W., *Differential Geometry*: *Cartan's Generalization of Klein's Erlangen Program*, Springer-Verlag, New York, 1997.

206. Simmons, G. J., The number of irreducible polynomials of degree $n$ over GF($p$), *Amer. Math. Monthly* 77 (1970), 743–745.

207. Sims, C. C., *Computation with Finitely Presented Groups*, Cambridge University Press, Cambridge, 1994.

208. Small, C., *Arithmetic of Finite Fields*, Monographs and Textbooks in Pure and Applied Mathematics, 148, Marcel Dekker, Inc., New York, 1991.

209. Solomon, R., *Lectures on Finite Groups*, Introduction to modern mathematics, Adv. Lect. Math. 33, pp. 363–390, Int. Press, Somerville, MA, 2015.

210. Spanier, E. H., *Algebraic Topology*, corrected reprint of the 1966 original, Springer-Verlag, New York, 1981.

211. Stallings, J. R., On torsion-free groups with infinitely many ends, *Ann. of Math.* (2) 88 (1968), 312–334.

212. Stewart, I., *Galois Theory*, 3rd ed., Chapman & Hall/CRC, Boca Raton, 2004.

213. Stillwell, J., *Classical Topology and Combinatorial Group Theory*, 2nd ed., Springer-Verlag, New York, 1993.

214. ———, *Mathematics and Its History*, Springer-Verlag, New York, 1989.

215. Strade, H., *Simple Lie Algebras over Fields of Positive Characteristic* I, de Gruyter, Berlin, 2004.

216. Suzuki, M., *Group Theory* I, Springer-Verlag, New York, 1982.

217. Swan, R. G., *The Theory of Sheaves*, University of Chicago Press, Chicago, 1964.

218. ———, Vector bundles and projective modules, *Trans. Amer. Math. Soc.* 105 (1962), 264–277.

219. Tennison, B. R., *Sheaf Theory*, London Mathematical Society Lecture Note Series, 20, Cambridge University Press, Cambridge, 1975.

220. Tignol, J.-P., *Galois' Theory of Algebraic Equations*, World Scientific Publishing Co., Inc., River Edge, 2001.

221. Vakil, R., *Math 216: Foundations of Algebraic Geometry*, http://math.stanford.edu/vakil/216blog/FOAGjun1113public.pdf, Stanford University, 2013.

222. van der Waerden, B. L., *Geometry and Algebra in Ancient Civilizations*, Springer-Verlag, New York, 1983.

223. ———, *A History of Algebra*, Springer-Verlag, New York, 1985.

224. ———, *Modern Algebra*, Vols. I, II, 4th ed., Ungar, New York, 1966.

225. ———, *Science Awakening*, Wiley, New York, 1963.

226. Voevodsky, V., *On motivic cohomology with $\mathbb{Z}/\ell$-coefficients*, Ann. of Math. (2) 174 (2011), no. 1, 401-438.

227. Weibel, C., *An Introduction to Homological Algebra*, Cambridge University Press, Cambridge, 1994.

228. Weyl, H., *The Classical Groups; Their Invariants and Representations*, Princeton, 1946.

229. Wehrfritz, B. A. F., *Infinite Linear Groups. An Account of the Group-Theoretic Properties of Infinite Groups of Matrices*, Ergebnisse der Matematik und ihrer Grenzgebiete, Band 76, Springer-Verlag, New York-Heidelberg, 1973.

230. ———, *Group and Ring Theoretic Properties of Polycyclic Groups. Algebra and Applications*, 10, Springer-Verlag, London, 2009.

231. Weiss, E., *Cohomology of Groups*, Academic Press, Orlando, 1969.

232. Williams, K. S., Note on non-Euclidean principal ideal domains, *Math. Mag.* 48 (1975), 176–177.

233. Zariski, O., and Samuel, P., *Commutative Algebra* I, van Nostrand, Princeton, 1958.

234. ———, *Commutative Algebra* II, van Nostrand, Princeton, 1960.

# Index

**Page numbers in italic refer to Part 1.**

This book is the second part of the new edition of *Advanced Modern Algebra* (the first part published as Graduate Studies in Mathematics, Volume 165). Compared to the previous edition, the material has been significantly reorganized and many sections have been rewritten. The book presents many topics mentioned in the first part in greater depth and in more detail. The five chapters of the book are devoted to group theory, representation theory, homological algebra, categories, and commutative algebra, respectively. The book can be used as a text for a second abstract algebra graduate course, as a source of additional material to a first abstract algebra graduate course, or for self-study.

ISBN 978-1-4704-2311-7

9 781470 423117

GSM/180

For additional information and updates on this book, visit

**www.ams.org/bookpages/gsm-180**

**www.ams.org**