# Algebra

Second Edition

Volume 3

P. M. COHN, FRS
*University College London*

JOHN WILEY & SONS
Chichester · New York · Brisbane · Toronto · Singapore

Copyright © 1991 by John Wiley & Sons Ltd.
Baffins Lane, Chichester
West Sussex PO19 1UD, England

All rights reserved.

No part of this book may be reproduced by any means,
or transmitted, or translated into a machine language
without the written permission of the publisher.

*Other Wiley Editorial Offices*

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Jacaranda Wiley Ltd, G.P.O. Box 859, Brisbane,
Queensland 4001, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

John Wiley & Sons (SEA) Pte Ltd, 37 Jalan Pemimpin 05-04,
Block B, Union Industrial Building, Singapore 2057

***Library of Congress Cataloging-in-Publication Data:***
(Revised for volume 3)
Cohn, P. M. (Paul Moritz)
Algebra.
Includes bibliographical references (v. 3, p. )
and indexes.
1. Algebra. I. Title.
QA154.2.C63 1982 512.9 81-21932
ISBN 0 471 10168 0 (v. 1)
ISBN 0 471 10169 9 (v. 1:pbk.)

***British Library Cataloguing in Publication Data:***
Cohn, P. M. (Paul Moritz)
Algebra. – 2nd. ed.
Vol. 3
1. Abstract algebra
I. Title
512.02

ISBN 0 471 92840 2

Typeset by Thomson Press (India) Ltd, New Delhi
Printed in Great Britain by Biddles Ltd, Guildford

# Contents

**Preface to the Second Edition** ix

**Conventions on terminology and notes to the reader** xi

## 1 Universal algebra

1.1 Algebras and homomorphisms . . . . . 1
1.2 Congruences and the isomorphism theorems. . . . . 4
1.3 Free algebras and varieties . . . . . 11
1.4 Abstract dependence relations . . . . . 20
1.5 The diamond lemma . . . . . 25
1.6 Ultraproducts . . . . . 27
1.7 The natural numbers. . . . . 31
Further exercises . . . . . 38

## 2 Multilinear algebra

2.1 Graded algebras . . . . . 40
2.2 Free algebras and tensor algebras . . . . . 43
2.3 The Hilbert series of a graded algebra or module . . . . . 52
2.4 The exterior algebra on a module . . . . . 58
Further exercises . . . . . 66

## 3 Homological algebra

3.1 Additive and abelian categories . . . . . 69
3.2 Functors on abelian categories . . . . . 78
3.3 The category $\mathscr{M}_R$ . . . . . 87
3.4 Homological dimension . . . . . 94
3.5 Derived functors . . . . . 99
3.6 Ext and Tor and the global dimension of a ring. . . . . 110
Further exercises . . . . . 120

## 4 Further group theory

4.1 Group extensions. . . . . 124
4.2 The Frattini subgroup and the Fitting subgroup . . . . . 135

4.3 Hall subgroups . . . . . 141
4.4 The transfer . . . . . 146
4.5 Free groups . . . . . 148
4.6 Commutators . . . . . 154
4.7 Linear groups . . . . . 159
Further exercises . . . . . 165

**5 Further field theory**

5.1 Algebraic dependence . . . . . 168
5.2 Simple transcendental extensions . . . . . 171
5.3 Separable and $p$-radical extensions . . . . . 176
5.4 Derivations . . . . . 181
5.5 Linearly disjoint extensions . . . . . 186
5.6 Infinite algebraic extensions . . . . . 196
5.7 Galois cohomology . . . . . 202
5.8 Kummer extensions . . . . . 206
Further exercises . . . . . 210

**6 Algebras**

6.1 The Krull–Schmidt theorem . . . . . 213
6.2 The projective cover of a module . . . . . 218
6.3 Semiperfect rings . . . . . 221
6.4 Equivalence of module categories . . . . . 227
6.5 The Morita context . . . . . 234
6.6 Projective, injective and flat modules . . . . . 239
6.7 Hochschild cohomology and separable algebras . . . . . 248
Further exercises . . . . . 256

**7 Central simple algebras**

7.1 Simple Artinian rings . . . . . 258
7.2 The Brauer group . . . . . 266
7.3 The reduced norm and trace . . . . . 273
7.4 Quaternion algebras . . . . . 280
7.5 Crossed products . . . . . 283
7.6 Change of base field . . . . . 289
7.7 Cyclic algebras . . . . . 295
Further exercises . . . . . 298

**8 Quadratic forms and ordered fields**

8.1 The Clifford algebra of a quadratic space . . . . . 300
8.2 The spinor norm . . . . . 306
8.3 Formally real fields . . . . . 309
8.4 The Witt ring of a field . . . . . 322

8.5 The symplectic group . . . . . 330
8.6 The orthogonal group . . . . . 337
8.7 Quadratic forms in characteristic 2 . . . . . 344
Further exercises . . . . . 347

**9 Noetherian rings and polynomial identities**

9.1 Rings of fractions . . . . . 349
9.2 Principal ideal domains . . . . . 355
9.3 Skew polynomials and Laurent series . . . . . 359
9.4 Goldie's theorem . . . . . 367
9.5 PI-algebras . . . . . 375
9.6 Varieties of PI-algebras and Regev's theorem . . . . . 380
9.7 Generic matrix rings and central polynomials . . . . . 384
9.8 Generalized polynomial identities . . . . . 389
Further exercises . . . . . 393

**10 Rings without finiteness assumptions**

10.1 The endomorphism ring of a vector space . . . . . 395
10.2 The density theorem revisited . . . . . 398
10.3 Primitive rings . . . . . 401
10.4 Semiprimitive rings and the Jacobson radical . . . . . 404
10.5 Algebras without a unit element . . . . . 409
10.6 Semiprime rings and nilradicals . . . . . 413
10.7 Prime PI-algebras . . . . . 419
10.8 Simple algebras with minimal right ideals . . . . . 423
10.9 Firs and semifirs . . . . . 425
Further exercises . . . . . 430

**11 Skew fields**

11.1 Generalities . . . . . 432
11.2 The Dieudonné determinant . . . . . 436
11.3 Free fields . . . . . 443
11.4 Valuations on skew fields . . . . . 448
11.5 Pseudo-linear extensions . . . . . 455
Further exercises . . . . . 459

**Bibliography** . . . . . 461

**List of notations** . . . . . 464

**Index** . . . . . 468

# Preface to the Second Edition

Volume 2 of the Second Edition, which included (besides new material) about half of the old Vol. 2, appeared in 1989. With the present volume the work is now complete (although many topics of current interest have had to be omitted). This volume includes the parts of Chs. 1, 3, 4, 6, 7, 8, 10, 12 left over, as well as further new material.

Ch. 1 takes over abstract dependence relations and the development of the natural numbers (viewed as algebra with a unary operator) from the old Ch. 1, together with the basic notions on universal algebra: the isomorphism theorems, free algebras and varieties as well as the ultraproduct theorem and the diamond lemma.

The material on graded algebras from Ch. 3 has been expanded to bring out the role of the universal derivation bimodule and now includes more on Hilbert series (formerly called 'gocha') and forms the new Ch. 2. Chs. 4 (homological algebra) and 6 (further field theory) become 3 and 5 respectively, with minor revisions and some rearrangement. Topics left over from Chs. 7 and 8 (real fields and quadratic forms) now form Ch. 8, where a discussion of the symplectic and orthogonal groups has been added. Ch. 12, on Noetherian and PI-rings, has become Ch. 9, while Ch. 10 (Artinian rings) now appears in Chs. 6, 7 and 10.

The main innovations (besides those already mentioned) are as follows. Ch. 4, on group theory, brings some items of general interest and importance (Hall subgroups, free groups, commutators), but also topics which find an echo elsewhere in the book, such as linear groups. Ch. 6, on algebras, deals with the Krull–Schmidt theorem, projective covers, Morita equivalence and related matters but just stops short of the representation theory of algebras, which would have required more space than was available. The new Ch. 10, on general rings, goes into a little more detail on the various radicals without attempting to the exhaustive, bringing merely 'what the educated algebraist ought to know'.

The final chapter, on skew fields, gives a simplified treatment of the Dieudonné determinant and establishes the existence of 'free fields'. While this topic might be regarded as unduly specialized by some, the proofs have now been so far simplified as to justify their inclusion in a graduate text; in any case the specializ-

ation lemma, on which the proof is based, is of independent interest. There is also a discussion of valuations on skew fields and the book ends with an example of a skew field extension of different left and right dimensions (Artin's problem), as an illustration of the preceding theory.

As in the earlier volumes, there are again numerous exercises of varying difficulty. There are also brief historical remarks, often consisting just of a name and a date. To avoid confusion with citations from the bibliography, references to the latter are given by a name and date in brackets (square for a book and round for a paper).

Once again I have received many useful comments from students and other readers of the first edition and I would like to thank them here. I also owe a debt to friends and colleagues with whom I have discussed details of the new edition: G. M. Bergman, M. L. Roberts, L. H. Rowen and A. H. Schofield. My thanks go to all of them, particularly Mark Roberts, who has read and commented on much of the MS as well as providing welcome help in reading proof-sheets. As always, the staff of John Wiley and Sons have been efficient and helpful in getting this volume published and I should like to express my thanks to them.

University College London P. M. COHN
August 1990

# Conventions on terminology and notes to the reader

References to Algebra 1 or 2 (Volumes 1 or 2) are in the form A.1, A.2 respectively, followed by the section number.

A property is said to hold for *almost all* members of a set if it holds for all but a finite number. The complement of a subset $Y$ in a set $X$ is written $X \setminus Y$. As a rule mappings are written on the right; in particular this is done when mappings have to be composed, so that $\alpha\beta$ means: first $\alpha$, then $\beta$. If $\alpha$ is a mapping from a set $X$ and $Y$ is a subset of $X$, then the restriction of $\alpha$ to $Y$ is written $\alpha|Y$.

If $G$ is a group and $H$ a subgroup, then the subsets $xH$ of $G$ (for $x \in G$) are called the *right cosets* of $H$ in $G$ and a subset $T$ of $G$ containing just one element from each right coset is called a *left transversal* of $H$ in $G$. We write $N \lhd G$ to indicate that $N$ is a normal subgroup of $G$ and for a subset $X$ of $G$, $\langle X \rangle$ or $\mathrm{gp}\{X\}$ is the subgroup generated by $X$.

All rings and monoids have a unit element or one, usually denoted by 1; by contrast an algebra (over a coefficient ring) need not have a one. A ring is *trivial* if it consists of 0 alone; this happens just when $1 = 0$. An element $a$ of a ring is called a *zero-divisor* if $a \neq 0$ and $ab = 0$ or $ba = 0$ for some $b \neq 0$; if $a$ is neither 0 nor a zero-divisor it is said to be *regular* (cf. p. 349). A non-trivial ring without zero-divisors is called an *integral domain*; this term is not taken to imply commutativity. A ring in which the non-zero elements form a group under multiplication is called a *skew field*; in the commutative case this reduces to a *field*, but sometimes (in Ch. 11) this term is also used in the general case. In any ring $R$, the set of all non-zero elements is denoted by $R^\times$; this notation is mainly used for integral domains, where $R^\times$ is a monoid. A skew field finite-dimensional over its centre is called a *division algebra*, but the term 'algebra' by itself is *not* taken to imply finite dimensionality. A ring is said to have *invariant basis number* (IBN) if any matrix with a two-sided inverse is square (cf. A.2, **4.4**). For any matrices $A$, $B$, their *diagonal sum* is defined as $A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$.

References to the bibliography are by name of author and date in square

brackets for books, and in round brackets for papers. As in Vol. 2, all results in a section are numbered consecutively; further, we again abbreviate 'if and only if' as iff and use ■ to indicate the end or absence of a proof.

The chapters are to a large extent independent (so no interdependence chart has been given), but the reader may have to turn back for the occasional result; this is usually clearly indicated.

# 1
# Universal algebra

Most algebraic systems such as groups, vector spaces, rings, lattices etc. can be regarded from a common point of view as sets with operations defined on them, subject to certain laws. This allows many basic results, such as the isomorphism theorems, to be stated and proved quite generally. Of the general theory of universal algebra (by now quite extensive) we shall need very little, in addition to the basic concepts only the notion of an algebraic variety, i.e. a class of algebraic systems defined by identical relations, or *laws*. But there are one or two other topics, not strictly speaking part of the subject, which are discussed here, namely dependence relations and the diamond lemma. There is also the ultraproduct theorem, a result from logic with many uses (cf. Ch. 11). The chapter ends with an axiomatic development of the natural numbers, regarded as an algebraic system, in an account following Henkin (1960) (cf. also Cohn [1981]).

## 1.1 Algebras and homomorphisms

Algebraic structures show certain common features: they have operations defined on them, such as multiplication, which satisfy laws such as the associative law. The operations are mostly binary (like multiplication or addition), but sometimes they are unary, e.g. taking the square of a number, or ternary, e.g. the basic operation in a ternary ring (occurring in the study of projective planes, cf. M. Hall [1959]), or even noughtary, like the neutral element in a group. For any integer $n \geqslant 0$ we define an *n-ary operation* on a set $S$ to be a mapping of $S^n$ into $S$. The number $n$ is called the *arity* of the operation and we say *unary* for 1-ary, *binary* for 2-ary, *ternary* for 3-ary and *finitary* to mean: $n$-ary for some natural number $n$. A 0-ary operation on $S$ is just a particular element of $S$; this is also called a *constant operation.*

An algebra is to be thought of as a set with certain finitary operations defined on it, but in order to compare different algebras we need to establish a correspondence between their sets of operations. This is done by indexing the operations in each algebra by a given index set, which is kept fixed in any discussion. Its elements are called *operators*, each with a given arity.

Thus by an *operator domain* we understand a set $\Omega$ and a mapping $a:\Omega \to \mathbf{N}_0$. The elements of $\Omega$ are called *operators*; if $\omega \in \Omega$, then $a(\omega) \in \mathbf{N}_0$ is called the *arity*

of $\omega$. We shall write $\Omega(n) = \{\omega \in \Omega \mid a(\omega) = n\}$, and refer to the members of $\Omega(n)$ as *n-ary* operators.

An $\Omega$-*algebra* is defined as a pair $(A, \Omega)$ consisting of a set $A$ together with a family of operations indexed by $\Omega$:

$$\omega : A^n \to A \qquad \text{for each } \omega \in \Omega(n),\ n = 0, 1, \ldots . \tag{1}$$

The set $A$ is called the *carrier* of the algebra. Strictly speaking we should denote the algebra by $(A, \Omega, \varphi)$, where $\varphi$ is the family of mappings $\varphi_n : \Omega(n) \to \mathrm{Map}(A^n, A)$ defined by (1), but usually we shall not distinguish notationally between an algebra and its carrier. The set $\Omega$ is called the operator domain, or also the *signature* of the algebra. We give some examples.

1. Groups. A *group* $(G, \cdot, ^{-1}, 1)$ is given by a binary operation (multiplication), a unary operation (inversion) and a constant operation (the neutral element), satisfying certain laws which are familiar to the reader (cf. **1·3** below).

2. Rings. A *ring* $(R, +, -, \times, 0, 1)$ is given by two binary operations $(+, \times)$, two constant operations $(0, 1)$ and a unary operation $(-)$, again satisfying well known laws.

3. Lattices. A *lattice* may be defined as a partially ordered set in which each pair of elements has a sup and an inf, or as an algebra $(L, \vee, \wedge)$ with two binary operations satisfying certain laws (cf. A.2, **2·1**). For Boolean algebras we require in addition a constant operation 0 and a unary operation $'$ (this leads to another constant operation $1 = 0'$, an instance of a derived operation, cf. **1·3**).

4. Vector spaces. Let $k$ be a field. A *vector space* over $k$ is an algebra $(V, +, 0, k)$ with a binary operation $+$, a constant operation 0 and a family of unary operations indexed by $k$: $\omega_\alpha : u \mapsto \alpha u$ $(u \in V,\ \alpha \in k)$, satisfying the laws familiar from linear algebra. For an infinite field $k$ this is an example of an algebra with an infinite signature.

5. A 1-element set has a unique $\Omega$-algebra structure, for any $\Omega$. This is called the *trivial* $\Omega$-algebra.

6. The empty set is an $\Omega$-algebra precisely when $\Omega$ has no constant operators.

Given an $\Omega$-algebra $A$ and $\omega \in \Omega(n)$, we can apply $\omega$ to any $n$-tuple $a_1, \ldots, a_n \in A$ and obtain another element of $A$, which is written $a_1 \cdots a_n \omega$. In the case $n = 0$ this just singles out an element of $A$, denoted by $\omega$; the zero element in a ring is an example.

Many algebraic concepts can be formulated for general $\Omega$-algebras. Thus given an $\Omega$-algebra $A$, an $\Omega$-*subalgebra* is an $\Omega$-algebra $B$ whose carrier is a subset of that of $A$ and which is closed under the operations of $\Omega$, as defined in $A$. It is clear from the definition that a given subset of $A$ can be defined as a subalgebra of $A$ in at most one way. To give an example, the ring $\mathbf{Z}$ of integers has no proper subrings, because the constant operation 1 already generates the whole of $\mathbf{Z}$. The subset $\{0\}$ is again a ring, the trivial ring, but it is not a subring because the operation 1 has different values on $\mathbf{Z}$ and on $\{0\}$.

It is not hard to see that the intersection of any family of subalgebras is again a subalgebra. Hence for any subset $X$ of an algebra $A$ we can form the intersection of all subalgebras containing $X$. This is called the subalgebra of $A$ *generated* by $X$; it may also be obtained by applying the operations of $\Omega$ to the elements of $X$ and repeating this operation a finite number of times. If the subalgebra generated by $X$ is the whole of $A$, then $X$ is called a *generating set* of $A$.

A mapping $f: A \to B$ between $\Omega$-algebras $A, B$ is said to be *compatible* with $\omega \in \Omega(n)$ if for all $a_1, \ldots, a_n \in A$,

$$(a_1 f) \cdots (a_n f)\omega = (a_1 \cdots a_n \omega) f. \tag{2}$$

If $f$ is compatible with each $\omega \in \Omega$, it is called a *homomorphism* from $A$ to $B$. If a homomorphism from $A$ to $B$ has an inverse which is again a homomorphism, it is called an *isomorphism*, and $A, B$ are then said to be *isomorphic*. For example, all 1-element $\Omega$-algebras are isomorphic. As in more special cases, an isomorphism of an algebra with itself is called an *automorphism* and a homomorphism of an algebra into itself is an *endomorphism*.

We observe that a homomorphism is determined once it is known on a generating set. This is the content of

PROPOSITION 1.1 *Let $f, g: A \to B$ be two homomorphisms between $\Omega$-algebras $A$ and $B$. If $f$ and $g$ agree on a generating set, then they are equal.*

*Proof.* The set $\{x \in A \mid xf = xg\}$ is easily seen to be a subalgebra of $A$. By hypothesis it contains a generating set of $A$, hence it is the whole of $A$ and so $f = g$, as we had to show. ■

From any family $(A_i)$ of $\Omega$-algebras we can form the direct product $P = \prod A_i$; its carrier is the Cartesian product of the $A_i$, and the operations are carried out componentwise. Thus if $\pi_i: P \to A_i$ are the projections from the Cartesian product to the factors, then any $\omega \in \Omega(n)$ is defined on $P$ by the equation

$$(a_1 \cdots a_n \omega)\pi_i = (a_1 \pi_i) \cdots (a_n \pi_i)\omega. \tag{3}$$

It is easily checked that this defines an $\Omega$-algebra structure on $P$ and the form of (3) shows that each projection $\pi_i$ is a homomorphism from $P$ to $A_i$.

Of course the $A_i$ need not be all distinct. If e.g. $A_i = A$ for all $i \in I$, we obtain the *direct power* of $A$ indexed by $I$, which is denoted by $A^I$. Its members may be regarded as functions $f: I \to A$ and the operations are defined componentwise; e.g. if an addition is defined in $A$, then in $A^I$ we have

$$(f + g)(i) = f(i) + g(i), \qquad i \in I.$$

## Exercises

(1) Show that the set of all subalgebras of an $\Omega$-algebra is a complete lattice (i.e. a lattice in which every subset has a sup and an inf, cf. A.2, **2·1**).

(2) Verify the equivalence of the two definitions of subalgebra generated by $X$, given in the text, i.e. show that the set obtained from $X$ by repeatedly applying $\Omega$ is the least subalgebra containing $X$.

(3) Show that if $\Omega$ is finite, there are only finitely many $\Omega$-algebras on a given finite set as carrier. Is there a bound in terms of the size of the carrier alone?

(4) Show that every homomorphism which is bijective is an isomorphism.

(5) Let $A$ be an $\Omega$-algebra with a carrier of $n$ elements. Show that $A$ has at most $n!$ automorphisms and at most $n^n$ endomorphisms. Find bounds on the number of auto- and endomorphisms if $\Omega$ includes a constant operator. Find bounds if $A$ has an $r$-element generating set.

(6) Let $A$ be an $\Omega$-algebra. Show that the set $\mathrm{Map}(A)$ of all mappings of $A$ into itself may be regarded as an $\Omega$-algebra. Further show that $\mathrm{End}(A)$, the set of all endomorphisms, is a subalgebra of $\mathrm{Map}(A)$ provided the following condition is satisfied by $A$: Given $\theta \in \Omega(m)$, $\omega \in \Omega(n)$ and any $m \times n$ matrix over $A$, the element of $A$ obtained by applying $\theta$ to each column and $\omega$ to the result is the same as the element obtained by applying $\omega$ to each row and $\theta$ to the result.

## 1.2 Congruences and the isomorphism theorems

Let $A$ and $B$ be any sets. By a *correspondence* from $A$ to $B$ we understand a subset of the Cartesian product $A \times B$. For example, a mapping $f: A \to B$ may be defined by a correspondence $\Gamma_f$ from $A$ to $B$ which has the properties of being (i) everywhere defined and (ii) single-valued:

(i) *for each* $a \in A$ *there exists* $b \in B$ *such that* $(a, b) \in \Gamma_f$,
(ii) *if* $(a, b), (a, b') \in \Gamma_f$ *then* $b = b'$.

This correspondence is sometimes called the *graph* of the mapping $f$.

We shall define two operations on correspondences. For any $\Gamma \subseteq A \times B$ we have the *inverse*, defined as

$$\Gamma^{-1} = \{(b, a) \in B \times A \mid (a, b) \in \Gamma\};$$

next, if $\Gamma \subseteq A \times B$ and $\Delta \subseteq B \times C$, then their *composition* is given by

$$\Gamma \circ \Delta = \{(a, c) \in A \times C \mid (a, x) \in \Gamma \text{ and } (x, c) \in \Delta \text{ for some } x \in B\}.$$

Further, if $\Gamma \subseteq A \times B$ and $A' \subseteq A$, we define $A'\Gamma = \{b \in B \mid (a, b) \in \Gamma \text{ for some } a \in A'\}$.

On every set $A$ we have the identity correspondence, also called the *diagonal*, $1_A = \{(a, a) \mid a \in A\}$ and the *universal correspondence* $A^2 = \{(a, a') \mid a, a' \in A\}$. For example, the above conditions (i), (ii) on $\Gamma_f$ to be the graph of a mapping can be expressed as follows:

$$\Gamma_f \circ \Gamma_f^{-1} \supseteq 1_A \qquad \Gamma_f^{-1} \circ \Gamma_f \subseteq 1_B.$$

To give another example, an equivalence on $A$ may be defined as a subset $\Gamma$ of $A^2$ with the properties

(i) $\Gamma \circ \Gamma \subseteq \Gamma$ (transitivity)
(ii) $\Gamma^{-1} = \Gamma$ (symmetry)
(iii) $\Gamma \supseteq 1_A$ (reflexivity).

To use correspondences in the study of $\Omega$-algebras, we shall need to know their behaviour as subalgebras.

LEMMA 2.1 *Let $A, B, C$ be $\Omega$-algebras and $\Gamma$, $\Delta$ subalgebras of $A \times B, B \times C$ respectively. Then $\Gamma^{-1}$ is a subalgebra of $B \times A$, $\Gamma \circ \Delta$ is a subalgebra of $A \times C$ and for any subalgebra $A'$ of $A$, $A'\Gamma$ is a subalgebra of $B$.*

*Proof.* Take $\omega \in \Omega(n)$ and $(a_i, c_i) \in \Gamma \circ \Delta$ $(i = 1, \ldots, n)$, say $(a_i, b_i) \in \Gamma$, $(b_i, c_i) \in \Delta$, and put $a_1 \cdots a_n \omega = a$, $b_1 \cdots b_n \omega = b$, $c_1 \cdots c_n \omega = c$. Then since $\Gamma, \Delta$ are subalgebras, we have $(a, b) \in \Gamma$, $(b, c) \in \Delta$, hence $(a, c) \in \Gamma \circ \Delta$ and since this holds for all $\omega \in \Omega$, it shows $\Gamma \circ \Delta$ to be a subalgebra. The proof for $\Gamma^{-1}$ and $A'\Gamma$ is quite similar and may be left to the reader. ■

Let $S, T$ be any sets and $f: S \to T$ a mapping between them. Then the *image* of $f$ is defined as $S\Gamma_f$, also written $\operatorname{im} f$; the *kernel* of $f$ is defined as the correspondence

$$\ker f = \{(x, y) \in S^2 \mid xf = yf\}. \tag{1}$$

In terms of the graph $\Gamma_f$ of $f$ we have

$$\ker f = \Gamma_f \circ \Gamma_f^{-1}.$$

Clearly it is an equivalence on $S$; the different equivalence classes are just the inverse images of elements in the image, sometimes called the *fibres* of $f$.

Let us consider how the above definition is related to the kernel of a homomorphism of groups. Given a group homomorphism $f: G \to H$, the kernel of $f$ in the usual sense is the inverse image under $f$ of the unit element of $H$; this is a normal subgroup $N$ of $G$ and the different cosets of $N$ in $G$ are just the fibres of $f$. So $\ker f$ as defined in (1) is the set of cosets of $N$ in $G$. Since this collection is entirely determined by $N$, it makes sense to replace it by $N$, which is what is usually done in group theory. But for arbitrary sets we shall need the whole correspondence $\ker f$ as defined above and we cannot replace it by anything simpler. This is still true when we come to study kernels of homomorphisms of $\Omega$-algebras.

Let $S, T$ be any sets and $\Gamma$ a correspondence from $S$ to $T$. We shall use $\Gamma$ to define systems of subsets of $S, T$ between which there is an inclusion-reversing bijection, as follows.

For any subset $X$ of $S$ we define a subset $X^*$ of $T$ by

$$X^* = \{y \in T \mid (x, y) \in \Gamma \text{ for all } x \in X\},$$

and similarly, for any subset $Y$ of $T$ we define a subset $Y^*$ of $S$ by

$$Y^* = \{x \in S \mid (x, y) \in \Gamma \text{ for all } y \in Y\}.$$

We thus have mappings

$$X \mapsto X^*, \qquad Y \mapsto Y^* \tag{2}$$

of $\mathscr{P}(S)$, $\mathscr{P}(T)$ into each other with the properties:

$$X_1 \subseteq X_2 \Rightarrow X_1^* \supseteq X_2^*, \qquad Y_1 \subseteq Y_2 \Rightarrow Y_1^* \supseteq Y_2^*, \tag{3}$$

$$X \subseteq X^{**}, \qquad Y \subseteq Y^{**} \tag{4}$$

$$X^{***} = X^*, \qquad Y^{***} = Y^*. \tag{5}$$

Conditions (3) and (4) are immediate from the definitions. If (3) is applied to (4) we get $X^* \supseteq X^{***}$ and (4) applied with $X^*$ in place of $X$ gives $X^* \subseteq X^{***}$. Hence $X^{***} = X^*$ and similarly for $Y^*$. This proves (5) as a consequence of (3) and (4) alone.

A pair of mappings (2) between $\mathscr{P}(S)$ and $\mathscr{P}(T)$ satisfying (3), (4) and hence (5) is called a *Galois connexion.* An obvious example, which also accounts for the name, is the situation in field theory. If $F$ is a field and $G$ the group of all its automorphisms, then the pairs $(x, \alpha) \in F \times G$ such that $x^\alpha = x$ form a correspondence which establishes a Galois connexion between certain subfields of $F$ and subgroups of $G$. If $G$ is a finite group of automorphisms of $F$ and $k$ is the subfield of elements left fixed by $G$, then there is a correspondence between all subgroups of $G$ and all fields between $k$ and $F$ (cf, A.2, **3.6** and **5.6** below).

Let us define a *congruence* on an $\Omega$-algebra $A$ as an equivalence on $A$ which is also a subalgebra of $A^2$. For example $1_A$ and $A^2$ are congruences on $A$, and every other congruence $\mathfrak{q}$ lies between these two: $1_A \subseteq \mathfrak{q} \subseteq A^2$. The congruence on $\mathbf{Z}$ (as ring) determined by a given positive integer $m$ consists of the residue classes mod $m$, i.e. the sets of numbers leaving a given residue after division by $m$. As in this example, we shall sometimes, for any congruence $\mathfrak{q}$ on $A$, write $a \equiv b \pmod{\mathfrak{q}}$ to mean $(a, b) \in \mathfrak{q}$.

The next two results explain the significance of congruences for algebras.

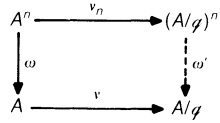
THEOREM 2.2 *Let $f: A \to B$ be a homomorphism of $\Omega$-algebras. Then* im $f$ *is a subalgebra of $B$ and* ker $f$ *is a congruence on $A$.*

*Proof.* It is easily checked that the graph $\Gamma_f$ of $f$ is a subalgebra of $A \times B$. By Lemma 2.1, im $f = A\Gamma_f$ is a subalgebra of $B$ and ker $f = \Gamma_f \circ \Gamma_f^{-1}$ is a subalgebra of $A^2$, therefore it is a congruence. ■

Given a group $G$ with a normal subgroup $N$, we can put a group structure on the set $G/N$ such that the natural mapping $G \to G/N$ is a homomorphism. In the same way we can, for any congruence $\mathfrak{q}$ on $A$, define an algebra structure on the set of $\mathfrak{q}$-classes $A/\mathfrak{q}$ such that the natural mapping $A \to A/\mathfrak{q}$ is a homomorphism with kernel $\mathfrak{q}$. This is the content of

THEOREM 2.3 *Let $A$ be an $\Omega$-algebra and $\mathfrak{q}$ a correspondence on $A$. Then there exists a unique $\Omega$-algebra structure on the set $A/\mathfrak{q}$ of all $\mathfrak{q}$-classes such that the natural mapping $v\colon A \to A/\mathfrak{q}$ is a homomorphism.*

*Proof.* The natural mapping $v: A \to A/\mathfrak{q}$ is well-defined because $\mathfrak{q}$ is an equivalence. It induces a mapping $v_n: A^n \to (A/\mathfrak{q})^n$ for $n = 0, 1, \ldots$ in an obvious fashion. Let us write $\bar{x}$ for the residue class (mod $\mathfrak{q}$) of $x \in A$. To complete the proof we must show that for each $\omega \in \Omega(n)$ there is just one way to complete the diagram




to a commutative square; thus we have to find a map $\omega': (A/\mathfrak{q})^n \to A/\mathfrak{q}$ such that

$$\bar{a}_1 \cdots \bar{a}_n \omega' = \overline{(a_1 \cdots a_n \omega)}. \tag{6}$$

This equation evidently defines $\omega'$ uniquely if we can show that the right-hand side is independent of the choice of $a_i$ in its $\mathfrak{q}$-class. Let $(a_i, a_i') \in \mathfrak{q}$; since $\mathfrak{q}$ is a subalgebra, we have

$$(a_1 \cdots a_n \omega, a_1' \cdots a_n' \omega) \in \mathfrak{q},$$

i.e.

$$\overline{a_1 \cdots a_n \omega} = \overline{a_1' \cdots a_n' \omega},$$

and this is what we had to show. ■

The algebra so defined on $A/\mathfrak{q}$ is again denoted by $A/\mathfrak{q}$ and is called the *quotient algebra* of $A$ by $\mathfrak{q}$, with the natural homomorphism $v: A \to A/\mathfrak{q}$. For example, as we have seen, $A$ always has the congruences $1_A$, $A^2$; the corresponding quotients are $A$ and the trivial $\Omega$-algebra consisting of a single element. An $\Omega$-algebra is said to be *simple* if it is non-trivial and has no quotients other than itself and the trivial algebra. It follows that an algebra $A$ is simple iff it is non-trivial and has no congruences other than $1_A$ or $A^2$.

The isomorphism theorems for groups all have precise analogues for $\Omega$-algebras. We begin with the factor theorem, which is also familiar in the case of groups (A.1, Th. 2 of **9.1**).

THEOREM 2.4 (Factor theorem) *Let $f: A \to B$ be a homomorphism of $\Omega$-algebras and $\mathfrak{q}$ a congruence on $A$ such that $\mathfrak{q} \subseteq \ker f$. Then there is a unique homomorphism $f': A/\mathfrak{q} \to B$ such that $f = vf'$, where $v$ is the natural homomorphism from $A$ to $A/\mathfrak{q}$. Further, $f'$ is injective if and only if $\mathfrak{q} = \ker f$.*

*Proof.* Let us again write $\bar{x}$ for the $\mathfrak{q}$-class of $x \in A$. If a homomorphism $f'$ with the stated properties exists, then it must satisfy

$$\bar{a}f' = af, \qquad a \in A. \tag{7}$$

Thus there can be at most one such mapping. To show that there is one we have to verify that the right-hand side of (7) depends only on the $\mathfrak{q}$-class containing $a$ and not on $a$ itself. Let $(a, a') \in \mathfrak{q}$; then $(a, a') \in \ker f$, hence $af = a'f$ as claimed. Thus there is a unique well-defined mapping $f'$ to satisfy (7), and it only remains to show that $f'$ is a homomorphism. Given $a_i \in A$, $\omega \in \Omega(n)$, we have

$$(a_1 \cdots a_n \omega)f = (a_1 f) \cdots (a_n f)\omega,$$

hence by (7),

$$\overline{(a_1 \cdots a_n \omega)}f' = (\bar{a}_1 f') \cdots (\bar{a}_n f')\omega,$$

and this shows $f'$ to be a homomorphism. It is injective iff no two distinct $\mathfrak{q}$-classes are identified by $f'$ and this is just the condition that $\mathfrak{q} = \ker f$. ■

THEOREM 2.5 (First isomorphism theorem) *Let $f: A \to B$ be a homomorphism of $\Omega$-algebras. Then*

$$A/\ker f \cong \operatorname{im} f. \tag{8}$$

*Thus $f$ may be factorized as $f = vf_1\mu$, where $v: A \to A/\mathfrak{q}$ is the natural homomorphism, $f_1$ is the isomorphism (8) and $\mu: \operatorname{im} f \to B$ is the inclusion homomorphism.*

*Proof.* By applying the factor theorem with $\mathfrak{q} = \ker f$ we find $f': A/\ker f \to B$ such that $f = vf'$, where $f'$ is injective. Its image is $\operatorname{im} f$, so there is an isomorphism $f_1: A/\ker f \to \operatorname{im} f$ to satisfy $f = vf_1\mu$, as claimed. ■

THEOREM 2.6 (Second isomorphism theorem) *Let $A$ be an $\Omega$-algebra, $A_1$ a subalgebra of $A$ and $\mathfrak{q}$ a congruence on $A$. Then the union of all $\mathfrak{q}$-classes meeting $A_1$ is a subalgebra $A_1^{\mathfrak{q}}$ of $A$, $\mathfrak{q}_1 = \mathfrak{q} \cap A_1^2$ is a congruence on $A_1$ and we have an isomorphism*

$$A_1/\mathfrak{q}_1 \cong A_1^{\mathfrak{q}}/\mathfrak{q}. \tag{9}$$

*Proof.* Let $v: A \to A/\mathfrak{q}$ be the natural homomorphism and $v_1$ its restriction to $A_1$. Then $v_1$ is a homomorphism of $A_1$ into $A/\mathfrak{q}$; its image is the set of $\mathfrak{q}$-classes

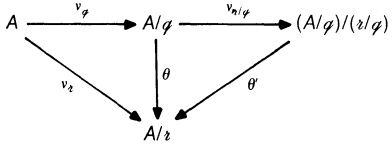
meeting $A_1$, namely $A_1^{\mathfrak{q}}/\mathfrak{q}$, and its kernel is $\mathfrak{q} \cap A_1^2 = \mathfrak{q}_1$. Applying Th. 2.5 we obtain (9). ■

Similarly, by applying the factor theorem with $B = A/\mathfrak{r}$ and the natural homomorphism $v_{\mathfrak{r}}: A \to A/\mathfrak{r}$ for $f$, we obtain

THEOREM 2.7 (Third isomorphism theorem) *Let $A$ be an $\Omega$-algebra and $\mathfrak{q}$, $\mathfrak{r}$ any congruences on $A$ such that $\mathfrak{q} \subseteq \mathfrak{r}$. Then there is a unique homomorphism $\theta: A/\mathfrak{q} \to A/\mathfrak{r}$ such that $v_{\mathfrak{q}}\theta = v_{\mathfrak{r}}$. Further, $\ker\theta$ is the set of pairs of $\mathfrak{q}$-classes that are identified in $A/\mathfrak{r}$. Denoting this set by $\mathfrak{r}/\mathfrak{q}$, we find that $\mathfrak{r}/\mathfrak{q}$ is a congruence on $A/\mathfrak{q}$ and $\theta$ induces an isomorphism*

$$\theta': (A/\mathfrak{q})/(\mathfrak{r}/\mathfrak{q}) \to A/\mathfrak{r}, \tag{10}$$

*such that $\theta = v_{\mathfrak{r}/\mathfrak{q}}\theta'$.* ■




If we fix $\mathfrak{q}$ and vary $\mathfrak{r}$, we obtain

COROLLARY 2.8 *Let $A$ be an $\Omega$-algebra and $\mathfrak{q}$ a congruence on $A$. There is a natural bijection between the set of congruences on $A$ which contain $\mathfrak{q}$ and the congruences on $A/\mathfrak{q}$,* and if $\mathfrak{r}$, $\mathfrak{r}/\mathfrak{q}$ *correspond in this way, then*

$$A/\mathfrak{r} \cong (A/\mathfrak{q})/(\mathfrak{r}/\mathfrak{q}). \quad ■$$

In particular, we see that $A/\mathfrak{q}$ is simple iff $\mathfrak{q}$ is a maximal proper congruence on $A$. We note the standard application of Zorn's lemma to obtain maximal subalgebras:

THEOREM 2.9 *Let $A$ be an $\Omega$-algebra, $A'$ a subalgebra and $S$ a subset of $A$. Then there exists a subalgebra $C$ of $A$ which is maximal subject to the conditions $C \supseteq A'$, $C \cap S = A' \cap S$.*

*Proof.* The family of all subalgebras $C$ such that $C \supseteq A'$, $C \cap S = A' \cap S$ is easily seen to be inductive; hence by Zorn's lemma there is a subalgebra which is maximal subject to these conditions. ■

Since congruences on $A$ are certainly subalgebras of $A^2$ and the collection of all congruences is inductive, we obtain

COROLLARY 2.10 *Let $A$ be an $\Omega$-algebra, $\Gamma$ a correspondence on $A$ and $\mathfrak{q}$ a congruence on $A$. Then there exists a congruence $\bar{\mathfrak{q}}$ on $A$ which is maximal subject to the conditions $\bar{\mathfrak{q}} \supseteq \mathfrak{q}$, $\bar{\mathfrak{q}} \cap \Gamma = \mathfrak{q} \cap \Gamma$.* ■

We conclude this section with a construction which is often used, the subdirect product. Let us again consider a direct product of $\Omega$-algebras: $P = \prod A_i$, with projections $\pi_i : P \to A_i$. It is easily seen that $P$ may be characterized by the properties:

(i) *for any $x, y \in P$, if $x\pi_i = y\pi_i$ for all $i$, then $x = y$,*
(ii) *given any family $(a_i)$, where $a_i \in A_i$, there exists $x \in P$ such that $x\pi_i = a_i$.*

We often encounter situations where only (i) holds. This means that we are dealing essentially with a certain subalgebra of the direct product $\prod A_i$. Let us define formally:

Let $(A_i)$ be a family of $\Omega$-algebras, with direct product $P = \prod A_i$ and projections $\pi_i : P \to A_i$. An algebra $A$ is called a *subdirect product* of the $A_i$ if there is an embedding of $A$ in $P$ such that the image is mapped by $\pi_i$ onto $A_i$.

We remark that any subalgebra $A$ of $\prod A_i$ is a subdirect product of the family $A_i'$, where $A_i'$ is the image of the restriction map $\pi_i | A$. Subdirect products usually arise as follows.

PROPOSITION 2.11 *Let $A$ be an $\Omega$-algebra and $(\mathfrak{q}_i)$ a family of congruences on $A$. Put $\mathfrak{q} = \cap \mathfrak{q}_i$, $A_i = A/\mathfrak{q}_i$. Then $A/\mathfrak{q}$ is a subdirect product of the family $(A_i)$.*

*Proof.* The map $\theta : A \to \prod A_i$ defined by

$$a\theta = (a_i) \qquad \text{where } a_i \text{ is the } \mathfrak{q}_i\text{-class of } a, \tag{11}$$

is a homomorphism and its kernel is clearly $\cap \mathfrak{q}_i = \mathfrak{q}$. Dividing by $\mathfrak{q}$, we obtain an embedding $A/\mathfrak{q} \to \prod A_i$ and by (11) $\theta\pi_i$ is surjective, hence $A/\mathfrak{q}$ is a subdirect product of the $A_i$. ■

Of course if the congruences $\mathfrak{q}_i$ intersect in $1_A$, then $A$ itself is a subdirect product of the $A_i$.

This proposition may also be expressed as follows. A congruence $\mathfrak{q}$ on $A$ is said to *separate* $a, b \in A$ if $a \not\equiv b \pmod{\mathfrak{q}}$. Given a class $\mathscr{P}$ of $\Omega$-algebras, an algebra $A$ is said to be *residually-$\mathscr{P}$* if for each pair of distinct elements $a$, $b$ of $A$ there is a congruence $\mathfrak{q}$ separating $a$ and $b$ such that $A/\mathfrak{q} \in \mathscr{P}$. Now Prop. 2.11 tells us that an algebra is residually-$\mathscr{P}$ iff it is a subdirect product of algebras in $\mathscr{P}$.

A family $(\mathfrak{q}_i)$ of congruences on $A$ is said to be *separating* if $\cap \mathfrak{q}_i = 1_A$. This just means that any pair of distinct elements is separated by some $\mathfrak{q}_i$. It is clear that any family of congruences on $A$ which includes $1_A$ is separating. If the set of all congruences $\neq 1_A$ on $A$ is separating, $A$ is called *subdirectly reducible*; otherwise $A$ is *subdirectly irreducible*. We note that the trivial algebra is

subdirectly reducible according to this definition. Let us also remark that $A$ is subdirectly irreducible iff in every subdirect product representation of $A$ at least one of the homomorphisms to the factors is an isomorphism. With this definition we have the following representation theorem due to G. Birkhoff 1944.

THEOREM 2.12 *Every $\Omega$-algebra $A$ is a subdirect product of subdirectly irreducible $\Omega$-algebras which are homomorphic images of $A$.*

*Proof.* Since the trivial algebra can be written as the empty product, we may take $A$ to be non-trivial. If $a, b \in A$, $a \neq b$, then by Cor. 2.10, there is a maximal congruence $\mathfrak{q}_0$ not containing $(a, b)$. In the algebra $A/\mathfrak{q}_0$ we have $a \not\equiv b \pmod{\mathfrak{q}_0}$, but $a \equiv b \pmod{\mathfrak{q}'}$ for all $\mathfrak{q}' \supset \mathfrak{q}_0$; hence $A/\mathfrak{q}_0$ is subdirectly irreducible. Moreover, if $(\mathfrak{q}_i)$ is the family of congruences formed for all such pairs $a, b$, then $\cap \mathfrak{q}_i = 1_A$ because any pair $a \neq b$ is separated by some $\mathfrak{q}_i$. Thus by Prop. 2.11, $A$ is a subdirect product of subdirectly irreducible algebras $A/\mathfrak{q}_i$, each a homomorphic image of $A$. ■

## Exercises

(1) Write down the conditions for a correspondence from $S$ to $T$ to be a bijection.

(2) Describe a partial ordering on a set $S$ in terms of the correspondence $\{(a, b) \in S^2 \mid a \leqslant b\}$.

(3) Fill in the details in the proof of Lemma 2.1.

(4) Verify that the kernel of a ring homomorphism (in the sense defined in the text) is the equivalence whose classes are the cosets of an ideal. Consider the isomorphism theorems of the text in the case of rings (cf. A.1, **10·1**).

(5) Verify that sets (without structure) can be regarded as the special case of $\Omega$-algebras with $\Omega = \varnothing$. Interpret the factor theorem and the isomorphism theorems for sets in this way.

(6) Show that $\mathbf{Z}$ as a ring is a subdirect product of the fields $\mathbf{F}_p$, where $p$ runs over all primes. Do the same for $\mathbf{F}_q$ where $q = p^n$ for a fixed prime $p$ and $n = 1, 2, \ldots$.

## 1.3 Free algebras and varieties

In order to study $\Omega$-algebras we shall need to form expressions in indeterminates, just as polynomials in one or more indeterminates are used to study rings. Let $X = \{x_1, x_2, \ldots\}$ be any set, our *alphabet*, usually taken to be countable, and $\Omega$ any operator domain. We define an $\Omega$-algebra $W(\Omega; X)$, the algebra of all $\Omega$-rows in $X$, as follows: An $\Omega$-*row* in $X$ is a finite sequence of elements in the set $\Omega \cup X$

(where $X$ is assumed disjoint from $\Omega$). The action of $\Omega$ is by juxtaposition; thus if $\omega \in \Omega(n)$ and $u_1, \ldots, u_n \in W(\Omega; X)$, then the effect of $\omega$ on the $n$-tuple $(u_1, \ldots, u_n)$ is the row

$$u_1 u_2 \cdots u_n \omega.$$

Clearly $X$ is a subset of $W(\Omega; X)$; the subalgebra generated by $X$ is called the $\Omega$-*word algebra* on $X$ and is denoted by $W_\Omega(X)$. Its elements are $\Omega$-*words* in the alphabet $X$. For example, if there is one binary operation $\alpha$, then $x_1 x_2 x_3 \alpha x_4 \alpha\alpha$ is an $\Omega$-word, while $x_1 \alpha\alpha x_2 \alpha x_3$ is an $\Omega$-row which is not an $\Omega$-word.

We shall need a simple test for finding which $\Omega$-rows are words. For this purpose we associate two integers with each $\Omega$-row. The *length* of $w \in W(\Omega; X)$, written $|w|$, is the number of terms in $w$; thus if $w = c_1 \cdots c_N$ where $c_i \in \Omega \cup X$, then $|w| = N$. Secondly we define the *valency* of $w$ as $v(w) = \sum_1^N v(c_i)$, where

$$v(c_i) = \begin{cases} 1 & \text{if } c_i \in X, \\ 1 - n & \text{if } c_i \in \Omega(n). \end{cases}$$

Intuitively the valency represents the 'element-balance': thus if $\omega \in \Omega(n)$, then $\omega$ requires an input of $n$ elements and has an output of one element, so that $v(\omega) = \text{output} - \text{input}$. This idea is exploited in the following result (due to K. Schröter and, independently, P. Hall), which provides a criterion for an $\Omega$-row to be a word:

PROPOSITION 3.1 *An $\Omega$-row $w = c_1 \cdots c_N$ in $X$ is an $\Omega$-word if and only if every left-hand factor $w_i = c_1 \cdots c_i$ of $w$ satisfies*

$$v(w_i) > 0 \qquad (i = 1, 2, \ldots, N), \tag{1}$$

*and*

$$v(w) = 1. \tag{2}$$

*Moreover, each word can be obtained in just one way from its constituents.*

*Proof.* We shall show more generally, by induction on the length $|w|$, that $w$ is a sequence of $r$ words iff (1) holds and

$$v(w) = r. \tag{3}$$

This includes the assertion for $r = 1$. When $|w| = 1$, (1) implies that $v(w) = 1$, so that $w \in X \cup \Omega(0)$, and conversely, so the result holds in this case; we may therefore take $|w| > 1$.

Suppose first that $w$ is an $\Omega$-word, say

$$w = u_1 \cdots u_n \omega, \qquad u_i \in W_\Omega(X), \qquad \omega \in \Omega(n).$$

by the induction hypothesis $v(u_i) = 1$ and $v(\omega) = 1 - n$, so $v(w) = n + 1 - n = 1$. Moreover, every left factor of each $u_i$ has positive valency, hence the same is

true for $w$. When $w$ is a sequence of $r$ words, (1) again holds and (3) follows by addition.

Conversely, let $w$ be an $\Omega$-row satisfying (1) and $|w| > 1$. We can write $w = w'c$, where $c \in \Omega \cup X$ and $v(w') = r' > 0$ by (1), and by induction $w'$ is then a sequence of $r'$ $\Omega$-words. Now either $c \in X \cup \Omega(0)$; then $w$ is a sequence of $r' + 1$ words and $v(w) = r' + 1$; or $c \in \Omega(n)$, where $n > 0$. Then, since $v(w) = r > 0$, we have $r' + 1 - n = r > 0$, hence $n = r' + 1 - r$ and $c$ is applied to the last $n$ words of $w'$ to produce a single word, so $w$ is a sequence of $r' - (n-1) = r$ $\Omega$-words, as we had to show. This analysis of $w$ also shows that it is built up from its constituents in just one way. ■

The uniqueness statement in Prop. 3.1 means that it is never necessary to insert brackets, because each expression is defined unambiguously. To give an example, let $+$ be a binary operation. Then the associative law may be written

$$x_1 x_2 + x_3 + = x_1 x_2 x_3 + +.$$

If $\lambda$ is a second binary operation, then the familiar distributive laws take the form

$$x_1 x_2 + x_3 \lambda = x_1 x_3 \lambda x_2 x_3 \lambda +, \qquad x_1 x_2 x_3 + \lambda = x_1 x_2 \lambda x_1 x_3 \lambda +.$$

Here it is essential to write the operation symbols on one side of the variables, say on the right. Equivalently the operation symbols can all be written on the left (the Łukasiewicz prefix notation). But with the usual infix notation $x_1 + x_2$ an ambiguity arises as soon as we form $x_1 + x_2 + x_3$.

Let $A$ be an $\Omega$-algebra. If in an element $w$ of $W = W_\Omega(X)$ we replace each element of $X$ by an element of $A$ we obtain a unique element of $A$. For $|w| = 1$ this is clear, so assume that $|w| > 1$ and use induction. We have $w = u_1 \cdots u_n \omega$ ($\omega \in \Omega(n)$, $u_i \in W$), where the $u_i$ are uniquely determined once $w$ is given, by Prop. 3.1. By induction each $u_i$ becomes $a_i \in A$ when we replace the elements of $X$ by elements of $A$, hence $w$ becomes $a_1 \cdots a_n \omega$, another element of $A$. This remarks can be used to establish the universal property of the $\Omega$-word algebra:

THEOREM 3.2 *Let $A$ be an $\Omega$-algebra and $X$ a set. Then any mapping $\theta: X \to A$ extends in just one way to a homomorphism $\bar{\theta}: W_\Omega(X) \to A$.*

*Proof.* Every $\Omega$-word is of the form

$$w = c_1 \cdots c_N, \qquad c_i \in \Omega \cup X.$$

We write $w\bar{\theta} = c_1' \cdots c_N'$, where

$$c' = \begin{cases} c & \text{if } c \in \Omega, \\ c\theta & \text{if } c \in X. \end{cases}$$

Thus $w\bar{\theta}$ is just the unique element of $A$ obtained by replacing each $x \in X$ by

$x\theta$. The remark preceding the theorem shows that $\bar{\theta}$ is well-defined, and it is easily seen to be a homomorphism extending $\theta$, which is unique by Prop. 1.1. ■

The content of this theorem is also expressed by saying that $W_\Omega(X)$ is the *free* $\Omega$-algebra on $X$ as a free generating set. Soon we shall meet free algebras in varieties of algebras; the free groups encountered in A.1, **9.9**, free modules of A.1, **10.4** and A.2, **4.4** and free algebras of A.2, **11.5** are examples.

Given any $\Omega$-algebra $A$, we can take a generating set $X$ of $A$ and apply the construction of Th. 3.2. This yields

COROLLARY 3.3 *Any $\Omega$-algebra $A$ can be expressed as a homomorphic image of an $\Omega$-word algebra $W_\Omega(X)$, for a suitable set $X$. Here $X$ can be taken to be any set corresponding to a generating set of $A$.* ■

The $\Omega$-words may also be thought of as operations. Any word in $x_1, \ldots, x_m \in X$ (and elements of $\Omega$) may be regarded as an $m$-ary operation, called a *derived operation*. For example, in groups the commutator $(x, y) = x^{-1}y^{-1}xy$ is a derived operation. The derived operations include the original operations $\omega \in \Omega$, in the form $x_1x_2 \cdots x_n\omega$, as well as $m$ operations $x_i$ $(i = 1, \ldots, m)$. They are the *projection operators*

$$x_1 \cdots x_m\delta_i = x_i. \tag{4}$$

Moreover, we have a composition of operations: if $f_1, \ldots f_n$ are any words in $x_1, \ldots, x_m$ and $g$ is a word in $n$ variables, then $f_1 \cdots f_n g$ is a word in $x_1, \ldots, x_m$ obtained *by composition* from $f_1, \ldots, f_n$, $g$.

On any set $A$ we can consider families of operations which include all projections and are closed under composition. Such a family is called a *clone* of operations on $A$. For example, if $A$ is an $\Omega$-algebra we have the clone generated by $\Omega$; this is the smallest clone including $\Omega$ and is obtained by repeatedly composing the elements of $\Omega$ and the projections.

In studying $\Omega$-algebras we are often not interested in the precise operations $\Omega$, but merely in the clone they generate, and they may be replaced by any other set of operations generating the same clone. For example, groups may be defined in terms of a constant operation $e$ and the single binary operation $xy^{-1}$, or in terms of $e$ and the single ternary operation $xy^{-1}z$, or even as non-empty algebras with the single binary operation $xy^{-1}$, besides the usual ways. This raises the question of finding relatively simple sets of operations.

Consider first unary operations. An operation $\omega: A^n \to A$ will be called *essentially unary* if it depends on at most one argument. More precisely, in terms of projections this means there is a unary operation $f: A \to A$ and $i$, $1 \leqslant i \leqslant n$, such that $\omega = \delta_i f$. For example, each projection operator is essentially unary. It is not hard to verify that the clone generated by any set of essentially unary operations consists entirely of essentially unary operations. For an

$\Omega$-algebra this case arises when $\Omega = \Omega(0) \cup \Omega(1)$. Such algebras can also be characterized by the fact that the union of two subalgebras is again a subalgebra. This shows for example that groups cannot be defined by unary operations alone (since the union of two subgroups is not usually a subgroup).

The distinction between binary and higher operations is much less precise, for as the next result, due to Sierpiński (1945) shows, every finitary operation can be composed from binary ones.

THEOREM 3.4 *Let $A$ be a finite set. Then every finitary operation on $A$ can be expressed in terms of binary operations on $A$.*

*Proof.* Let $|A| = n$; we may without loss of generality regard $A$ as the ring of integers mod $n$, $\mathbf{Z}/n$. The ring operations on $\mathbf{Z}/n$ are at most binary, so it will be enough to show that every operation can be expressed in terms of the ring operations and the $\delta$-function

$$\delta_r(x) = \begin{cases} 1 & \text{if } x = r, \\ 0 & \text{if } x \neq r. \end{cases} \tag{5}$$

This can be accomplished by a multivariable analogue of the Lagrange interpolation formula: given $f(x_1, \ldots, x_k)$, we have

$$f(x_1, \ldots, x_k) = \sum f(a_1, \ldots, a_k)\delta_{a_1}(x_1) \cdots \delta_{a_k}(x_k), \tag{6}$$

where the summation is over all $k$-tuples $(a_1, \ldots, a_k)$. It is of course important to realize that the $a_i$ on the right of (6) are parameters, not variables; thus $ax$, for any $a \in \mathbf{Z}/n$, can be built up from $x$ by repeated addition and so is (for any given $a$) a unary operation. ■

The theorem still holds when $A$ is infinite, but the proof in that case is quite different and is based on the fact that there is then a bijection from $A^2$ to $A$, which can be used to reduce $n$-ary operations to binary ones (cf. Cohn [1981] and Ex. (3)).

When we come to define a concrete class of algebras such as groups, we do so by specifying its operations: $\mu$ binary, $\nu$ unary and $\varepsilon$ 0-ary. the axioms for groups in terms of these operations take the form*:

(associativity) $$x_1x_2\mu x_3\mu = x_1x_2x_3\mu\mu, \tag{7}$$

(neutral) $$x\varepsilon\mu = \varepsilon x\mu = x, \tag{8}$$

(inverse) $$xx\nu\mu = x\nu x\mu = \varepsilon. \tag{9}$$

We see that the axioms take the form of equations holding identically for all

*As stated here, these laws are redundant: parts of (8) and (9) are a consequence of the remainder (as is well known, cf. A.1, 3.2); but this point does not concern us here.

values of the variables. Generally, by an *identity* or *law* over $\Omega$ in $X$ we understand a pair $(u, v) \in W^2$, or sometimes the equation formed from the pair:

$$u = v. \tag{10}$$

We shall say that the law (10) *holds* in the $\Omega$-algebra $A$ or that $A$ *satisfies* (10) if every homomorphism $W \to A$ maps $u$ and $v$ to the same element of $A$, in other words, if $u$ and $v$ define the same derived operation on $A$.

This relation between laws and algebras establishes a Galois connexion between the set of all laws in the given alphabet $X$ and the class of all $\Omega$-algebras. Given any set $\Sigma$ of laws, we can form $\mathscr{V}_\Omega(\Sigma)$, the class of all $\Omega$-algebras satisfying all the laws in $\Sigma$. This class $\mathscr{V}_\Omega(\Sigma)$ is called the *variety* generated by $\Sigma$. For example, groups form the variety of $(\mu, \nu, \varepsilon)$-algebras generated by (7)–(9). Likewise rings form a variety, but fields do not. In the other direction we can from any class $\mathscr{C}$ of $\Omega$-algebras form the set $\mathfrak{q}(\mathscr{C})$ of all laws holding in all algebras of $\mathscr{C}$. Now our Galois connexion relates each variety of $\Omega$-algebras to a correspondence on $W_\Omega(X)$ of the form $\mathfrak{q}(\mathscr{C})$.

Our next task will be to determine the precise form of the set $\mathfrak{q}(\mathscr{C})$. A subalgebra of $A$ is called *fully invariant* if it is mapped into itself by all endomorphisms of $A$; this definition also extends to congruences on $A$, as subalgebras of $A^2$.

THEOREM 3.5 *Let $W = W_\Omega(X)$ be the $\Omega$-word algebra on an infinite alphabet $X$. Then the Galois connexion between $\Omega$-algebras and laws establishes a natural bijection between varieties of $\Omega$-algebras and fully invariant congruences on $W$.*

*Proof.* For any class $\mathscr{C}$ of $\Omega$-algebras let $\mathscr{C}^* = \mathfrak{q}(\mathscr{C})$ be the set of all laws holding in all $\mathscr{C}$-algebras, and for any set $\Sigma$ of laws let $\Sigma^* = \mathscr{V}_\Omega(\Sigma)$ be the variety defined by $\Sigma$. We first show that $\mathscr{C}^*$ is a fully invariant congruence on $W$. The congruence properties are clear: in every $\mathscr{C}$-algebra we have $u = u$ for any $u \in W$; if $u = v$ holds, then so does $v = u$, and if $u = v$, $v = w$ hold, then $u = w$ holds too. Further, if $u_i = v_i (i = 1, \ldots, n)$ are laws holding in $A$ and $\omega \in \Omega(n)$, then

$$u_1 \cdots u_n \omega = v_1 \cdots v_n \omega$$

holds in $A$. Now let $(u, v) \in \mathscr{C}^*$ and let $\theta$ be any endomorphism of $W$. If $\alpha: W \to A$, where $A \in \mathscr{C}$, is any homomorphism, then so is $\theta\alpha$, whence $u\theta\alpha = v\theta\alpha$. Thus the law $u\theta = v\theta$ holds in $A$, so $(u\theta, v\theta) \in \mathscr{C}^*$ and this shows $\mathscr{C}^*$ to be a fully invariant congruence.

To complete the proof we have to show that

$$\mathscr{V}^{**} = \mathscr{V} \tag{11}$$

for any variety $\mathscr{V}$ and

$$\mathfrak{q}^{**} = \mathfrak{q} \tag{12}$$

for any fully invariant congruence $\mathfrak{q}$ on $W$.

By the definition of a variety, $\mathscr{V} = \Sigma^*$ for some $\Sigma \subseteq W^2$, hence $\mathscr{V}^{**} = \Sigma^{***} = \Sigma^* = \mathscr{V}$, which proves (11). To establish (12), we take a fully invariant congruence $\mathfrak{q}$ on $W$ and first show that

$$W/\mathfrak{q} \in \mathfrak{q}^*. \tag{13}$$

This will follow if we can show that all the laws corresponding to the elements of $\mathfrak{q}$ hold in $W/\mathfrak{q}$. Let $(u, v) \in \mathfrak{q}$, let $\alpha: W \to W/\mathfrak{q}$ be any homomorphism and denote the natural homomorphism $W \to W/\mathfrak{q}$ by $\nu$. We shall define an endomorphism $\alpha'$ of $W$ such that

$$w\alpha'\nu = w\alpha \qquad \text{for all } w \in W; \tag{14}$$

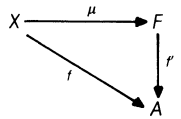
to do so we pick for each $x \in X$ an element $x_0 \in W$ such that $x_0\nu = x\alpha$ and define $x\alpha' = x_0$. By Th. 3.2 the mapping $\alpha': X \to W$ so defined extends to a homomorphism and (14) holds for all $w \in X$; hence by Prop. 1.1 it holds generally.

Now $\mathfrak{q}$ is fully invariant, hence $(u\alpha', v\alpha') \in \mathfrak{q}$ and so $u\alpha = u\alpha'\nu = v\alpha'\nu = v\alpha$, and this establishes (13).

To prove (12), we note that in any case $\mathfrak{q}^{**} \supseteq \mathfrak{q}$. If $(u, v) \notin \mathfrak{q}$, then $u = v$ is not a law in $W/\mathfrak{q}$, because $u \not\equiv v \pmod{\mathfrak{q}}$, but $W/\mathfrak{q} \in \mathfrak{q}^*$ by (13), and so $(u, v) \notin \mathfrak{q}^{**}$. Therefore equality holds in (12). ■

Let $\mathscr{C}$ be a class of $\Omega$-algebras; its members will be called *$\mathscr{C}$-algebras*. By a *free* $\mathscr{C}$-algebra on a set $X$ we understand an algebra $F$ in $\mathscr{C}$ with the following universal property: there is a mapping $\mu: X \to F$ such that every mapping $f: X \to A$ into a $\mathscr{C}$-algebra $A$ can be factored uniquely by $\mu$ to give a homomorphism from $F$ to $A$, i.e. there exists a unique homomorphism $f': F \to A$ such that

$$\mu f' = f. \tag{15}$$




*Remarks*

1. If $\mathscr{C}$ contains non-trivial algebras, then $\mu$ is an embedding. For given $a, b \in X$, $a \neq b$, we can map $X$ to a $\mathscr{C}$-algebra in such a way that $af \neq bf$; hence by (15), $a\mu \neq b\mu$.

2. If $\mathscr{C}$ admits subalgebras, then $F$ is generated by the image $X\mu$. For otherwise we could replace $F$ by the subalgebra generated by $X\mu$; since $F$ is unique up to isomorphism, it must itself be generated by $X\mu$.

3. If $\mathscr{C}$ admits subalgebras, $F$ is a free $\mathscr{C}$-algebra on $X$ and $X'$ is any subset

of $X$, then the subalgebra of $F$ generated by the image of $X'$ is the free $\mathscr{C}$-algebra on $X'$. For this subalgebra is easily seen to possess the universal property.

Not every class has free algebras, but they exist in varieties, by our next result.

PROPOSITION 3.6 *Let $\mathscr{V}$ be a variety of $\Omega$-algebras and $\mathfrak{q}$ the congruence on $W = W_\Omega(X)$ consisting of all the laws on $\mathscr{V}$. Then $W/\mathfrak{q}$ is the free $\mathscr{V}$-algebra on $X$.*

*Proof.* By (13), $W/\mathfrak{q}$ is a $\mathscr{V}$-algebra, so it only remains to verify the universal property. Let us write $v: W \to W/\mathfrak{q}$ for the natural mapping. Given any mapping $f: X \to A$ to a $\mathscr{V}$-algebra, by Th. 3.2 this extends to a homomorphism $\bar{f}: W \to A$. Given $u, v \in W$, if $u \equiv v \pmod{\mathfrak{q}}$, then $(u, v)$ is a law in $\mathscr{V}$ and so holds in $A$, hence $u\bar{f} = v\bar{f}$. Thus $\mathfrak{q} \subseteq \ker \bar{f}$, and by the factor theorem there is a homomorphism $f': W/\mathfrak{q} \to A$ such that $\bar{f} = vf'$. If $\mu: X \to W$ is the injection, we have $f = \mu\bar{f} = \mu v f'$, and $f'$ is unique, since it is given on a generating set of $W/\mathfrak{q}$. Thus $W/\mathfrak{q}$ satisfies all the conditions for a free $\mathscr{V}$-algebra. ■

There is another way of forming free algebras, which leads to a useful criterion, due to G. Birkhoff, for a class of algebras to be a variety.

THEOREM 3.7 *Let $\mathscr{C}$ be a class of $\Omega$-algebras. Then $\mathscr{C}$ is a variety if and only if it is closed under the operations of taking subalgebras, homomorphic images and direct products.*

*Proof.* The necessity of the conditions is easy to check: given any $\Omega$-algebra $A$, it is clear that any subalgebra and any homomorphic image of $A$ satisfies all the laws holding in $A$. Moreover, if a law holds in every member of a family $(A_i)$ of $\Omega$-algebras, then it also holds in their direct product. This shows that every variety satisfies the given conditions.

Conversely, let $\mathscr{C}$ be a class of $\Omega$-algebras closed under subalgebras, homomorphic images and direct products. Then $\mathscr{C}$ contains the trivial algebra (as the direct product of the empty family). If there are no other algebras in $\mathscr{C}$, then we have the variety defined by the law $x_1 = x_2$. So we may now assume that $\mathscr{C}$ contains a non-trivial algebra. We can form a free $\mathscr{C}$-algebra on a given set $X$ as follows. Consider the family $(A_i)$ of all $\mathscr{C}$-algebras with a generating set of cardinal not exceeding that of $X$. Take all mappings $f_\alpha: X \to A_\alpha$ where $A_\alpha$ is a $\mathscr{C}$-algebra and $Xf_\alpha$ a generating set of $A_\alpha$, and in the direct product $P = \prod A_\alpha$ consider the subalgebra $F$ generated by the elements $(xf_\alpha)$, $x \in X$. As a subalgebra of the direct product, $F$ is again in $\mathscr{C}$. We claim that $F$ satisfies the universal property relative to the mapping $\mu: x \mapsto (xf_\alpha)$. For if $f: X \to A$ is any mapping to a $\mathscr{C}$-algebra $A$, and $A'$ is the subalgebra generated by $Xf$, then the mapping $f$ cut down to $A'$ coincides with some $f_\alpha$ and so $A'$ is a homomorphic image of

$F$, the mapping $F \to A'$ being the projection on the appropriate factor. Thus we have a homomorphism $f' : F \to A$ such that $f = \mu f'$ and $f'$ is unique since it is prescribed on a generating set of $F$. Thus $F$ is the free $\mathscr{C}$-algebra on $X$.

Clearly we have

$$\mathscr{C} \subseteq \mathscr{C}^{**}, \tag{16}$$

and it remains to prove equality. Let $\mathfrak{q} = \mathscr{C}^*$ be the set of all laws holding in $\mathscr{C}$. By Prop. 3.6, the free $\mathscr{C}$-algebra is $W/\mathfrak{q}$. If $A \in \mathscr{C}^{**}$, we can write $A$ as a homomorphic image of $W$, for an appropriate $X$, say $f : W \to A$. By definition of $\mathscr{C}^{**} = \mathfrak{q}^*$, $A$ satisfies all the laws of $\mathfrak{q}$, hence $f$ can be factored by $\mathfrak{q}$; thus $A$ is a homomorphic image of $W/\mathfrak{q}$, the free $\mathscr{C}$-algebra on $X$, and $A$ is therefore itself a $\mathscr{C}$-algebra. Hence equality in (16) is established. ■

We have already remarked that rings and groups are examples of varieties. We now see that fields (commutative or not) do not form a variety, since they do not admit direct products; for if $E, F$ are any fields, their direct product $E \times F$ as a ring has zero-divisors and so cannot be a field (cf. **1.6**).

## Exercises

(1) Show that if some operation symbols are written on the left and others on the right of their arguments, then ambiguities can arise.

(2) Let $w = c_1 \cdots c_N$ be an $\Omega$-word of the form $u_1 \cdots u_n \omega \, (u_i \in W, \omega \in \Omega(n))$. Show that any proper subsequence $c_i c_{i+1} \cdots c_j$, where $j - i < N - 1$, which is itself an $\Omega$-word, can only occur within a single factor $u_k$.

(3) Assuming a bijection $\mu : A^2 \leftrightarrow A$ between a set $A$ and its Cartesian square $A^2$, show that every $n$-ary operation $\omega$ on $A$ can be expressed in terms of the binary operation $\mu$ and a suitable unary operation.

(4) Verify that the set of all essentially unary operations on a set is a clone. Deduce that any operation derived from essentially unary operations is again essentially unary.

(5) Let $A$ be any $\Omega$-algebra, $X$ a set and for each $x \in X$, let $\delta_x : A^X \to A$ be the projection on the $x$th factor. Show that the subalgebra of the direct power $A^{A^X}$ generated by all the $\delta_x$ $(x \in X)$ is the free algebra on $X$ for the variety generated by the algebra $A$ (i.e. the least variety containing $A$).

(6) Show that modular lattices form a variety. Similarly for distributive lattices, and Boolean algebras.

(7) Show that groups may be defined in terms of the operation $xy\alpha = xy^{-1}$ as non-empty algebras satisfying $xz\alpha yz\alpha\alpha = xy\alpha$, $xx\alpha yy\alpha y\alpha\alpha = y$. Show that abelian groups may be defined by $xxy\alpha\alpha = y$, $xy\alpha z\alpha = xz\alpha y\alpha$.

(8) Show that any variety of groups defined by a finite set of laws can also be defined by a single law.

(9) Show that the automorphism group of $W_\Omega(X)$ is isomorphic to the group of all permutations of $X$.

(10) Let $\mathscr{V}$ be a variety of $\Omega$-algebras and $F$ the free $\mathscr{V}$-algebra on a set $X$. Given a homomorphism $f: A \to B$ between $\mathscr{V}$-algebras which is surjective, and a homomorphism $\theta: F \to B$, find a homomorphism $\theta': F \to A$ such that $\theta = \theta' f$. (Hint: see the proof of Th. 3.5.)

## 1.4 Abstract dependence relations

The notion of linear dependence in a vector space has certain properties which we shall meet again later when we come to discuss algebraic dependence in **5.1**. These concepts also occur elsewhere e.g. in some geometrical constructions, and it is convenient to single them out here and discuss their properties in general.

Let $S$ be a set; by a *dependence relation* on $S$ we understand a rule which associates with each finite subset $X$ of $S$ certain elements of $S$, said to be *dependent on* $X$. Further, the following conditions are satisfied:

**D.0** *If* $X = \{x_1, \ldots, x_n\}$, *then each* $x_i$ *is dependent on* $X$.

**D.1** (*Transitivity*) *If* $z$ *is dependent on* $Y$ *and each member of* $Y$ *is dependent on* $X$, *then* $z$ *is dependent on* $X$.

**D.2** (*Exchange property*) *If* $y$ *is dependent on* $\{x_1, \ldots, x_n\}$, *but not on* $\{x_2, \ldots, x_n\}$, *then* $x_1$ *is dependent on* $\{y, x_2, \ldots, x_n\}$.

We note that by **D.**0, 1, if $y$ is dependent on $X'$ and $X' \subseteq X$, then $y$ is dependent on $X$. The notion of dependence can be extended to infinite subsets by saying that $y$ is dependent on an infinite subset $X$ of $S$ if there is a finite subset $X'$ of $X$ such that $y$ is dependent on $X'$. Then it is easily seen that if the exchange property **D.**2 holds for finite subsets, it also holds for infinite subsets. Similarly **D.**0, 1 hold for arbitrary subsets.

As an example let us consider linear dependence in a vector space over a field or, more generally , in a module $M$ over a ring $R$. Given a left $R$-module $M$ and a subset $X$ of $M$, an element $y$ of $M$ is said to be *linearly dependent* on the elements of $X$ if

$$y = \sum c_i x_i \qquad \text{for some } c_i \in R, \text{ where } x_i \in X. \tag{1}$$

The element $y$ is said to satisfy a *non-trivial relation* with the elements of $X$ if

$$by + \sum a_i x_i = 0 \qquad \text{for some } a_i, b \in R,\ b \neq 0, \text{ where } x_i \in X. \tag{2}$$

It is clear that (1) implies (2) (with $b = 1$, $a_i = -c_i$). Conversely, if (2) holds with an invertible element $b$, then (1) follows by taking $c_i = -b^{-1}a_i$. This shows in particular that (1), (2) are equivalent for vector spaces over a field.

Now **D.**0–2 are easily verified for linear dependence as defined by (1) or (2) over a field. We prove **D.**2 as an example. If $y$ is dependent on $x_1, \ldots, x_n$ then we have a relation (2), and if $y$ is not dependent on $x_1, \ldots, x_n$, then $a_1$ cannot vanish, so (2) shows $x_1$ to be dependent on $y, x_2, \ldots, x_n$.

For modules over a ring, (1), (2) are in general no longer equivalent. The relation (1) always satisfies **D.**0, 1 but not always **D.**2. For example, in $\mathbf{Z}^2$ (as $\mathbf{Z}$-module) $(2,2)$ depends on $(1,0)$ and $(0,1)$ but not on $(0,1)$ alone, yet $(1,0)$ does not depend on $(2,2)$ and $(0,1)$, i.e. there is no relation $(1,0) = m(2,2) + n(0,1)$, $m, n \in \mathbf{Z}$. Similarly the relation (2) satisfies **D.**0, 2 but not necessarily **D.**1 (cf. Ex. (2)).

Given a dependence relation on a set $S$, we define for each $X \subseteq S$ the *span* of $X$ as the set $\langle X \rangle$ of all elements of $S$ dependent on $X$; we also say that $\langle X \rangle$ is *spanned* by $X$. From **D.**0 we see that $\langle X \rangle \supseteq X$ and **D.**1 shows that every element dependent on $\langle X \rangle$ itself belongs to $\langle X \rangle$.

A family $X = \{x_i | i \in I\}$ of elements of $S$ is said to be *independent* if no $x_i$ is dependent on $\{x_j | j \neq i\}$; otherwise $X$ is called *dependent*. Clearly an independent family consists of distinct elements. Moreover, being independent is a property of finite character, for a family is dependent whenever this is true of some finite subfamily. An independent family which spans $S$ is called a *basis* of $S$. This of course generalizes the definitions given in (A.1, **4.1**) for vector spaces, and as in that case, we can recognize bases in the following way:

PROPOSITION 4.1 *Let $S$ be a set with a dependence relation. Then for any subset $X$ of $S$ the following three conditions are equivalent:*

(a) *$X$ is a maximal independent subset of $S$,*
(b) *$X$ is a minimal spanning set,*
(c) *$X$ is a basis of $S$.*

*Proof.* (a)⇔(c). Let $X$ be a maximal independent subset of $S$; we have to show that every element of $S$ is dependent on $X$. Given $y \in S$, if $y \in X$, then $y$ is dependent on $X$ by **D.**0, and if $y \notin X$, then $X \cup \{y\}$ is dependent, by maximality, so some element is dependent on the rest, say $x \in X$ is dependent on $X' \cup \{y\}$, where $X' = X \backslash \{x\}$. Since $X$ is independent, $x$ is not dependent on $X'$, hence by **D.**2, $y$ is dependent on $X' \cup \{x\} = X$, and this shows that $X$ spans $S$. Conversely, if $X$ is a basis, it is independent, but every element not in $X$ is dependent on $X$, so $X$ is maximal independent.

(b)⇔(c). Let $X$ be a minimal spanning set; we have to show that $X$ is independent. If $X$ were dependent, say $x \in X$ is dependent on the rest of $X$, then we could omit $x$ from $X$ and still have a spanning set, by **D.**1. This contradicts the minimality, hence $X$ is a basis. Conversely, if $X$ is a basis, it is independent, so no element is dependent on the rest, and $X$ is a minimal spanning set. ■

It is instructive to go through this proof in a particular case, and the reader is advised to do this, say for vector spaces over a skew field, taking first the case where the basis is finite and then the general case.

Our main aim will be to show that every set with a dependence relation has a basis, and that any two bases have the same number of elements. This will again show that a vector space, even infinite-dimensional, has a basis, and a uniquely defined dimension. We first use **D.**2 to derive a form of the exchange lemma (cf. A.1, **4.3**).

LEMMA 4.2 (Exchange lemma) *Let $S$ be a set with a dependence relation. If $X$ is an independent subset and $Y$ a spanning set of $S$, then we can complete $X$ to a basis of $S$ by elements from $Y$, i.e. there is a subset $Y'$ of $Y$ such that $X \cap Y' = \varnothing$ and $X \cup Y'$ is a basis of $S$. Moreover, if $Y$ is finite, then $X$ is finite and*

$$|X \cup Y'| \leqslant |Y|. \tag{3}$$

*Proof.* Consider the collection $\mathscr{I}$ of all independent subsets $Z$ of $S$ such that

$$X \subseteq Z \subseteq X \cup Y. \tag{4}$$

Since independence is a property of finite character, $\mathscr{I}$ is inductive, and so by Zorn's lemma $\mathscr{I}$ has a maximal member $B$, i.e. a maximal independent set $B$ such that $X \subseteq B \subseteq X \cup Y$. This shows that $B$ consists of $X$ together with some elements of $Y$, thus $B = X \cup Y'$; by omitting from $Y'$ elements that are also in $X$ we ensure that $X \cap Y' = \varnothing$. Now every element of $Y$ depends on $B$, by maximality, hence $\langle B \rangle \supseteq \langle Y \rangle = S$, therefore $B$ is a basis, as claimed.

Assume now that $Y$ is finite. We shall show that starting from any finite subset $X'$ of $X$ we can obtain a basis by adjoining at most $|Y| - |X'|$ elements of $Y$. This will show that $|X'| \leqslant |Y|$ and for $X' = X$ we reach the desired conclusion (3). We shall use induction on $|X'|$; when $|X'| = 0$, we can by the first part choose a subset of $Y$ which is a basis. Now suppose that $|X'| = r > 0$ and that we have a basis $B_{r-1} = \{x_1, \ldots, x_{r-1}, y_r, \ldots, y_n\}$, where $x_i \in X'$, $y_j \in Y$. With another element $x_r$ of $X'$ we form the spanning set $B_{r-1} \cup \{x_r\}$ which contains the independent subset $\{x_1, \ldots, x_r\}$. By the first part it contains a basis $B_r$ including $x_1, \ldots, x_r$, but $B_r$ cannot include all of $y_r, \ldots, y_n$ because the family $\{x_1, \ldots, x_r, y_r, \ldots, y_n\}$ is dependent ($x_r$ is dependent on the rest); hence $B_r$ has the required form and it contains at most $n$ elements. By induction we obtain a basis of at most $|Y|$ elements, consisting of $X'$ together with some elements of $Y$. In particular, $|X'| \leqslant |Y|$ and since $X'$ was any finite subset of $X$, it follows that $X$ is finite and if the basis is $X \cup Y'$, then $|X \cup Y'| \leqslant |Y|$. ■

By the last part, the number of elements in an independent family is bounded by the number of elements in a finite spanning set. In particular, given two bases $B$, $C$, if $B$ is finite, then $|C| \leqslant |B|$ and so by symmetry, $|B| = |C|$. Thus we have

COROLLARY 4.3 *Let $S$ be a set with a dependence relation. If $S$ has a finite spanning set $C$, then any independent family has at most $|C|$ members and any two bases have the same number of elements.* ■

Our next task is to show that any two bases, finite or not, have the same cardinal. For the case of infinite bases this is true in a rather wider context, which we shall now explain, as it is sometimes useful.

Suppose that instead of a dependence relation we have a relation satisfying only **D.**0 and **D.**1 (not necessarily **D.**2); let us call this a *spanning relation.* We can as before define the *span* of a subset $X$ as the set $\langle X \rangle$ of all elements dependent on $X$. If we examine the proof of Prop. 4.1 we find that for a spanning relation (c)$\Leftrightarrow$(b)$\Rightarrow$(a), but a maximal independent set need no longer be a basis. In fact the term 'basis' in this context tends to mislead and we shall only speak of minimal spanning sets. The example of the rational numbers $\mathbf{Q}$ (as $\mathbf{Z}$-module) shows that minimal spanning sets need not exist. But when they do, we have the following relation, generalizing the situation for modules (cf. A.2, Prop. 4.4.4).

PROPOSITION 4.4 *Let $S$ be a set with a spanning relation and let $X$ be a minimal spanning set of $S$. If $X$ is infinite, of cardinal $\alpha$, then any spanning set of $S$ has at least $\alpha$ elements. In particular, $S$ has no finite spanning set and any two minimal spanning sets of $S$ have the same cardinal.*

The proof is quite similar to the module case and again uses the relation $\aleph_0\beta = \beta$ for infinite cardinals $\beta$ (A.2, Prop. 1.2.6). Let $Y$ be any spanning set of $S$, of cardinal $\beta$. Each $y \in Y$ is dependent on $X$ and hence is dependent on a finite subset $X_y$ of $X$. We claim that

$$X = \bigcup_{y \in Y} X_y. \tag{5}$$

For the span of the right-hand side contains each $y \in Y$ and so must be $S$. Thus $\cup X_y$ is a subset of $X$ spanning $S$, and (5) follows by the minimality of $X$. If $Y$ were finite, then (5) would express $X$ as a finite union of finite sets, but this contradicts the fact that $X$ is infinite. We thus have

$$\alpha = |X| \leqslant \sum |X_y| \leqslant \aleph_0 \beta = \beta.$$

Hence $\alpha \leqslant \beta$; if $Y$ is also minimal, we can interchange the roles of $X$ and $Y$ and so conclude that $\alpha = \beta$. ■

This applies in particular to modules, since as we observed earlier, the relation '$y$ lies in the submodule generated by $X$' is a spanning relation. We thus obtain Prop 4.4.4 of A.2 as a special case.

Applying Prop. 4.4 to the case of dependence relations, we obtain

THEOREM 4.5 *Let $S$ be any set with a dependence relation. Then $S$ has a basis; more precisely, if $X$ is an independent subset and $Y$ is a spanning set of $S$ such that $X \subseteq Y$, then there is a basis $B$ such that*

$$X \subseteq B \subseteq Y, \tag{6}$$

*and any two bases have the same cardinal.*

*Proof.* Let $X, Y$ be as stated; then $X \cup Y = Y$ and by Lemma 4.2 there exists a basis satisfying (6). In particular, taking $X = \varnothing$, $Y = S$, we see that there always is a basis. The last part follows by Cor. 4.3 and Prop. 4.4. ■

As a consequence we see again that every vector space has a basis (cf. A.2, **4.2**), and any two bases have the same cardinal. For example, **R** considered as a **Q**-module has a basis, but the proof (involving Zorn's lemma) is non-constructive, and no explicit construction of such a basis is known. Any **Q**-basis of **R** is called a *Hamel basis* of the real numbers.

Occasionally we shall have to deal with a relation satisfying **D.**0, 2 but only a weakened form of **D.**1 obtained by assuming $Y$ independent. Thus our relation satisfies

**D.**1′ *If $z$ is dependent on the independent set $Y$ and each element of $Y$ is dependent on $X$, then $z$ is dependent on $X$.*

We still find that bases are just maximal independent sets, for the proof (a)⇔(c) Prop. 4.1 did not use **D.**1, and maximal independent sets always exist, by Zorn's lemma, so every independent subset is contained in a basis. Now the proof of Lemma 4.2, in the case where $X, Y$ are finite bases, uses **D.**1 only in the weaker form **D.**1′, so we still have the following conclusion:

COROLLARY 4.6 *Let $S$ be a set with a relation satisfying* **D.**0, 1′, 2. *Then any independent subset is contained in a basis, and if $S$ has a finite basis, any two bases have the same cardinal.* ■

The axiomatic study of dependence relations turns out to have connexions with many other fields, particularly graph theory, combinatorial theory, block designs etc. See D. J. A. Welsh [1976], N. White [1986].

## Exercises

(1) Prove the exchange property for infinite subsets: If $y$ is dependent on $X \cup \{z\}$ but not on $X$, then $z$ is dependent on $X \cup \{y\}$.

(2) Verify that the relation defined by (2) over a commutative integral domain satisfies **D.**0–2. Give an example of a commutative ring with zero-divisors for which **D.**1 fails.

(3) Check that in the proof of Prop. 4.1, (c)⇒(a), (b) only uses **D**.0, (a)⇒(c) only uses **D**.0, 2 and (b)⇒(c) only uses **D**.0, 1. Give examples to show that the axioms listed here cannot be omitted.

(4) Let $S$ be a set with a dependence relation and for each subset $X$ write $r(X)$ for the cardinal of a basis of $X$. Show that (i) $r(\varnothing) = 0$, (ii) $r(X) \leqslant r(X \cup \{y\}) \leqslant r(X) + 1$, (iii) if $r(X) = r(X \cup \{y\}) = r(X \cup \{z\})$, then $r(X \cup \{y, z\}) = r(X)$.

Show that any function $r(X)$ on the subsets of a set $S$ with integer values, satisfying (i)–(iii) leads to a dependence relation on $S$.

(5) Show that for any $n \geqslant 1$ there is a minimal spanning set for **Z** (as **Z**-module) of $n$ elements.

(6) How many dependence relations are there on a set of three elements?

(7) Two dependence relations on a set $S$ are called *equivalent* if we can pass from one to the other by a permutation of $S$. Show that there are three equivalence classes of dependence relations on a set of three elements. How many classes are there on a four-element set?

(8) Show that any dependence relation on a set $S$ is determined by the collection of bases of $S$.

(9) Show that any dependence relation on a set $S$ is determined by the collection of all minimal dependent subsets.

(10) Show that every additive mapping of **R** into itself is **Q**-linear, but not necessarily **R**-linear. (Hint: use a Hamel basis.)

## 1.5 The diamond lemma

In many algebraic problems the elements of a set are defined as equivalence classes of formal expressions, where two expressions are considered equivalent if one can pass from one to the other by a series of 'moves'. The problem is to decide when two given expressions are equivalent. For example, the elements of a particular $\Omega$-algebra $A$ are given by $\Omega$-words in a generating set; the defining relations in $A$ allow the passage between certain words and we have to decide when two given words represent the same element of $A$ (the *word problem* for $A$).

The situation may be represented by a graph as follows: the vertices of our graph are the different formal expressions, and each move, from $u$ to $v$ say, is represented by an edge from $u$ to $v$. Now the equivalence classes are the connected components of our graph. Frequently the moves are of two sorts: direct moves (e.g. in a group, removing a factor $xx^{-1}$) and their inverses (inserting a factor $xx^{-1}$ in a certain place); this means that we have a directed graph. An expression is *reduced* if it admits no direct moves and the main result of this

section, the *diamond lemma*, gives conditions under which each equivalence class contains a single reduced expression. The conditions are of a form that frequently applies, and it leads to a simple solution of our problem: To test if two expressions are equivalent we apply direct moves until each is in reduced form; if these reduced forms are equal, then and only then are the two expressions equivalent.

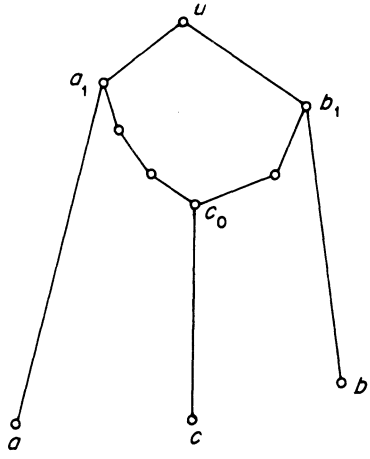
LEMMA 5.1 (Diamond lemma, M. H. A. Newman 1942) *Let $A$ be a set with an equivalence relation defined on it by moves as above, such that the following conditions are satisfied.*

(i) *Finiteness condition. For each $u \in A$ there exists an integer $r = r(u)$ such that no chain of direct moves applied to $u$ has more than $r$ terms.*

(ii) *Confluence condition. If $u \in A$ can be transformed to $x$ by one direct move and to $y$ by another, then there exists $v \in A$ which can be reached from each of $x$, $y$ by an appropriate series of direct moves.*

*Then each equivalence class of $A$ contains exactly one reduced element.*

*Proof.* By (i) we can from each element of $A$ reach a reduced element by a




finite series of direct moves, so each equivalence class contains a reduced element. Given $u \in A$, suppose that we reach a reduced element $a$ in $m$ moves, passing through the elements $u = a_0, a_1, \ldots, a_m = a$ and that $u = b_0, b_1, \ldots, b_n = b$ is another such chain leading to a reduced element $b$ in $n$ steps; we have to show that $a = b$. By (ii) we can reach a common element $c_0$ by direct moves applied to $a_1$ or to $b_1$, and by direct moves applied to $c_0$ reach a reduced element $c$. We shall use induction on the least value of $r(u)$. Clearly $r(a_1) < r(u)$, hence by induction we have $c = a$, and similarly $c = b$, therefore $a = b$, as claimed. ■

A typical application is the existence proof of a normal form for the elements

of a free group (cf. Ex. (4) and Ch. 4). For a discussion of the applications to rings, with many illuminating examples, see Bergman (1978).

### Exercises

(1) In Lemma 5.1 (ii) assume that if $u$ is transformed to $x$ by one direct move and to $y$ by another, where $x \neq y$, then there exists $v \in A$ which can be reached from each of $x, y$ by just one direct move. Show that all reduction chains from a given element to a reduced element have the same length. Show that the extra condition cannot be omitted.

(2) (M. H. A. Newman) Show that the conclusion of Lemma 5.1 still applies if (ii) holds but instead of (i) we have merely the minimum condition: no element admits an infinite succession of direct moves. (Hint. repeat the construction in the proof of Lemma 5.1 and use the minimum condition.)

(3) Let $A$ be a ring with an endomorphism $\alpha$. Show that in the ring $R$ generated by $A$ and a symbol $x$ satisfying $ax = xa^{\alpha}$ for all $a \in A$, every element can be uniquely written as a 'polynomial' in $x$: $\sum_0^n x^i a_i (a_i \in A)$, and hence that $A$ is embedded in $R$ (cf. **9.3**).

(4) Let $X$, $X'$ be disjoint sets with a bijection $x \leftrightarrow x'$ between them. Write $Z = X \cup X'$ and on the set $Z^*$ of all strings of letters from $Z$ (including the empty string 1) define a product by juxtaposition (this is just the free monoid on $Z$, cf. A.2, **11.6**). Define direct moves as the replacement of $fxx'g$ or $fx'xg$ by $fg$ and their inverses as inverse moves. Show that the diamond lemma applies and deduce the existence of a normal form for the elements of a free group (cf. **4.5**).

(5) Let $S$ be a semigroup without idempotent (i.e. $x^2 \neq x$ for all $x \in S$) and satisfying

$$ua = ub \Rightarrow va = vb \qquad \text{for all } u, v, a, b \in S. \tag{1}$$

By adjoining formal solutions of the equations

$$xa = b, \qquad a, b \in S, \tag{2}$$

show that $S$ can be embedded in a semigroup in which (2) has a solution for all $a, b$. (Hint: adjoin a new symbol $p$ to $S$ and consider all words in $S \cup \{p\}$ with direct move $pa \to b$. Verify the conditions of Lemma 5.1 and show that distinct elements of $S$ cannot be equivalent. Now show that the resulting semigroup again satisfies (1) and repeat the process.)

## 1.6 Ultraproducts

Let us again consider the direct product construction. Given a direct product $P = \prod A_i$ of $\Omega$-algebras, we have seen that if a law $u = v$ holds in each factor $A_i$ then it holds in the product. On the other hand, consider the statement occurring in the definition of a field:

$$\text{for all } a \neq 0 \text{ there exists } a' \text{ such that } aa' = 1. \tag{1}$$

This may well hold in each factor $A_i$ and yet fail to hold in the direct product, as is seen by taking the direct product of two fields: the element $(1, 0)$ is different from 0 but does not have an inverse.

In order to remedy the situation we introduce certain homomorphic images of direct products, called ultraproducts, which have the property that every sentence of first-order logic which holds in all the factors, also holds in the product. For a complete proof we would need a detailed description of what constitutes a sentence in first-order logic (an 'elementary sentence') and this would take us rather far afield. However, the construction itself is easily explained and has many uses in algebra. We describe it below and refer for further details to Bell and Slomson [1971], Barwise [1977] and Cohn [1981].

We shall need the concept of an ultrafilter. Let $I$ be a non-empty set. By a *filter* on $I$ one understands a collection $\mathscr{F}$ of subsets of $I$ such that

**F.1** $I \in \mathscr{F}$, $\varnothing \notin \mathscr{F}$,
**F.2** *if* $X, Y \in \mathscr{F}$, *then* $X \cap Y \in \mathscr{F}$,
**F.3** *if* $X \in \mathscr{F}$ *and* $X \subseteq X' \subseteq I$, *then* $X' \in \mathscr{F}$.

For example, given any subset $A$ of $I$, if $A \neq \varnothing$, then the set of all subsets of $I$ containing $A$ is a filter, called the *principal* filter generated by $A$. More generally, if $(A_\lambda)$ is any family of subsets of $I$, then the collection of all subsets containing a set of the form

$$A_{\lambda_1} \cap \cdots \cap A_{\lambda_r} \tag{2}$$

forms a filter, provided that none of the sets (2) is empty. This condition on the $A_\lambda$ is called the *finite intersection property*. Thus any family of subsets of $I$ with the finite intersection property is contained in a filter on $I$. Such a family is also called a *filter base*.

An *ultrafilter* on $I$ is a filter which is maximal among all the filters on $I$. An alternative characterization is given by

LEMMA 6.1 *A filter $\mathscr{F}$ on $I$ is an ultrafilter if and only if for each subset $A$ of $I$, either $A$ or its complement $A'$ belongs to $\mathscr{F}$.*

Of course, $A, A'$ cannot both belong to $\mathscr{F}$, because then $\mathscr{F}$ would contain $\varnothing = A \cap A'$.

*Proof.* Let $\mathscr{F}$ be an ultrafilter. If $A \notin \mathscr{F}$, then by **F.3**, no member of $\mathscr{F}$ can be contained in $A$ and so each member of $\mathscr{F}$ meets $A'$. It follows that the family of all sets containing some $F \cap A'$ $(F \in \mathscr{F})$ is a filter containing $\mathscr{F}$. By maximality it equals $\mathscr{F}$ and so $A' \in \mathscr{F}$. Conversely, if for each $A \subseteq I$, either $A$ or $A'$ belongs to $\mathscr{F}$, consider a filter $\mathscr{F}_1 \supset \mathscr{F}$, $B \in \mathscr{F}_1 \setminus \mathscr{F}$. By assumption, $B' \in \mathscr{F}$ and so $\varnothing = B \cap B' \in \mathscr{F}_1$ which is a contradiction. ■

The existence of ultrafilters is clear from Zorn's lemma:

THEOREM 6.2 *Every filter on a set $I$ is contained in an ultrafilter.*

*Proof.* Let $\mathscr{F}$ be a filter and consider the set of all filters containing $\mathscr{F}$. This set is easily seen to be inductive, hence it has a maximal member, which is the required ultrafilter. ■

For example, the principal filter generated by a one-element subset is an ultrafilter. When $I$ is finite, every ultrafilter is of this form, but for infinite sets we can always find non-principal ultrafilters as follows. Let $I$ be an infinite set. The *cofinite* subsets of $I$, i.e. the subsets with a finite complement, clearly have the finite intersection property and are therefore contained in an ultrafilter, which is easily seen to be non-principal.

We shall use filters to construct certain homomorphic images of direct products. Let $A_i (i \in I)$ be a family of $\Omega$-algebras and let $\mathscr{F}$ be any filter on the index set $I$. Then the *reduced product*

$$\prod A_i/\mathscr{F} \tag{3}$$

is the homomorphic image of the direct product $P = \prod A_i$ defined by the rule: For any $x, y \in P$,

$$x \equiv y \Leftrightarrow \{i \in I \mid x\pi_i = y\pi_i\} \in \mathscr{F}, \tag{4}$$

where $\pi_i$ is the natural projection on $A_i$. Let us call a subset of $I$ $\mathscr{F}$*-large* if it belongs to $\mathscr{F}$. Then the definition states that $x \equiv y$ iff $x$ and $y$ agree on an $\mathscr{F}$-large set. We have to verify that we obtain an $\Omega$-algebra in this way, i.e. that the correspondence defined on $P$ by (4) is a congruence. Reflexivity and symmetry are clear and transitivity follows by F.2. Now take $\omega \in \Omega(n)$ and let $x_\nu = y_\nu (\nu = 1, \ldots, n)$, say $x_\nu$ and $y_\nu$ agree on $A_\nu \in \mathscr{F}$. Then $A_1 \cap \cdots \cap A_n \in \mathscr{F}$ and on this set $x_1 \cdots x_n \omega$ and $y_1 \cdots y_n \omega$ agree. Thus we have

THEOREM 6.3 *A reduced product of $\Omega$-algebras is an $\Omega$-algebra.* ■

More generally, this holds for any $\mathscr{V}$-algebras, where $\mathscr{V}$ is a variety, because the reduced product is a homomorphic image of a product.

A reduced product formed with an ultrafilter is called an *ultraproduct*, or *ultrapower* if all factors are the same. Now the ultraproduct theorem for $\Omega$-algebras asserts that an ultraproduct $\prod A_i/\mathscr{F}$ of $\Omega$-algebras formed with an ultrafilter $\mathscr{F}$, is again an $\Omega$-algebra such that any elementary sentence holds in the ultraproduct precisely if it holds in each factor $A_i$ for $i$ in an $\mathscr{F}$-large set. As already indicated, we shall not prove the full form, but merely a special case, which illustrates the method and which is sufficient for our purposes.

THEOREM 6.4 *An ultraproduct of skew fields is a skew field.*

*Proof.* Let $D_i (i \in I)$ be a family of skew fields and $K = \prod D_i/\mathscr{F}$ their ultraproduct, formed with an ultrafilter $\mathscr{F}$ on $I$. By Th. 6.3 and the remark following it, $K$ is a ring. Let $a \in K$ and suppose that $a \neq 0$. Taking a representative $(a_i)$ for $a$ in $\prod D_i$, we have

$$J = \{i \in I \mid a_i = 0\} \notin \mathscr{F}, \tag{5}$$

because $a \neq 0$. Therefore $J' \in \mathscr{F}$ by Lemma 6.1, and we can define $(b_i)$ by

$$b_i = \begin{cases} a_i^{-1} & \text{if } i \in J' \\ 0 & \text{if } i \in J. \end{cases}$$

Let us denote the image of $(b_i)$ in $K$ by $b$. Since $a_i b_i = b_i a_i = 1$ for $i \in J'$ and $J' \in \mathscr{F}$, we find that $ab = ba = 1$. Hence every non-zero element of $K$ has an inverse and so $K$ is a skew field, as claimed. ■

It is instructive to take a reduced product and see where the proof fails: it was for (5) that we needed the property of ultrafilters singled out in Lemma 6.1.

To illustrate Th. 6.4 and the ultraproduct theorem mentioned earlier, let us take a sentence $\Psi$ in the language of fields and suppose that we can find fields of arbitrarily large characteristic for which $\Psi$ holds. Then $\Psi$ also holds in their ultraproduct, and this will be of characteristic 0, if it was formed with a non-principal ultrafilter. For suppose that $k_i$ is a field of characteristic $p_i$, where $p_1 \leqslant p_2 \leqslant \cdots$ and $p_i \to \infty$ as $i \to \infty$. Then the sentence

$\Phi_n$. $$\underbrace{1 + 1 + \cdots + 1}_{n} = 0$$

holds in only finitely many of the $k_i$, for each $n$, and hence its negation $\neg \Phi_n$ holds in their ultraproduct; this shows the latter to be of characteristic 0. Since $\Psi$ holds in each $k_i$ it also holds in the ultraproduct. Thus we have

PROPOSITION 6.5 *An elementary sentence which holds in a field $k_i$ of finite characteristic $p_i$, where the $p_i$ are unbounded, also holds in certain fields of characteristic* 0. ■

For example, consider the statement: every non-degenerate binary quadratic form is universal. This may be stated as:

$$\forall a, b, c \exists x, y [a \neq 0 \wedge b \neq 0 \Rightarrow ax^2 + by^2 = c].$$

It holds for all finite fields of characteristic not two (cf. A.2, **7.2**); hence it also holds in certain fields of characteristic 0.

As a second illustration we observe that a field of characteristic $p$ may be

defined by the sentence '$\neg \Phi_1 \wedge \Phi_p$'. Hence a field of finite characteristic is defined by the 'infinite disjunction'

$$\neg \Phi_1 \wedge [\Phi_2 \vee \Phi_3 \vee \Phi_5 \vee \cdots].$$

This is not an elementary sentence as it stands. But we can assert further that it is not equivalent to any set of elementary sentences. For if it were, it would hold in all fields of finite characteristic and hence also in some fields of characteristic 0, which is not so.

### Exercises

(1) Show that any ultrafilter which includes a finite set must be principal.

(2) Let $A$ be an infinite set. Show that for every non-empty subset $B$ of $A$ there is an ultrafilter $\mathscr{F}_B$ including $B$. What is the condition on $B$ for $\mathscr{F}_B$ to include all cofinite subsets?

(3) Let $I$ be a set and $\mathscr{P}(I)$ the Boolean algebra of all subsets of $I$ (cf. A.2, **2.3**). Defining ideals of Boolean algebras as inverse images of 0 in homomorphisms, show that a filter on $I$ is just the complement of a non-zero ideal in $\mathscr{P}(I)$. Which ideals correspond to ultrafilters?

(4) Show that any formula $\Phi(x)$ holds in an ultraproduct $\prod A_i/\mathscr{F}$ iff it holds in all the factors $A_i$ for an $\mathscr{F}$-large set of indices. (Hint: verify that the formulae for which this is true include all 'atomic formulae' and are closed under $\vee, \wedge, \neg, \forall, \exists$. Hence the result holds for sentences = 'formulae without free variables'.)

(5) (Compactness theorem of model theory) Let $\mathscr{T}$ be a set of elementary sentences about $\Omega$-algebras. Show that if each finite subset $P$ of $\mathscr{T}$ has a model (i.e. there is an algebra in which each sentence of $P$ holds), then $\mathscr{T}$ has a model. (Hint: for each $P \subseteq \mathscr{T}$ take a model $A_P$ and form a suitable ultraproduct of the $A_P$.)

(6) Show that an integral domain $R$ which is embeddable in a direct product of skew fields is embeddable in a skew field. (Hint: let $K_i$ ($i \in I$) be the family of skew fields and for $c \in R^\times$ let $I_c$ be the set of indices $i$ for which $c$ is inverted in $K_i$. Verify that the $I_c$ form a filter base.)

## 1.7 The natural numbers

In Volume 1 we took the natural numbers for granted and showed how the negative numbers and fractions could be formed from them. For a rigorous development it is necessary either to provide an axiomatic foundation for the natural numbers, or to deduce their properties from some other domain such as set theory. The latter alternative would of course make it necessary to include an axiomatic foundation of set theory and this would involve us far deeper in foundational questions than is appropriate here. Such a study would occupy a

whole volume by itself and would not greatly help us in our understanding of algebra. We shall therefore confine ourselves to a derivation of the properties of natural numbers from the Peano axioms and a brief discussion of their significance, as well as their relevance to algebra. As we shall see, the framework of universal algebra is particularly appropriate for this purpose.

We begin by writing down a system of axioms for the natural numbers. This is not so much to give a rigorous foundation as to make explicit the properties of numbers we are using. The notions of set theory will be used freely, in the intuitive form introduced in Volume 1. We shall also use individual numbers (e.g. to label the axioms) without hesitation: no axioms are needed for the numbers up to 12, or for that matter up to $10^{100}$. The purpose of the axioms is to allow us to deal with the set of *all* numbers.

The axioms, as stated essentially by G. Peano in 1889, are:

**N.1** 1 *is a natural number.*
**N.2** *Every natural number n has a unique successor n′, which is again a natural number.*
**N.3** 1 *is not the successor of any number.*
**N.4** *Distinct numbers have distinct successors:* $m \neq n \Rightarrow m' \neq n'$.
**N.5** (*Principle of induction*) *A set of natural numbers containing* 1 *and with each number its successor contains all natural numbers.*

The set of all natural numbers will be denoted by **N**. We can think of **N** as an algebra with a single unary operation, the *successor function* $x \mapsto x'$. Let us call an algebra with a single unary operation an *induction algebra.* By **N.2**, **N** is an induction algebra; moreover it is generated by the single element 1, by **N.5**. To elucidate the structure of **N** we begin with a general lemma on induction algebras.

LEMMA 7.1 *Let A be an induction algebra. Then the subalgebra B generated by an element b of A consists of b and successors of elements of B.*

*Proof.* The set $B_1$ consisting of $b$ and successors of elements of $B$ is contained in $B$; it contains $b$ and the successor of any element of $B_1$ and so is a subalgebra. Since $B$ is the least subalgebra containing $b$, it follows that $B_1 = B$, as we had to show. ■

For example, if we take $A = \mathbf{N}$, $b = 1$ and remember **N.5**, we see from the lemma that every number different from 1 is the successor of a number. Thus if $n \neq 1$, then there is a number which we shall denote by $n-1$ such that $(n-1)' = n$. By **N.4**, $n-1$ is uniquely determined by $n$; it is called the *immediate predecessor* of $n$.

For any $n \in \mathbf{N}$ we denote by $[n|$ the subalgebra generated by $n$.

LEMMA 7.2 *For all $n \in N$, we have $n \notin [n'|$.*

*Proof.* Let us show first that

$$n' \neq n. \tag{1}$$

For $n = 1$ this holds by **N.3**; if it holds for any $n \neq 1$, then it holds for $n'$ by **N.4**, hence it holds for all $n$, by induction (i.e. **N.5**).

Now by Lemma 7.1, $[1'|$ consists entirely of successors of numbers, whereas 1 is not a successor, hence $1 \notin [1'|$. Suppose now that $n \notin [n'|$ but that $n' \in [n''|$. By (1), $n' \neq n''$, so $n'$ must be the successor of an element in $[n''|$; but this can only be $n$ (by **N.4**), so $n \in [n''|$ and $[n''| \subset [n'|$, therefore $n \in [n'|$, which contradicts the hypothesis. By induction we conclude that $n \notin [n'|$ for all $n$. ■

Let us write $|n]$ for the complement of $[n'|$ in **N**. By Lemma 7.2, $n \in |n]$; the elements of $|n]$ other than $n$ will be called the *predecessors* of $n$. Clearly they include the immediate predecessor of $n$, if $n \neq 1$. With these preparations we can prove a result on which the box principle is based.

THEOREM 7.3 *Let $m, n \in \mathbf{N}$. There is an injective mapping from $|m]$ to $|n]$ if and only if $|m] \subseteq |n]$. Further there is a bijection between $|m]$ and $|n]$ if and only if $m = n$.*

*Proof.* If $|m] \subseteq |n]$, then the inclusion mapping is the required injection, and for $m = n$ this is a bijection. Conversely, assume that $f : |m] \to |n]$ is an injective mapping; we must show that $|m] \subseteq |n]$. When $m = 1$, then since $1 \notin [n'|$, we have $1 \in |n]$ and so $|1] \subseteq |n]$. We may therefore assume that $m \neq 1$ and use induction on $m$. Since $m \neq 1$, there is an immediate predecessor $m - 1$; we define a mapping $g : |m - 1] \to |n]$ by the rule

$$kg = \begin{cases} kf & \text{if } kf \neq n, \\ mf & \text{if } kf = n. \end{cases}$$

To check that this is a well-defined mapping we note that there is at most one number $k$ such that $kf = n$, because $f$ is injective. Denote this number by $k_0$; if $k_0 = m$ or if $k_0$ is not defined (because the image of $f$ does not include $n$), then $g$ is just $f$ restricted to $|m - 1]$. Otherwise $g$ differs from $f$ only at $k_0$ and there it has the value $mf$ which it assumes nowhere else, for the domain of $g$ does not include $m$. Thus $g$ is well-defined; moreover $g$ is injective and it does not assume the value $n$. It follows that $n \neq 1$ and that $g$ is an injective mapping from $|m - 1]$ to $|n - 1]$. By the induction hypothesis, $|m - 1] \subseteq |n - 1]$; taking successors on both sides we obtain $|m] \subseteq |n]$.

If there is a bijection between $|m]$ and $|n]$, we conclude that $|m] = |n]$ and therefore $m = n$, because $n$ is the unique element of $|n]$ without a sucessor. ■

A set $S$ is called *finite* if there is a bijection between $S$ and $|n]$, for some $n \in \mathbf{N}$. By what has been said, there can be at most one such $n$ and this is called the *cardinal* of $S$. Thus for any finite set there is a unique natural number which is its cardinal.

Theorem 7.3 also leads to the familiar ordering of the natural numbers. Let us write $m \leqslant n$ to mean: $|m] \subseteq |n]$. It is clear that this relation is reflexive and transitive, and by the last part of Th. 7.3 we see that $m \leqslant n$, $n \leqslant m$ implies $m = n$. Thus we have a partial ordering. From the definition it is clear that $m \leqslant n$ implies $m' \leqslant n'$, so it only remains to show that '$\leqslant$' is a total ordering. Given $m, n \in \mathbf{N}$, if $m = 1$, then clearly $m \leqslant n$; similarly if $n = 1$, then $n \leqslant m$. Now if $m, n \neq 1$, we can form $m - 1$, $n - 1$ and by induction either $m - 1 \leqslant n - 1$ or $n - 1 \leqslant m - 1$. Taking successors we find that $m \leqslant n$ or $n \leqslant m$. We shall also adopt the usual notation of writing $m < n$ or $n > m$ to mean '$m \leqslant n$ but $m \neq n$' and $m \geqslant n$ to mean $n \leqslant m$.

In contrapositive form Th. 7.3 shows that if $m \notin |n]$, so that $m > n$, then there can be no injective mapping from $|m]$ to $|n]$. In particular, taking $m = n' (= n + 1)$, we see that when $n'$ objects are distributed over $n$ boxes, at least one box contains more than one element. This is just Dirichlet's Box Principle, already encountered in (A.2, p.2), where it was stated without proof.

Our next aim is to establish a universal property of $\mathbf{N}$, which forms the basis of the process of definition by recursion.

THEOREM 7.4 $\mathbf{N}$ *is the free induction algebra on* 1. *Thus if* $A$ *is any induction algebra and* $a \in A$, *then there is a unique homomorphism* $\alpha : \mathbf{N} \to A$ *such that* $1\alpha = a$.

*Proof*. In detail the assertion states that $A$ is a set with a unary operation $x \mapsto x'$, and given $a \in A$, there is a unique mapping $x \mapsto x\alpha$ from $\mathbf{N}$ to $A$ such that

$$1\alpha = a, \qquad x'\alpha = (x\alpha)' \qquad x \in \mathbf{N}.$$

By Prop. 1.1 there can be at most one such mapping. To prove its existence we form the direct product $\mathbf{N} \times A$; this is again an induction algebra, with operation $(x, y)' = (x', y')$. Let $H$ be the subalgebra of $\mathbf{N} \times A$ generated by $(1, a)$; further denote by $p, q$ the projection mappings of $\mathbf{N} \times A$ on the factors $\mathbf{N}$ and $A$ respectively, and by $p_1, q_1$ their restrictions to $H$.

The image of $H$ under $p_1$ is a subalgebra of $\mathbf{N}$ containing 1, and hence is $\mathbf{N}$ itself, by N.5. Thus for each $x \in \mathbf{N}$,

$$\text{there exists } y \in A \text{ such that } (x, y) \in H. \tag{2}$$

We claim that for each $x \in \mathbf{N}$ there is exactly one such $y$. For by Lemma 7.1, $H$ consists of $(1, a)$ together with successors of elements of $H$. Any successor is of the form $(x', y')$; here the first component is different from 1, by N.3, so the $y$ determined in (2) by $x = 1$ is unique. Let $\mathbf{N}_1$ be the subset of all $x \in \mathbf{N}$ occurring as first component in just *one* element of $H$, i.e. for which the $y \in A$ determined

in (2) is unique. As we have seen, $1 \in \mathbf{N}_1$; if we can show that $\mathbf{N}_1$ contains with each element its successor, it will follow from **N.5** that $\mathbf{N}_1 = \mathbf{N}$.

Let $x \in \mathbf{N}_1$ and suppose that $y_1, y_2 \in A$ are such that $(x', y_i) \in H$ $(i = 1, 2)$; we have to show that $y_1 = y_2$. Since $x' \neq 1$, $(x', y_i)$ is a successor in $H$, say $(x', y_i) = (u_i, v_i)' = (u_i', v_i')$, where $(u_i, v_i) \in H$. Equating first components, we find $x' = u_1' = u_2'$, and by **N.4**, $x = u_1 = u_2$, i.e. $(x, v_i) \in H$ for $i = 1, 2$. But $x \in \mathbf{N}_1$ and this means that $v_1 = v_2$; hence $v_1' = v_2'$ and so $y_1 = y_2$ as we had to show. Therefore $x' \in \mathbf{N}_1$ and by **N.5** we conclude that $\mathbf{N}_1 = \mathbf{N}$.

We have now shown that for each $x \in \mathbf{N}$ there is a unique $y \in A$ such that $(x, y) \in H$. Writing $x\alpha$ for $y$, we find that $1\alpha = a$, $x'\alpha = (x\alpha)'$ and this is what we had to prove. ■

Functions on $\mathbf{N}$ are frequently defined recursively, for example the sum of the squares of the first $n$ natural numbers may be defined as the function $g: \mathbf{N} \to \mathbf{N}$ such that

$$g(1) = 1, \qquad g(n+1) = g(n) + (n+1)^2.$$

It may appear intuitively obvious that this defines a function, but this needs to be proved. The basic reason is that $\mathbf{N}$ is a *free* induction algebra on 1, as we shall now see. The statement in Th. 7.5 is a little more general, but the proof begins with the case exemplified above.

THEOREM 7.5 *Given $a \in \mathbf{N}$ and any function $f$ from $\mathbf{N}^2$ to $\mathbf{N}$, there exists a unique function $\varphi$ from $\mathbf{N}$ to itself, satisfying the equations:*

$$\varphi(1) = a, \qquad \varphi(n') = f(n, \varphi(n)). \tag{3}$$

*Proof.* Suppose first that $f$ is independent of its first argument. Then we have to find $\varphi: \mathbf{N} \to \mathbf{N}$ to satisfy

$$\varphi(1) = a, \qquad \varphi(n') = f(\varphi(n)). \tag{4}$$

In this case the result follows immediately from Th. 7.4, taking $A$ there to be the set $\mathbf{N}$ with $f$ as its successor function.

In the general case we take $A = \mathbf{N}^2$ with successor function

$$(x, y) \mapsto (x', f(x, y)),$$

and apply Th. 7.4 with the element $(1, a)$ in place of $a$. We obtain a mapping $\varphi: \mathbf{N} \to \mathbf{N}^2$; if its projections on the factors are $\varphi_1$ and $\varphi_2$, then $\varphi_1(1) = 1$, $\varphi_1(n') = n'$, hence $\varphi_1(x) = x$ for all $x \in \mathbf{N}$, and now

$$\varphi_2(1) = a, \qquad \varphi_2(n') = f(n, \varphi_2(n)).$$

Thus $\varphi_2$ is the required function. ■

This result emphasizes the difference between 'proof by induction' and

'definition by recursion'. Whereas the former is embodied in **N**.5, the latter also relies on **N**.3, 4, and a further argument (such as that leading to Th. 7.5) is needed to prove it. From the algebraist's point of view the situation can be summed up by saying that a *proof* by induction depends on the fact that **N**, as induction algebra, is *generated* by 1, whereas the method of *definition* by recursion depends on the fact that **N** is generated *freely* by 1.

As an application let us see how Th. 7.5 may be used to define addition and multiplication on **N**. Given $a \in \mathbf{N}$, there is a mapping $\alpha_A : \mathbf{N} \to \mathbf{N}$ such that

$$1\alpha_a = a', \qquad x'\alpha_a = (x\alpha_a)'.$$

If we write $a + x$ in place of $x\alpha_a$, these equations take on a more familiar form:

$$a + 1 = a', \qquad a + x' = (a + x)'. \tag{5}$$

From this definition it is easy to prove the associative and commutative laws of addition:

$$(a + b) + c = a + (b + c), \tag{6}$$

$$a + b = b + a. \tag{7}$$

We shall prove (6) as an example and leave (7) to the reader. For $c = 1$, (6) reduces to $(a + b) + 1 = a + (b + 1)$, i.e. $(a + b)' = a + b'$, which is true by the definition (5). If we assume that (6) holds for $c = n$, then $(a + b) + n' = [(a + b) + n]' = [a + (b + n)]' = a + (b + n)' = a + (b + n')$, by (5) and the case $c = n$. Hence (6) holds for $c = n'$, and so by induction it holds for all $n$.

Similarly we can define multiplication by constructing for each $a \in \mathbf{N}$, a mapping $\mu_a : \mathbf{N} \to \mathbf{N}$ such that $1\mu_a = a$, $x'\mu_a = x\mu_a + a$. The existence and uniqueness follow again from Th. 7.5, and we write as usual $x\mu_a = ax$, so that the definition takes the form

$$a1 = a, \qquad a(x + 1) = ax + a. \tag{8}$$

The associative and commutative laws can again be proved without difficulty, as well as the distributive law. If we adjoin a new element, denoted by 0, to **N**, to satisfy $a + 0 = a$, $a0 = 0$, we have a monoid under addition and the procedure of A.1, **6.2** enables us to embed it in a group, which is the additive group of integers **Z**. All that is needed is a proof of the cancellation law: $a + c = b + c$ implies $a = b$; this presents no difficulty and may be left to the reader. Now the multiplication on **N** can easily be extended to the set **Z** of all integers and the distributive law verified; thus we have obtained the ring structure on **Z**. It is also not difficult to extend the ordering of **N** to **Z**; it is still preserved under addition but only under multiplication by a positive number (i.e. an element of **N**).

Let us consider the Peano axioms from the point of view of first order logic. We remark that **N**.1–4 are elementary sentences, whereas **N**.5 is not. Without going into detail this can be explained by saying that when expressed formally,

**N.5** involves quantification over *sets* of numbers and not merely numbers. It is easy to give examples of structures other than **N** which satisfy **N.1–4**. For example, consider the set $T$ consisting of the disjoint union of two induction algebras, $T_1$ isomorphic to **N** and $T_2$ isomorphic to **Z**. Then $T$ satisfies **N.1–4**, though of course not **N.5**.

This leaves open the question whether it may not be possible to replace **N.5** by elementary sentences, so as to characterize the natural numbers by elementary sentences alone. This equation can be answered negatively by forming an ultrapower $I$ of **N**, with a non-principal ultrafilter. $I$ is again an induction algebra and by the ultraproduct theorem it satisfies all the elementary sentences holding in **N**, but $I$ is not isomorphic to **N**, because it is again totally ordered, by the rule: $x \leqslant y$ iff $x_i \leqslant y_i$ for all components in a large set, but it is not well-ordered. To see this, we consider the sequence $a_r = (a_{rn})$ of elements of $I$, where $a_{1n} = n$ and for $r \geqslant 1$,

$$a_{r+1n} = [\tfrac{1}{2}(a_{rn} + 1)],$$

where $[x]$ denotes the greatest integer $\leqslant x$. Thus $a_1 = (1, 2, 3, 4, \ldots)$, $a_2 = (1, 1, 2, 2, 3, 3, 4, 4, \ldots)$, $a_3 = (1, 1, 1, 1, 2, 2, 2, 2, 3, \ldots)$ etc. This is a strictly descending sequence, because the set

$$\{n \in \mathbf{N} \mid a_{r+1m} \geqslant a_{rn}\}$$

is finite, for $r = 1, 2, \ldots$.

An induction algebra $A$ satisfying the same elementary sentences as **N** is said to be *elementarily equivalent* to **N**, or also a *model* of **N**; if $A$ is not isomorphic to **N**, it is called a *non-standard model*. Such non-standard models of **N** allow one to introduce 'infinite' numbers, just as a non-standard model of **R** will contain 'infinitesimal' numbers. Non-standard analysis is a powerful tool with many applications, but it lies outside the scope of this work (cf. e.g. Stroyan and Luxemburg [1976])

## Exercises

(1) Prove the commutative law of addition on **N**.

(2) Prove the associative and commutative laws of multiplication, the distributive law and the cancellation law.

(3) Give a direct proof that an induction algebra generated by a single element 1 satisfies either **N.3** or **N.4**.

(4) Show that in any induction algebra the union of two subalgebras is again a subalgebra.

(5) Let $f$ be the function on the non-negative integers, defined by $f(0) = 0$, $f(x') = x$. Describe the function $g(x, y)$ defined by $g(x, 0) = x$, $g(x, y') = f(g(x, y))$.

(6) Let $p_n$ be the number of equivalence relations on a set of $n$ elements. Obtain the following recursion formula for $p_n$:

$$p_{n+1} = \sum \binom{n}{i} p_i, \qquad p_0 = 1.$$

Show also that $\sum p_n x^n/n! = \exp[(\exp x) - 1]$.

(7) Define the natural ordering on $\mathbf{Z}$ in terms of the ordering on $\mathbf{N}$ and prove its compatibility with addition and multiplication.

(8) Show that there is no total ordering on $\mathbf{Z}/m$, the set of integers mod $m$, preserving addition and multiplication.

(9) Use Th. 7.3 to give a proof that $\mathbf{N}$ is not finite.

(10) Give a direct proof by induction that there exists no surjective mapping from $|m]$ to $|n]$ if $m < n$.

## Further exercises on Chapter 1

(1) Let $\mathscr{V}$ be a variety of $\Omega$-algebras and assume that there is a $\mathscr{V}$-algebra $C$ whose carrier is finite with more than one element. Let $F_n$ be the free $\mathscr{V}$- algebra on an $n$-element set; by establishing a correspondence between $C^n$ and homomorphisms from $F_n$ to $C$ show that any generating set of $F_n$ has at least $n$ elements. Deduce that $F_m$ and $F_n$ are not isomorphic when $m \neq n$.

(2) Show that the free distributive lattice on three generators has 18 elements. (Hint: form the different expressions in $x_1, x_2, x_3$; try the cases of one and two generators first.)

(3) Show that the free distributive lattice with 0, 1 on $n$ free generators is $2^{2^n}$.

(4) Define $\mathbf{N}$ as an algebra with a single unary operation $\lambda$ by the rule

$$n\lambda = \begin{cases} n-1 & \text{if } n > 1, \\ 1 & \text{if } n = 1. \end{cases}$$

Show that any non-trivial homomorphic image of $\mathbf{N}$ is isomorphic to $\mathbf{N}$, but that $\mathbf{N}$ is not simple.

(5) Let $S$ be a set with a dependence relation. A subset of $S$ is said to be *closed* if it contains all the elements dependent on it. Show that the intersection of any family of closed sets is closed. Find all dependence relations for which the union of any two closed sets is closed.

(6) Let $S$ be a set with a spanning relation (**D**.0, 1) and let $A$ be the set of all $a \in S$ which can be omitted from any spanning set, i.e. if $X \cup \{a\}$ spans $S$, then so does $X$. Show that $A$ is the intersection of all maximal closed proper subsets of $S$ (cf. **4.1** below).

(7) Show that a non-zero vector space of dimension $\alpha$ over a countable field has cardinal $\max\{\alpha, \aleph_0\}$. Deduce that the Hamel basis of $\mathbf{R}$ has $2^{\aleph_0}$ elements.

(8) (H. Whitney) Let $S$ be a finite set with a dependence relation $\Phi$. Let $\Phi^*$ be the set of complements of bases of $S$ for the relation $\Phi$. Show that there is a dependence relation on $S$ whose collection of bases is $\Phi^*$. (Hint: show that there is a dependence relation whose independent sets are the complements of the spanning sets of $S$.)

(9) (Mac Lane–Ingleton) Let $S = \{1, \ldots, 9\}$ and consider the dependence relation with minimal dependent families $\{1,2,3\}$, $\{4,5,6\}$, $\{1,7,5\}$, $\{1,8,6\}$, $\{2,7,4\}$, $\{2,9,6\}$, $\{3,8,4\}$, $\{3,9,5\}$. Show that there is no set of vectors in a vector space with this dependence relation. (Hint: use Pappus' theorem to show that $\{7,8,9\}$ would have to be dependent.)

(10) A *matroid* may be defined as a finite set $M$ with a family $\mathscr{C}$ of subsets called *circuits* such that (i) no proper subset of a circuit is a circuit and (ii) if $X, Y$ are distinct circuits and $a \in X \cap Y$, then $X \cup Y \backslash \{a\}$ contains a circuit. Verify that the minimal cycles in a graph form a matroid. Show also that the minimal dependent subsets in a vector space form a matroid.

Given a matroid $M$, let us call $a \in M$ *dependent* on a set $X \subseteq M$ if $X$ contains a subset $X'$ such that $X' \cup \{a\}$ is a circuit. Show that this defines a dependence relation on $M$. (Thus the theory of dependence relations is coextensive with the theory of matroids.)

(11) (G. M. Bergman) On an infinite-dimensional vector space $V$, define a *filter of subspaces* as a set of subspaces of $V$ satisfying **F**.1–3 (where **F**.1 now reads: $V \in \mathscr{F}$, $0 \notin \mathscr{F}$), and define an ultrafilter again as a maximal filter. Show that if $\mathscr{F}$ is an ultrafilter, then every subspace of $V$ either lies in $\mathscr{F}$ or has a complement in $\mathscr{F}$. Let $\mathscr{A}$ be the set of linear maps $V \to V$ which are 'continuous', i.e. the inverse image of an $\mathscr{F}$-space is an $\mathscr{F}$-space, and let $\mathscr{N}$ be the set of maps with kernel in $\mathscr{F}$. Show that $\mathscr{N}$ is an ideal in $\mathscr{A}$ and $\mathscr{A}/\mathscr{N}$ is a skew field.

(12) A ring $R$ is called *prime* if $R \neq 0$ and $aRb = 0$ implies $a = 0$ or $b = 0$ (cf. Ch. 9 below). Show that an ultraproduct of prime rings is prime, but an ultraproduct of simple rings need not be simple. Is an ultrapower of a simple ring necessarily simple? What about $\mathfrak{M}_n(K)$, where $K$ is a skew field?

(13) Let $k$ be a field and $\mathscr{D}$ an ultrafilter on $\mathbf{N}$. Show that there is a monoid homomorphism $\prod \mathfrak{M}_n(k)/\mathscr{D} \to k^{\mathbf{N}}/\mathscr{D}$, whose kernel is an ideal.

(14) For any set $S$ denote by $\{S\}$ the set whose single member is $S$ (as usual). Let $M$ be the induction algebra generated by $\varnothing$ in this way, with successor operation $S' = S \cup \{S\}$. Show that $M$ satisfies **N**.1–5 with $\varnothing$ in place of 1.

(15) Use Th. 7.5 to define exponentiation on $\mathbf{N}$ by $a^1 = a$, $a^{b'} = a \cdot a^b$. What goes wrong if we try to define this operation on $\mathbf{Z}/m$?

# 2
# Multilinear algebra

As an instance of free algebras we have free rings. It is convenient to have a coefficient domain, usually a non-trivial commutative ring $K$, and we shall then be dealing with free $K$-algebras; for the case when $K$ is a field, such algebras were already briefly discussed in A.2, **11.5**. Free algebras form the subject of **2.2**, which also deals with the related notions of tensor algebra and symmetric algebra on a $K$-module and describes derivations. All these algebras are examples of graded rings and **2.1** is devoted to properties of the grading. A graded algebra has an important invariant, its Hilbert series, essentially a power series whose coefficients indicate the dimensions of the components. In **2.3** we show that the Hilbert series of a commutative Noetherian ring is a rational function and also prove the Golod–Shafarevich theorem, giving a sufficient condition for a graded algebra to be infinite-dimensional. The applications, to construct a finitely generated algebra which is nil but not nilpotent, and a finitely generated infinite $p$-group, are sketched in the exercises. The final section deals with exterior algebras, leading to a simple development of determinants.

## 2.1 Graded algebras

In a polynomial ring the *degree* of a polynomial is a useful concept which allows us to analyse the ring in different ways. Thus we may write the polynomial ring as a direct sum of its homogeneous components. There are many other rings which share this property, and it is convenient to describe its general form before applying it. Throughout, the coefficient ring $K$ is an arbitrary commutative ring. The case when there is no coefficient ring is included by taking $K = \mathbf{Z}$, since every ring is a $\mathbf{Z}$-algebra. To avoid trivialities, we shall usually assume that $K$ is non-trivial.

DEFINITION A *graded $K$-algebra* is a family $(A_n)$ of $K$-modules indexed by $\mathbf{Z}$, with a bilinear mapping

$$A_m \times A_n \to A_{m+n}, \tag{1}$$

denoted by $(a, b) \mapsto ab$, such that for any $a \in A_m$, $b \in A_n$, $c \in A_p$,

$$a(bc) = (ab)c \tag{2}$$

and there is an element $e \in A_0$ such that $ae = ea = a$ for all $a \in A_n (n \in \mathbf{Z})$. $A_n$ is called the *component* of degree $n$; if $a \in A_n$, we say that $a$ is *homogeneous* of degree $n$ and write $\deg a = n$.

We shall assume that distinct $A_n$ have only 0 in common, so that every non-zero element has a unique degree, while 0 has all degrees. From the definition it is clear that $A_0$ is a $K$-algebra and each $A_n$ is an $A_0$-bimodule.

Given a graded $K$-algebra $(A_n)$, let us put $A = \oplus_n A_n$; we can define a multiplication on $A$, using (1) and the distributive laws. Thus if $a = a_m + a_{m+1} + \cdots + a_M$ and $b = b_n + b_{n+1} + \cdots + b_N$, where $a_i, b_i \in A_i$, then

$$ab = a_m b_n + a_m b_{n+1} + a_{m+1} b_n + \cdots + a_M b_N.$$

It is clear that multiplication is associative, by (2), and distributive, with $e$ as unit element; thus $A$ is a $K$-algebra. Sometimes $A$ is called an *internally graded* algebra, corresponding to the *externally graded* algebra $(A_n)$. Of course we may be given an algebra which is internally graded, i.e. $A$ may have the form $A = \oplus_n A_n$, where $(A_n)$ is externally graded.

*Examples*

1. Let $A = K[x_1, \ldots, x_d]$ be a polynomial ring in $d$ indeterminates, and denote by $A_n$ the submodule of all homogeneous polynomials of total degree $n$; then $A$ is internally graded by the $A_n$. We observe that in this case $A_n = 0$ for $n < 0$; this is expressed by saying that $A$ is *positively* graded. In particular, for $d = 1$ we have $A = K[x]$; here $A_n = Kx^n$.
2. The ring of 'Laurent polynomials' $k[t, t^{-1}]$ is an example of a graded ring with non-zero components for all $n \in \mathbf{Z}$. By contrast, the ring of all Laurent series has no obvious grading (it has a filtration, cf. Ex. (4) of Further Exercises).
3. Any ring $R$ may be graded by putting $A_0 = R, A_n = 0$ for $n \neq 0$. Such a ring is said to be *concentrated in degree* 0.

Let $A$ be a graded ring. A *graded $A$-module* is a family of $K$-modules $(M_n)$ such that $A_r M_s \subseteq M_{r+s}$ for all $r, s \in \mathbf{Z}$. We can again form an internally graded module by writing $M = \oplus M_n$; an element $x$ of $M$ is called *homogeneous* of *degree $n$* if $x \in M_n$. Every element of $M$ can be uniquely expressed as a sum of finitely many homogeneous elements. We remark that for any graded ring $A$, a graded $A$-module $M$ which is finitely generated can always be generated by a finite set of homogeneous elements; we need merely take all the components of the given set of generators.

The tensor product (over $K$) of two graded $K$-algebras is again a graded $K$-algebra in an obvious sense: Given $A = (A_n)$, $B = (B_n)$, we have $A \otimes B = C$, where

$$C_n = \bigoplus_r A_r \otimes B_{n-r}. \tag{3}$$

For example, the polynomial ring $K[x, y]$ may be constructed as the tensor product of $K[x]$ and $K[y]$. We observe that when $A$ and $B$ are both positively graded, the sum in (3) is necessarily finite, but this is not essential for the definition to make sense.

A graded algebra $A$ is said to be *anticommutative* if

$$a_m b_n = (-1)^{mn} b_n a_m \qquad \text{for } a_m \in A_m, b_n \in A_n. \tag{4}$$

Thus to interchange $a_m$ and $b_n$ we change the sign if $m, n$ are both odd and leave it the same otherwise. For example, if $A_n = 0$ for all odd $n$, then (4) reduces to the commutative law. Some writers omit the 'anti' and call (4) the commutative law for graded algebras; although there is some justification for this practice, we shall not follow it but refer to (4) as the *anticommutative law*. We remark that if $A$ is anticommutative and $K$ admits division by 2 (e.g. if $K$ is a field of characteristic not 2), then $a^2 = 0$ for any $a$ of odd degree; this is sometimes made part of the general definition.

Let $A = (A_n)$, $B = (B_n)$ be two graded algebras. By a *linear mapping of degree* $r$ from $A$ to $B$ we understand a family of $K$-linear mappings $f_n : A_n \to B_{n+r}$. A linear mapping of degree 0 from $A$ to $B$ which is a ring homomorphism is called a *homomorphism of graded algebras*. Similarly we can define linear mappings of degree $r$ between graded $A$-modules.

The definition of subalgebra and ideal should be clear: if we regard $A$ as internally graded, a *graded* subalgebra $B$ of $A$ (also called a *homogeneous* subalgebra of $A$) is a subalgebra $B$ in the ordinary sense such that $B = \sum(B \cap A_n)$; likewise a graded ideal $\mathfrak{a}$ of $A$ is an ideal in the usual sense such that $\mathfrak{a} = \sum(\mathfrak{a} \cap A_n)$. This condition can be expressed by saying that for any finite family $a_n \in A_n (n \in \mathbf{Z})$ we have $\sum a_n \in B$ (or $\mathfrak{a}$) iff each $a_n$ lies in $B$ (or $\mathfrak{a}$). The reader should verify that the image and kernel of a homomorphism of graded algebras are a graded subalgebra and ideal in this sense. Conversely, given a graded algebra $A$ and a graded ideal $\mathfrak{a}$ in $A$, the quotient $A/\mathfrak{a}$ is again graded, and the natural homomorphism $A \to A/\mathfrak{a}$ is a homomorphism of graded algebras. Finally we observe that if $X$ is a subset of $A$ consisting of homogeneous elements, then the ideal $\mathfrak{a}$ of $A$ generated by $X$ is graded, for every element of $\mathfrak{a}$ can be written as a sum of terms $uxv$, again in $\mathfrak{a}$, where $x \in X$ and $u, v$ are homogeneous.

In a positively graded ring $A$, the set $A_+ = \Sigma_1^\infty A_n$ is an ideal called the *augmentation ideal*; it is such that $A = A_0 \oplus A_+$. If $A_0 \cong K$ under the mapping $\alpha \mapsto \alpha e$ ($\alpha \in K$), $A$ is an augmented $K$-algebra in the sense defined in A.2, **5.1**.

In the commutative case Noetherian graded rings have the following convenient description.

PROPOSITION 1.1 *A positively graded commutative ring $A$ is Noetherian if and only if $A_0$ is Noetherian and $A$ is finitely generated as a ring over $A_0$.*

*Proof.* If $A$ is generated by $x_1, \ldots, x_d$ over $A_0$, then it is a homomorphic image

of the polynomial ring $A_0[x_1,\dots,x_d]$ and hence is Noetherian by the Hilbert basis theorem (A.2, **9.4**). Conversely, assume that $A$ is Noetherian; then so is $A_0 = A/A_+$ and $A_+$ is finitely generated as ideal. By an earlier remark we may take a family of homogeneous generators $x_1,\dots,x_d$ of $A_+$, of degrees $r_1,\dots,r_d$ say, and write $C$ for the subring generated by the $x_i$ over $A_0$. To complete the proof, we show that $A_n \subseteq C$, by induction on $n$. Clearly $A_0 \subseteq C$; if $n>0$ and $A_r \subseteq C$ for $r<n$, take $y \in A_n$. We can write $y$ as a linear combination of the $x$'s, say $y = \sum a_i x_i$, where $a_i \in A_{n-r_i}$ (and $A_s = 0$ for $s<0$ is understood). Since $r_i > 0$, we can apply induction to conclude that $a_i \in C$, hence $y \in C$ and so $A_n \subseteq C$. It follows that $A = C$, and so $A$ is finitely generated as ring over $A_0$. ■

In all that has been said about graded algebras, the set $\mathbf{Z}$ used to index the components could be replaced by any additive group or even a monoid. The only other case we shall actually need is that of algebras graded by $\mathbf{Z}/2$, the group of two elements, briefly: mod 2 graded algebras. Such an algebra has the form $A = A_0 \oplus A_1$, where $A_0A_1 + A_1A_0 \subseteq A_1, A_0A_0 + A_1A_1 \subseteq A_0$; it is often called a *superalgebra.*

### Exercises

(1) Let $f: A \to B$ be a linear mapping of degree $r$ between graded algebras. Show that if $f$ is a homomorphism, then $r$ must be 0.

(2) Verify that for any homomorphism of graded algebras, the kernel and image are graded.

(3) Let $R$ be a $K$-algebra of the form $R = K \oplus \mathfrak{a}$, where $\mathfrak{a}$ is an ideal (i.e. an augmented $K$-algebra). Describe $R$ as a graded algebra, with a grading monoid of two elements.

(4) Show that $\mathbf{C}$ may be regarded as a mod 2 graded algebra over $\mathbf{R}$. What is the generalization to binomial field extensions, i.e. extensions generated by a root of $x^n = a$?

(5) Define a graded tensor product $A \hat{\otimes} B$ of graded algebras $A, B$ as a tensor product with multiplication $(a_i \hat{\otimes} b_r)(a'_j \hat{\otimes} b'_s) = (-1)^{jr} a_i a'_j \hat{\otimes} b_r b'_s$. Show that the graded tensor product of anticommutative algebras is anticommutative.

## 2.2 Free algebras and tensor algebras

We have already in A.2, **11.5** encountered the free algebra $K\langle X\rangle$ on a set $X$ over a field $K$. We recall that this consisted of all polynomials in the elements of $X$ over $K$, care being taken to preserve the order of the terms from $X$. In what follows we shall take $K$ to be any non-trivial commutative ring and to simplify the notation, take $X$ to be finite, $X = \{x_1,\dots,x_d\}$. Then each element of $K\langle X\rangle$ can be uniquely written as

$$f = \sum \alpha_{i_1 \dots i_r} x_{i_1} \cdots x_{i_r}, \tag{1}$$

where the sum is over all distinct sequences $I = (i_1, \ldots, i_r)$, where $1 \leqslant i_\rho \leqslant d$, with coefficients in $K$ and almost all 0. Writing $I = (i_1, \ldots, i_r)$ and $x_I = x_{i_1} \cdots x_{i_r}$, we can express (1) more briefly as

$$f = \sum \alpha_I x_I. \tag{1}$$

If $g = \sum \beta_J x_J$ is another element of $K\langle X \rangle$, then

$$fg = \sum \alpha_I \beta_J x_I x_J. \tag{2}$$

We can embed $X$ in $K\langle X \rangle$ by identifying $x_i$ with the element (1) for which $\alpha_I = 1$ if $I = i$ and 0 otherwise. Relative to this mapping $K\langle X \rangle$ has the universal property characterizing free algebras. For, given any mapping $\varphi: x_i \to a_i$ of $X$ into a $K$-algebra $A$, we can extend it in just one way to a homomorphism from $K\langle X \rangle$ to $A$, namely by mapping

$$\sum \alpha_I x_I \mapsto \sum \alpha_I a_I,$$

where $a_I = a_{i_1} \cdots a_{i_r}$ if $I = (i_1, \ldots, i_r)$. In the special case where $X$ consists of a single element $x$, the free algebra on $x$ is just the polynomial ring $K[x]$, which of course is commutative, but when $X$ has more than one element, $K\langle X \rangle$ is non-commutative, because e.g. $x_1 x_2 \neq x_2 x_1$.

In the above construction of $K\langle X \rangle$ we formed first products and then sums. It is also possible to carry out these procedures in the opposite order, and it will be instructive to do this in a slightly more general context. Let $U$ be a $K$-module and denote by $U^n$ the $n$-fold tensor product of $U$ with itself over $K$:

$$U^n = U \otimes U \otimes \cdots \otimes U \qquad (n \text{ factors}). \tag{3}$$

In particular, $U^1 = U$ and for $n = 0$ we shall interpret $U^0$ as $K$. By the associative law for tensor products we have for any $r, s \geqslant 0$,

$$U^r \otimes U^s \cong U^{r+s}. \tag{4}$$

We shall use this isomorphism to define a multiplication on the direct sum

$$\mathbf{T}(U) = \bigoplus_{n=0}^{\infty} U^n. \tag{5}$$

In other words $(U^n)$ is a graded ring for the multiplication (3) and $\mathbf{T}(U)$ is the corresponding internally graded ring. This algebra $\mathbf{T}(U)$ is called the *tensor $K$-algebra* on the module $U$. Every element $a \in \mathbf{T}(U)$ is unique of the form $a = \sum a_n$, where $a_n \in U^n$ is the homogeneous component of degree $n$ of $a$. If $b = \sum b_n$, where $b_n \in U^n$, is another element, then $ab = \sum a_m b_n$ and the homogeneous component of degree $n$ of $ab$ is $\sum a_p b_{n-p}$. We state the universal property of $\mathbf{T}(U)$ as

THEOREM 2.1 *Let $U$ be a $K$-module, where $K$ is a commutative ring. Then the tensor algebra $\mathbf{T}(U)$ is the universal $K$-algebra for $K$-linear mappings of $U$*

*into $K$-algebras. Thus there is a homomorphism $\lambda: U \to \mathbf{T}(U)$ such that every $K$-linear mapping $f: U \to A$ into a $K$-algebra $A$ can be uniquely factored by $\lambda$.*

*Proof.* We shall take $\lambda$ to be the embedding which identifies $U$ with its image $U^1$ in $\mathbf{T}(U)$. Given a $K$-linear mapping $f$ of $U$ into a $K$-algebra $A$, we define for each $n \geqslant 1$, a mapping

$$f^n: (u_1, \ldots, u_n) \mapsto (u_1 f) \cdots (u_n f).$$

Since $f$ is $K$-linear, this mapping is multilinear and so gives rise to a linear mapping $f^{(n)}: U^n \to A$. On putting these mappings $f^{(n)}$ together we obtain a mapping $\bar{f}: \mathbf{T}(U) \to A$, which is a homomorphism because it is linear and for any $a = a_1 \otimes \cdots \otimes a_r$, $b = b_1 \otimes \cdots \otimes b_s$ we have $a\bar{f} = a_1 f \cdots a_r f$, $b\bar{f} = b_1 f \cdots b_s f$, and $ab = a_1 \otimes \cdots \otimes a_r \otimes b_1 \otimes \cdots \otimes b_s$, hence

$$\begin{aligned}(ab)\bar{f} &= (a_1 \otimes \cdots \otimes a_r \otimes b_1 \otimes \cdots \otimes b_s)\bar{f} \\ &= (a_1 f) \cdots (a_r f)(b_1 f) \cdots (b_s f) \\ &= (a\bar{f})(b\bar{f}).\end{aligned}$$

Moreover, it is determined by $f$ because it must agree with $f$ on $U^1$ and this set generates $\mathbf{T}(U)$. ■

If $U$ is the free $K$-module on $X = \{x_1, \ldots, x_d\}$, $\mathbf{T}(U)$ clearly coincides with the free algebra $K\langle X \rangle$.

Theorem 2.1 provides us with a 'free' $K$-algebra on a given $K$-module; this means that the coefficients (from $K$) lie in the centre of the algebra, but sometimes we have a more general situation. To describe it we need to generalize the notion of a $K$-algebra. Let $E$ be a general ring (not necessarily commutative); by an *$E$-ring* we understand a ring $R$ which is an $E$-bimodule, such that $x(yz) = (xy)z$ for any $x, y, z$ in $R$ or $E$. Even when $E$ is commutative, an $E$-ring is more general than an $E$-algebra, because the action of $E$ need not centralize the ring. The difference may be succinctly described by saying that whereas a $K$-algebra ($K$ commutative) is a ring $R$ with a homomorphism from $K$ to the centre of $R$, an $E$-ring is a ring $R$ with a homomorphism from $E$ to $R$. Frequently it is convenient to assume that both $E$ and $R$ are $K$-algebras and that the homomorphism $E \to R$ is $K$-linear. In this case $R$ will be called an $E$-ring over $K$ or sometimes an $E_K$-ring. When the coefficient ring $E$ is a $K$-algebra, $K$ is usually embedded in $E$ as a subring, and by an identification we shall assume $K$ to be a subring of $E$. An $E$-bimodule over $K$ or $E_K$-bimodule is understood to be an $E$-bimodule $U$ such that $\alpha u = u\alpha$ for all $u \in U$, $\alpha \in K$.

From any $E_K$-bimodule $U$ we can form a tensor $E_K$-ring $\mathbf{T}(U)$ on $U$ as before. We again define $U^n$ by (3), where the tensor products are over $E$ and $U^0 = E$, $U^1 \cong U$. Now (4) is clear and $\mathbf{T}(U)$ can again be defined by (5), this time as an $E_K$-ring. As before we have

THEOREM 2.2 *Let $E$ be a $K$-algebra ($K$ commutative) and $U$ an $E_K$-bimodule. Then the tensor $E_K$-ring $\mathbf{T}(U)$ is the universal $E_K$-ring for bimodule homomorphisms of $U$ into $E_K$-rings. Thus there is a homomorphism $\lambda: U \to \mathbf{T}(U)$ such that every $E$-bimodule mapping $f: U \to A$ into an $E_K$-ring $A$ can be uniquely factored by $\lambda$.*

*Proof.* The proof is similar to that of Th. 2.1 and may be left to the reader. ■

When $E = K$, this reduces to the case of tensor $K$-algebras considered in Th. 2.1. To give a more general example, for any set $X$ the free $E_K$-bimodule $F$ on $X$ is a direct sum of copies of $E$ indexed by $X$. Now $\mathbf{T}(F)$, the tensor $E_K$-ring on $F$, also denoted by $E_K\langle X\rangle$, may be described as the ring generated by $E$ and $X$ with defining relations

$$\alpha x = x\alpha \qquad \text{for all } x \in X, \alpha \in K. \tag{6}$$

For simplicity we shall often suppress explicit reference to $K$ in what follows. As in Ex. (11) of **1.3** we can easily verify that the construction of $\mathbf{T}(U)$ from $U$ is in fact a functor from $E$-bimodules to $E$-rings. If $\alpha: U \to V$ is a homomorphism of $E$-bimodules and $\lambda: V \to \mathbf{T}(V)$ is the canonical embedding, then $\alpha\lambda$ is a homomorphism from $U$ to $\mathbf{T}(V)$ and by the universal property of $\mathbf{T}(U)$ we obtain a homomorphism $\mathbf{T}(\alpha): \mathbf{T}(U) \to \mathbf{T}(V)$, which is easily seen to possess the functorial property.

As an application of Th. 2.2 let us consider derivations. A *derivation* over $E$ on an $E$-ring $A$ is a linear mapping $\delta$ of $A$ into an $A$-bimodule such that

$$(xy)^\delta = x \cdot y^\delta + x^\delta \cdot y \qquad (x, y \in A).$$

For example, the polynomial ring $K[x]$ has the derivation $f \mapsto f'$, where $f' = \mathrm{d}f/\mathrm{d}t$ is the usual derivative, already encountered in A.1, **6.8**. It is often useful to consider a slightly more general situation. Let $\alpha: C \to A, \beta: C \to B$ be two homomorphisms between $E$-rings and let $M$ be an $(A, B)$-bimodule. By an $(\alpha, \beta)$-*derivation* over $E$ we understand a mapping $\delta: C \to M$ which is linear and satisfies

$$(xy)^\delta = x^\alpha y^\delta + x^\delta y^\beta \qquad \text{for all } x, y \in C. \tag{7}$$

When $A = B = C$ and $\alpha = \beta = 1$, this reduces to the previous case. If we put $A = B = M = E = K, C = K[t]$ and for a fixed $p \in K$ define $\alpha: C \to K$ as the evaluation mapping $c \mapsto c(p)$, then the mapping $c \mapsto c'(p)$ (the derivative of $c$, evaluated at $p$) is an $(\alpha, \alpha)$-derivation.

We note that any $(\alpha, \beta)$-derivation maps 1 into 0. For if we put $x = y = 1$ in (7) and observe that $1^\alpha = 1^\beta = 1$, we find that $1^\delta = 0$. It is easily checked that the kernel of $\delta$ is a subalgebra of $C$, called the algebra of $\delta$-*constants*.

Given an $(A, B)$-bimodule $M$ and homomorphisms $\alpha: C \to A, \beta: C \to B$, each $m \in M$ defines a mapping

$$x \mapsto x^\alpha m - m x^\beta \qquad x \in C. \tag{8}$$

This is easily checked to be an $(\alpha, \beta)$-derivation; it is called the *inner derivation* induced by $m$. An *outer derivation* is one that is not inner.

The study of derivations can be reduced to that of homomorphisms as follows. Let $\begin{pmatrix} A & M \\ 0 & B \end{pmatrix}$ be the ring of all matrices

$$\begin{pmatrix} a & x \\ 0 & b \end{pmatrix} \qquad a \in A, b \in B, x \in M,$$

with the usual matrix addition and multiplication. The $(A, B)$-bimodule property ensures that we get an $E$-ring in this way:

$$\begin{pmatrix} a & x \\ 0 & b \end{pmatrix}\begin{pmatrix} a' & x' \\ 0 & b' \end{pmatrix} = \begin{pmatrix} aa' & ax' + xb' \\ 0 & bb' \end{pmatrix},$$

and the associative law holds because in a product of three factors,

$$a(a'x'' + x'b'') + x(b'b'') = (aa')x'' + (ax' + xb')b''.$$

Now it is easily verified that the mapping of $C$ into $\begin{pmatrix} A & M \\ 0 & B \end{pmatrix}$ defined by

$$x \mapsto \begin{pmatrix} x^{\alpha} & x^{\delta} \\ 0 & x^{\beta} \end{pmatrix} \qquad x \in C, \tag{9}$$

is an $E$-ring homomorphism iff $\alpha, \beta$ are $E$-ring homomorphisms and $\delta$ is an $(\alpha, \beta)$-derivation over $E$.

Let $U$ be any $E$-module and let $\alpha: U \to A, \beta: U \to B$ be $E$-linear mappings into $E$-rings $A, B$. By Th. 2.2 they extend to unique homomorphisms $\alpha', \beta'$ of $\mathbf{T}(U)$ into $A, B$ respectively. Further, let $\delta: U \to M$ be an $E$-linear mapping of $U$ into an $(A, B)$-bimodule $M$. Then for $x \in U$, (9) defines an $E$-linear mapping of $U$ into $\begin{pmatrix} A & M \\ 0 & B \end{pmatrix}$ which again extends to a unique homomorphism of $\mathbf{T}(U)$ into $\begin{pmatrix} A & M \\ 0 & B \end{pmatrix}$. Suppose that $x \in \mathbf{T}(U)$ maps to $\begin{pmatrix} x_1 & x_2 \\ 0 & x_3 \end{pmatrix}$ (it is easily seen that the (2, 1)-entry must be 0). Then $x \mapsto x_1$ is a homomorphism of $\mathbf{T}(U)$ into $A$ extending $\alpha$, hence it must be $\alpha'$. Similarly $x \mapsto x_3$ must be $\beta'$, while $x \mapsto x_2$ is an $(\alpha', \beta')$-derivation $\delta'$ say, by the above remarks. This derivation extends $\delta$ and is uniquely determined by it. Hence we have

PROPOSITION 2.3 *Let $U$ be an $E$-bimodule, let $A, B$ be any $E$-rings and $M$ an $(A, B)$-bimodule. Given any $E$-linear mappings $\alpha: U \to A, \beta: U \to B, \delta: U \to M$, there exist unique homomorphisms $\alpha': \mathbf{T}(U) \to A, \beta': \mathbf{T}(U) \to B$ extending $\alpha$ and $\beta$ respectively, and there is a unique $(\alpha', \beta')$-derivation over $E, \delta': \mathbf{T}(U) \to M$, extending $\delta$.* ■

The result can be applied to free algebras by taking $E = K$ and $U$ the free $K$-module on $x_1, \ldots, x_d$. Another important special case is that where $E = K$, $M = A = B = \mathbf{T}(U)$, $\beta = 1$. The conclusion then takes the following form:

COROLLARY 2.4 *Let $U$ be a $K$-module and $\mathbf{T}(U)$ its tensor algebra. Given any $K$-linear mappings $\alpha: U \to \mathbf{T}(U)$, $\delta: U \to \mathbf{T}(U)$, there is a unique endomorphism $\alpha'$ of $\mathbf{T}(U)$ extending $\alpha$ and a unique $(\alpha', 1)$-derivation $\delta'$ of $\mathbf{T}(U)$ extending $\delta$.* ■

Any derivation followed by a module homomorphism is again a derivation and we can ask whether there is a module with a derivation which is universal. For simplicity we shall assume that $A = B = C = R$ say and $\alpha = \beta = 1$. Thus $R$ is an $E$-ring and we are dealing with derivations from $R$ to $R$-bimodules. These derivations form the objects of a category in which the morphisms are commutative triangles and we are looking for an initial object in this category, i.e. an $R$-bimodule $\Omega$ with a derivation $\delta: R \to \Omega$ such that every derivation from $R$ can be uniquely factored by $\delta$. Such a bimodule indeed exists and can be described explicitly.

PROPOSITION 2.5 (Eilenberg) *Let $R$ be any $E$-ring with multiplication mapping $\mu: R \otimes_E R \to R$. This mapping gives rise to an exact sequence*

$$0 \to \Omega \to R \otimes_E R \xrightarrow{\mu} R \to O, \tag{10}$$

*where $\Omega = \ker \mu$ is an $R$-bimodule, generated as left (or right) $R$-module by the elements $x \otimes 1 - 1 \otimes x$ $(x \in R)$.*

This $R$-bimodule $\Omega$ is the *universal derivation bimodule* for $R$, with the universal derivation

$$\delta: x \mapsto x \otimes 1 - 1 \otimes x \qquad (x \in R). \tag{11}$$

*Proof.* It is clear that the multiplication $\mu: x \otimes y \mapsto xy$ is an $R$-bimodule homomorphism and its kernel $\Omega$ contains the elements $x \otimes 1 - 1 \otimes x$. Suppose that $\sum x_i \otimes y_i \in \Omega$; then $\sum x_i y_i = 0$ in $R$ and so

$$\sum x_i \otimes y_i = \sum (x_i \otimes 1 - 1 \otimes x_i) y_i = \sum x_i (1 \otimes y_i - y_i \otimes 1).$$

This shows $\Omega$ to be generated by the elements $x \otimes 1 - 1 \otimes x$ as left or right $R$-module.

Now the mapping (11) is a derivation:

$$\begin{aligned}(xy)^\delta &= xy \otimes 1 - 1 \otimes xy = x(y \otimes 1 - 1 \otimes y) + (x \otimes 1 - 1 \otimes x)y \\ &= x \cdot y^\delta + x^\delta y.\end{aligned}$$

It is universal, for if $d: R \to M$ is any derivation, we can define a homomorphism $f: \Omega \to M$ as follows: if $\sum x_i \otimes y_i \in \Omega$, then $\sum x_i y_i = 0$, hence $\sum x_i^d \cdot y + \sum x_i \cdot y_i^d = 0$

and so we may put

$$(\sum x_i \otimes y_i)f = \sum x_i^d \cdot y_i = -\sum x_i \cdot y_i^d. \tag{12}$$

This is an $R$-bimodule homomorphism, since $(\sum a(x_i \otimes y_i))f = -\sum a x_i \cdot y_i^d$, $(\sum (x_i \otimes y_i)a)f = \sum x_i^d \cdot y_i a$. Moreover, $\delta f : x \mapsto x \otimes 1 - 1 \otimes x \mapsto x^d$. Hence $d = \delta f$ and $f$ is unique since it is prescribed on the generating set $\{x \otimes 1 - 1 \otimes x\}$. ■

The module $\Omega$ in (10), regarded as universal derivation bimodule of $R$, is often denoted by $\Omega_E(R)$. We shall be particularly interested in this bimodule when $R$ is the tensor $E$-ring on an $E$-bimodule $U$, $R = \mathbf{T}_E(U)$. In that case we can give an explicit description.

PROPOSITION 2.6 *Let $E$ be a $K$-algebra, $U$ an $E$-bimodule and $R = \mathbf{T}_E(U)$ the tensor $E$-ring on $U$. Then the universal derivation bimodule of $R$ is given by*

$$\Omega_E(R) = R \otimes U \otimes R \tag{13}$$

*and the exact sequence*

$$0 \to R \otimes U \otimes R \xrightarrow{\alpha} R \otimes R \xrightarrow{\mu} R \to 0 \tag{14}$$

*is split exact as sequence of left (or right) $R$-modules.*

Here and in the proof that follows, all tensor products are understood over $E$.

*Proof.* As before, we write $\Omega = \ker \mu$ and consider the canonical derivation $\delta : R \to \Omega$ given by $x^\delta = 1 \otimes x - x \otimes 1$. We saw in the proof of Prop. 2.5 that $\Omega$ is generated as left or right $R$-module by the $x^\delta$ as $x$ ranges over $R$. But $R$ is generated by $U$ as $E$-ring, therefore $\Omega$ is generated as $E$-bimodule by the $u^\delta$, where $u \in U$. Thus the restriction map $\delta | U : U \to \Omega$ gives rise to an $R$-bimodule map $\alpha : R \otimes U \otimes R \to R \otimes R$ such that $(1 \otimes u \otimes 1)^\alpha = u^\delta$ and clearly $\operatorname{im} \alpha = \Omega$. We complete the proof by finding a left $R$-linear map $\beta : R \otimes R \to R \otimes U \otimes R$ such that $\alpha\beta = 1$. Consider the $E$-bimodule map from $U$ to $R \otimes U \otimes R$ given by $u \mapsto 1 \otimes u \otimes 1$. By Cor. 2.4 this extends to a derivation over $E$, $d : R \to R \otimes U \otimes R$. Since $d$ is left $E$-linear, it induces a left $R$-linear map $\beta : R \otimes R \to R \otimes U \otimes R$, where $(x \otimes y)^\beta = x \cdot y^d$. For any $u \in U, x \in R$ we have

$$\begin{aligned}(1 \otimes u \otimes x)^{\alpha\beta} &= (1 \otimes ux - u \otimes x)^\beta = (ux)^d - u \cdot x^d \\ &= u^d x = 1 \otimes u \otimes x,\end{aligned}$$

and since the elements $1 \otimes u \otimes x$ generate $R \otimes U \otimes R$ as left $R$-module, we see that $\alpha\beta = 1$. This shows (14) to be split exact, as claimed. In a similar way the mapping $x \otimes y \mapsto x^d y$ is a right $R$-linear map splitting $\alpha$. ■

To construct free objects in other varieties of algebras it is often simplest to

take free algebras and apply the factor theorem. We illustrate the method by the case of symmetric algebras, which is of some interest in itself.

We begin by describing the process of 'abelianizing' a ring, which is analogous to the corresponding notion for groups (cf. A.1, **9.5**).

THEOREM 2.7 *To any ring $R$ there corresponds a commutative ring $R^{ab}$ with a homomorphism $v: R \to R^{ab}$ which is universal for homomorphisms of $R$ into commutative rings. Thus each homomorphism $f$ from $R$ to a commutative ring can be uniquely factored by $v$.*

*Proof.* Let $\mathfrak{c}$ be the *commutator ideal* of $R$, i.e. the ideal generated by all the commutators $xy - yx$, where $x, y \in R$, and write $R^{ab} = R/\mathfrak{c}$, with the natural homomorphism $v: R \to R^{ab}$. Then any homomorphism $f$ from $R$ to a commutative ring maps $xy - yx$ to 0, for all $x, y \in R$, hence $\ker f \supseteq \mathfrak{c}$, and so by the factor theorem (Th. 1.2.4) $f$ can be factored uniquely by $v$. ■

We remark that if $X$ is a generating set of $R$, then the commutator ideal of $R$ is already generated by the elements $xy - yx$ for all $x, y \in X$. For let $\mathfrak{b}$ be the ideal generated by these commutators and write $\lambda: R \to R/\mathfrak{b}$ for the natural mapping. Then $R/\mathfrak{b}$ is generated by the elements $x\lambda, x \in X$. Now fix $x_1 \in X$; $x_1\lambda$ commutes with every $y\lambda, y \in X$, so the centralizer of $x_1\lambda$ contains a generating set of $R/\mathfrak{b}$ and hence is the whole ring, i.e. each $x_1\lambda$ lies in the centre of $R/\mathfrak{b}$. Hence the centre of $R/\mathfrak{b}$ contains a generating set and so is the whole ring, i.e. $R/\mathfrak{b}$ is commutative. This means that $\mathfrak{b} = \ker \lambda \supseteq \mathfrak{c}$, hence $\mathfrak{b} = \mathfrak{c}$ as claimed.

Given a $K$-module $U$, let $\mathbf{T}(U)$ be its tensor algebra. We define the *symmetric algebra* on $U$ as the algebra

$$\mathbf{S}(U) = \mathbf{T}(U)^{ab}.$$

By what has just been said, we see that $\mathbf{S}(U)$ can be obtained from $\mathbf{T}(U)$ by imposing the relations

$$xy = yx \qquad \text{for all } x, y \in U, \tag{15}$$

because $\mathbf{T}(U)$ is generated by $U$. It is clear from the definition that $\mathbf{S}(U)$ is universal for $K$-linear mappings from $U$ to commutative $K$-algebras, so we have

THEOREM 2.8 *For any $K$-module $U$ there is a commutative $K$-algebra $\mathbf{S}(U)$ with a $K$-linear mapping $\mu: U \to \mathbf{S}(U)$ which is universal for $K$-linear mappings from $U$ to commutative $K$-algebras. $\mathbf{S}(U)$ is obtained from the tensor algebra $\mathbf{T}(U)$ by adjoining the relations* (15). ■

Let $\delta: U \to \mathbf{S}(U)$ be any $K$-linear mapping. We ask: when can $\delta$ be extended to a derivation of $\mathbf{S}(U)$? The answer is 'always'. For what we require is a

homomorphism from $\mathbf{S}(U)$ to $\begin{pmatrix} \mathbf{S}(U) & \mathbf{S}(U) \\ 0 & \mathbf{S}(U) \end{pmatrix}$ extending the mapping

$$f: u \mapsto \begin{pmatrix} u & u^{\delta} \\ 0 & u \end{pmatrix}.$$

Now a simple verification shows that $uf$ and $vf$ commute, for any $u, v \in U$, so by the universal property of $\mathbf{S}(U)$ there is a unique $K$-algebra homomorphism from $\mathbf{S}(U)$ extending $f$. This proves

PROPOSITION 2.9 *Let $U$ be a $K$-module and $\mathbf{S}(U)$ its symmetric algebra. Then any $K$-linear mapping $\delta: U \to \mathbf{S}(U)$ extends to a unique derivation of $\mathbf{S}(U)$.* ■

For an example of a symmetric algebra, let $U$ be the free $K$-module on $x_1, \ldots, x_d$; then $\mathbf{S}(U)$ is just the polynomial ring $K[x_1, \ldots, x_d]$. For this polynomial ring clearly satisfies the universal property expressed by Th. 2.8, and as we know, this characterizes $\mathbf{S}(U)$ up to isomorphism.

## Exercises

(1) Verify that $U \mapsto \mathbf{T}(U)$ is a functor. Likewise for $\mathbf{S}(U)$.

(2) Show that $\mathbf{S}(U \oplus V) \cong \mathbf{S}(U) \otimes \mathbf{S}(V)$.

(3) Given a surjective homomorphism $\mu: U \to V$ of $K$-modules, show that $\mathbf{T}(\mu): \mathbf{T}(U) \to \mathbf{T}(V)$ is surjective. Likewise for $\mathbf{S}(U)$.

(4) Let $K$ be an integral domain. By considering degrees, show that $K\langle X \rangle$ is again an integral domain.

(5) Prove Leibniz's formula for $(1,1)$-derivations:

$$(ab)\delta^n = \sum \binom{n}{i} (a\delta^i)(b\delta^{n-i}).$$

More generally, show that for a $(1, \beta)$-derivation $\delta$,

$$(ab)\delta^n = \sum a\delta^i \cdot b f_i^n(\beta, \delta),$$

where $f_i^n(\beta, \delta)$ is the coefficient of $t^i$ in the expansion of $(t\beta + \delta)^n$.

(6) Writing $[x, y] = xy - yx$, verify the identity $[xy, z] = x[y, z] + [x, z]y$, which expresses the fact that the mapping $u \mapsto [u, z]$ is a derivation. Use this result to give another proof of the remark following Th. 2.7, that the commutator ideal of a ring $R$ is generated by all $[x, y]$, where $x, y$ range over a generating set of $R$.

(7) Let $U$ be a $K$-module and $\alpha$ an endomorphism of $\mathbf{S}(U)$. Show that any mapping $\delta: U \to \mathbf{S}(U)$ extends to an $(\alpha, \alpha)$-derivation of $\mathbf{S}(U)$.

(8) Let $\alpha: C \to A, \beta: C \to B$ be $K$-algebra homomorphisms and $M$ an $(A, B)$-bimodule. Show that the $(\alpha, \beta)$-derivations into $M$ form a $K$-module $\mathrm{Der}(C, M)$. Taking $C = K\langle X \rangle$, use Prop. 2.3 to deduce that $\mathrm{Der}(K\langle X \rangle, M) \cong M^X$.

(9) Let $K$ be non-trivial and suppose that $K\langle X \rangle$ and $K\langle Y \rangle$ are isomorphic as $K$-algebras. Use Ex. (8) to show that $X$ and $Y$ are equipotent. ($|X|$ is called the *rank* of the free algebra $k\langle X \rangle$; this exercise shows it to be uniquely determined by $K\langle X \rangle$.)

(10) Show that for any commutative ring $K$, the free $K$-algebra $K\langle X \rangle$ admits a unique antiautomorphism leaving $K$ and $X$ elementwise fixed. More generally, show that any antiautomorphism of a $K$-algebra $E$ leaving $K$ elementwise fixed extends to an antiautomorphism of $E_K\langle X \rangle$ leaving $X$ elementwise fixed.

(11) Find extensions of Prop. 2.5, 6 to $(\alpha, \beta)$-derivations.

## 2.3 The Hilbert series of a graded algebra or module

In the study of finite-dimensional spaces or algebras over a field the dimension provides a useful comparison, based on the formula for a linear mapping $f$ from a space $U$:

$$\dim \ker f + \dim \mathrm{im}\ f = \dim U. \tag{1}$$

It is an easy consequence that for a short exact sequence of vector spaces

$$0 \to V' \to V \to V'' \to 0 \tag{2}$$

we have the formula

$$\dim V = \dim V' + \dim V''. \tag{3}$$

Let us generally define an *additive function* on a class of additive groups as a $\mathbf{Z}$-valued function $\lambda$ such that for any short exact sequence (2), we have

$$\lambda(V) = \lambda(V') + \lambda(V''). \tag{4}$$

As (3) shows, the dimension is an additive function; more generally, for modules of finite composition length the length is an additive function. The defining property extends to longer sequences as follows.

PROPOSITION 3.1 *Let $\lambda$ be an additive function on modules. Then for any exact sequence of modules*

$$0 \to M_0 \to M_1 \to \cdots \to M_r \to 0 \tag{5}$$

*we have*

$$\sum (-1)^i \lambda(M_i) = 0.$$

*Proof.* Let $f_i: M_i \to M_{i+1}$ be the maps in (5); by exactness we have $\ker f_i = \operatorname{im} f_{i-1} = N_i$ say, where $N_0 = N_{r+1} = 0$. From the short exact sequence $0 \to N_i \to M_i \to N_{i+1} \to 0$ we obtain $\lambda(M_i) = \lambda(N_i) + \lambda(N_{i+1})$, and the result follows on taking the alternating sum. ■

When we are dealing with a positively graded module $M = \oplus M_n$ over a graded ring we have an infinite sequence $\lambda(M_n)$ of numbers and we define the *Hilbert series*, also called the *Poincaré series*, of $M$ as the formal power series

$$H(M) = \sum \lambda(M_n) t^n.$$

It is a remarkable fact that for finitely generated modules over a commutative Noetherian ring this function is rational in $t$, for any functional $\lambda$.

THEOREM 3.2 (Hilbert–Serre) *Let $A = \oplus A_n$ be a positively graded commutative Noetherian ring, generated as $A_0$-algebra by $x_1, \ldots, x_d$, homogeneous of positive degrees $r_1, \ldots, r_d$ respectively. If $M = \oplus M_n$ is any finitely generated positively graded $A$-module, then each homogeneous component $M_n$ is finitely generated as $A_0$-module and the Hilbert series of $M$, for any additive functional $\lambda$, has the form*

$$H(M) = \frac{f(t)}{\prod_{i=1}^{d} (1 - t^{r_i})} \qquad \text{where } f \in \mathbf{Z}[t]. \tag{6}$$

*Proof.* By hypothesis, $M$ is finitely generated over $A$, by $u_1, \ldots, u_e$ say, where each $u_i$ may be taken homogeneous, of degree $s_i$ say. Any $v \in M_n$ has the form $v = \sum u_i a_i$, where $a_i$ is homogeneous of degree $n - s_i$, hence $M_n$ is spanned over $A_0$ by all $u_i c_i$, where $c_i$ runs over all products of $x_1, \ldots, x_d$ of degree $n - s_i$.

To prove (6) we have to show that $\prod(1 - t^{r_i}) H(M)$ is a polynomial in $t$. We shall use induction on $d$, the number of generators of $A$ over $A_0$. When $d = 0, A = A_0$ and $M$ is a finitely generated $A_0$-module, hence $M_n = 0$ for all large $n$, so $H(M)$ is a polynomial in $t$ and the result holds.

Now assume that $d > 0$ and that the result holds for $d - 1$. Multiplication by $x_d$ is a linear mapping of $M$ of degree $r = r_d$, so we have an exact sequence

$$0 \to N_n \to M_n \xrightarrow{x_d} M_{n+r} \to L_{n+r} \to 0. \tag{7}$$

Write $L = \oplus L_n$, $N = \oplus N_n$; these modules are a quotient and submodule of $M$ respectively and so are finitely generated. Both are annihilated by $x_d$ and so are $A_0[x_1, \ldots, x_{d-1}]$-modules. Applying $\lambda$ to (7) we obtain

$$\lambda(N_n) - \lambda(M_n) + \lambda(M_{n+r}) - \lambda(L_{n+r}) = 0.$$

If we multiply by $t^{n+r}$ and sum over $n$, the first term gives $t^r H(N)$; similarly for

the second term, while the third term gives $H(M)$ except for the first $r$ terms; thus we obtain $H(M)$ by adding a suitable polynomial. The same applies to the fourth term and so we have altogether $t^rH(N) - t^rH(M) + H(M) - H(L) = g$, i.e.

$$(1 - t^r)H(M) + [t^rH(N) - H(L)] = g \tag{8}$$

where $g$ is a polynomial in $t$. Now $N, L$ are $A_0[x_1, \ldots, x_{d-1}]$-modules, as we saw; by induction on $d$, $\prod_{i=1}^{d-1}(1 - t^{r_i})[H(L) - t^rH(N)]$ is a polynomial, hence by (8), so is $\prod_{i=1}^{d}(1 - t^{r_i})H(M)$, and (6) follows. ■

Consider the special case where all the $x_i$ are of degree 1, thus $r_i = 1$ for all $i$:

COROLLARY 3.3 *If in Th. 3.2. $x_1, \ldots, x_d$ are all of degree 1 and $\delta = \delta(M)$ is the order of the pole of $H(M)$ at $t = 1$, then $\delta \leqslant d$ and there exists a unique polynomial $h$ of degree $\delta - 1$ over* $\mathbf{Q}$ *such that $\lambda(M_n) = h(n)$ for all large $n$.*

*Proof.* By Th. 3.2 we have $H(M) = (1 - t)^{-d}f(t)$. By cancelling any power of $1 - t$ dividing $f$, we can write this as $H(M) = (1 - t)^{-\delta}g(t)$, where $g(1) \neq 0$. Thus if $g = \sum a_i t^i$, then $a = \sum a_i \neq 0$. Since $(1 - t)^{-\delta} = \sum \binom{\delta + j - 1}{\delta - 1} t^j$, we have $H(M) = \sum_{ij} \binom{\delta + j - 1}{\delta - 1} a_i t^{i+j}$; hence if $g$ is of degree $N$, then

$$\lambda(M_n) = \sum_{i=0}^{N} a_i \binom{n + \delta - i - 1}{\delta - 1} \qquad \text{for } n \geqslant N.$$

The sum on the right is a polynomial $h$ in $n$; the $i$th term in the sum has leading term $a_i n^{\delta - 1}/(\delta - 1)!$, hence the leading term of $h$ is $an^{\delta - 1}/(\delta - 1)!$, and this is non-zero, by hypothesis. Clearly $h$ is unique, since its value is prescribed for all $n \geqslant N$. ■

When the base ring $A_0$ is Artinian as well as Noetherian, each component $M_i$ is of finite composition length $l(M_i)$. This length function is clearly additive and we may apply Th. 3.2 and Cor. 3.3 with $\lambda = l$. In that case the polynomial $h$ of Cor. 3.3 is called the *Hilbert polynomial* or *characteristic polynomial* of $M$. We note that it is integer-valued for integer arguments, but not necessarily with integer coefficients. It can be shown that $h(\xi)$ is a linear combination of binomial coefficients $\binom{\xi}{r}$, $r = 0, 1, \ldots, \delta - 1$ (cf. Ex. (2, 3)).

To illustrate Cor. 3.3 let us take $k$ to be a field and $\lambda(M)$ the dimension of $M$ as vector space over $k$. For the polynomial ring $A = k[x_1, \ldots, x_d]$, $\lambda(A_n)$ is the number of products of the $x$'s of degree $n$. A typical such product is

$$x_1^{\nu_1} \cdots x_d^{\nu_d}. \tag{9}$$

We form (9) by taking a row of $n+d-1$ blank squares and putting down $x_1$ in the first $v_1$ squares, then a dot, then $x_2$ in the next $v_2$ squares etc. Such a product is uniquely determined by choosing the $d-1$ squares to contain the dots, hence their number is

$$\lambda(A_n) = \binom{n+d-1}{d-1},$$

and the Hilbert series for $A_n$ is

$$H(A) = \sum_n \binom{n+d-1}{d-1} t^n = (1-t)^{-d}. \tag{10}$$

Let us take $d=3$ and as our graded module $M = A/(x_1^2 + x_2^2 - x_3^2)$. The component $M_n$ for $n \geqslant 2$ is spanned by $x_1^r x_2^{n-r}$ and $x_1^r x_2^{n-r-1} x_3$, in all $2n+1$ monomials, hence $\lambda(M_n) = 2n+1$ and this holds for $n = 0, 1$ as well. Thus $H(M) = \sum(2n+1)t^n = 2(1-t)^{-2} - (1-t)^{-1} = (1+t)(1-t)^{-2}$. Here $\delta(M)$ is 2; it can be shown that in general $\delta(M)$ is the dimension of the variety whose ideal is the annihilator of $M$, in this case the cylinder $x_1^2 + x_2^2 = x_3^2$ (cf. Hartshorne [1977]).

The Hilbert series (10) for the polynomial ring can be obtained more simply by using

PROPOSITION 3.4 *If $A, B$ are any graded $k$-algebras (over a field $k$), and $\lambda$ is the dimension function, then the Hilbert series satisfies*

$$H(A \otimes B) = H(A)H(B).$$

*Proof.* Put $C = A \otimes B$; since $C_n = \sum A_i \otimes B_{n-i}$, we have $\lambda(C_n) = \sum \lambda(A_i)\lambda(B_{n-i})$, and so $H(C) = \sum H(C_n)t^n = \sum \lambda(A_i)\lambda(B_{n-i})t^n = H(A)H(B)$. ■

It is clear that for a polynomial ring in one variable we have $H(k[x]) = \sum t^n = (1-t)^{-1}$, hence $H(k[x_1, \ldots, x_d]) = (1-t)^{-d}$.

For the free algebra $F = k\langle x_1, \ldots, x_d \rangle$ it is clear that $\lambda(F_n) = d^n$, hence (cf. also Ex. (1)):

$$H(F) = \sum d^n t^n = (1-dt)^{-1}. \tag{11}$$

Another important result (due to Golod and Shafarevich, who use it to construct infinite 'class-field towers' in algebraic number theory) is an estimate for the Hilbert series of an algebra, which in some cases enables one to recognize from the presentation that the algebra is infinite-dimensional.

Let $k$ be a field and $A$ any $k$-algebra, generated by $x_1, \ldots, x_d$ say. The defining relations will be polynomials in the $x$'s, and if $A$ is graded, these defining relations may be taken to be homogeneous. We may always assume that there are no relations of degree 1, for if $\sum \alpha_i x_i = 0$ is a non-trivial relation, we can make a

linear transformation to new variables $y_1, \ldots, y_d$ such that $y_1 = \sum \alpha_i x_i$. Then $y_1 = 0$ is one of the defining relations, so $A$ is already generated by $y_2, \ldots, y_d$. Thus a graded algebra always has a set of defining relations which are homogeneous of degree $\geqslant 2$. Now our intuition tells us that the fewer relations there are, the larger $\dim A$ will be, and here relations of low degree carry more weight than those of higher degree. How can these ideas be made precise to give a usable estimate? The theorem of Golod and Shafarevich provides a very satisfactory answer. To state it we shall use the following convention: we compare two power series over $\mathbf{Q}$ by writing $\sum \alpha_n t^n \geqslant \sum \beta_n t^n$ to mean: $\alpha_n \geqslant \beta_n$ for all $n$.

THEOREM 3.5 (Golod–Shafarevich) *Let $F = k\langle x_1, \ldots, x_d \rangle$ be the free algebra on $d$ generators over a field $k$, graded by the degree in the $x$'s, and let $A = F/\mathfrak{a}$ be the quotient of $F$ by an ideal $\mathfrak{a}$ with a homogeneous generating set of $r_n$ elements of degree $n$ ($n \geqslant 2$). Then $A$ is infinite-dimensional, provided that*

$$(1 - dt + \sum r_n t^n)^{-1} \geqslant 1. \tag{12}$$

We observe that the series in brackets has integer coefficients and constant term 1, hence its inverse is again a series with integer coefficients.

*Proof.* (Vinberg) Assume that (12) holds and write $r = \sum r_n t^n$; we first show that $g = (1 - dt + r)^{-1}$ is not a polynomial. Suppose that $g$ is a polynomial; then $(1 - dt + r)g = 1$, and this shows that $r$ cannot be a polynomial. Now

$$(1 + r)g = 1 + dtg, \tag{13}$$

and $r \geqslant 0$ by definition, while the same is true of $g$, by (12). Moreover, as we saw, $r$ is not a polynomial, hence neither is $(1 + r)g$, whereas the right-hand side of (13) clearly is a polynomial. This contradiction shows that $g$ cannot be a polynomial.

Let $V$ be the vector space spanned by the defining relations of $\mathfrak{a}$, so by definition, $H(V) = r = \sum r_n t^n$, and since $H(F) = (1 - dt)^{-1}$, the inequality (12) takes the form

$$(H(F)^{-1} + H(V))^{-1} \geqslant 1. \tag{14}$$

Take a graded complement $B$ of $\mathfrak{a}$ in $F$ (as vector space):

$$F = \mathfrak{a} \oplus B. \tag{15}$$

We claim that

$$\mathfrak{a} = \mathfrak{a}F_1 + BV. \tag{16}$$

For any element of $\mathfrak{a}$ has the form $\sum f_\lambda p_\lambda g_\lambda$, where $f_\lambda, g_\lambda \in F, p_\lambda \in V$ and $f_\lambda, g_\lambda, p_\lambda$ are all homogeneous. If $\deg g_\lambda > 0$, we can write $g_\lambda = \sum h_i x_i$ and then find that $f_\lambda p_\lambda g_\lambda = \sum f_\lambda p_\lambda h_i x_i \in \mathfrak{a}F_1$. If $\deg g_\lambda = 0$, the term reduces to $f_\lambda p_\lambda$ and this lies in $FV$, but $FV = \mathfrak{a}V + BV$ and $\mathfrak{a}V \subseteq \mathfrak{a}FF_1 \subseteq \mathfrak{a}F_1$. Thus (16) follows. Comparing

dimensions, we find

$$H(\mathfrak{a}) \leqslant H(\mathfrak{a})dt + H(B)H(V),$$

i.e.

$$H(\mathfrak{a})(1 - dt) \leqslant H(B)H(V).$$

Using the equation $H(F) = H(\mathfrak{a}) + H(B)$ (from (15)) to eliminate $H(\mathfrak{a})$ and remembering (11), we obtain

$$(H(F) - H(B))H(F)^{-1} \leqslant H(B)H(V),$$

which simplifies to

$$H(B)(H(F)^{-1} + H(V)) \geqslant 1. \tag{17}$$

If the left-hand side of (17) is denoted by $G$, then we can write this as

$$H(B) = G(H(F)^{-1} + H(V))^{-1}. \tag{18}$$

By (17), $G \geqslant 1$, while the second factor on the right of (18) is $\geqslant 1$ by (14); moreover $H(B)$ is not a polynomial, for as (18) shows, it is the product of two series $\geqslant 1$, of which the second is not a polynomial. Therefore $H(B)$ is not a polynomial, and since $H(A) = H(B)$, by (15) and the definition of the algebra $A$, it follows that $A$ is infinite-dimensional. ■

The main application of this result in algebra is the construction of a finitely generated nilalgebra which is not nilpotent, i.e. an algebra $A$ (without 1) with a finite generating set, such that every element of $A$ is nilpotent, but $A^n \neq 0$ for all $n$ (Ex. (5)). This answers a question raised by Kurosh in 1941. By an adaptation of this example a finitely generated infinite $p$-group can be constructed, thus answering an earlier question of Burnside in 1902, see Ex. (6).

## Exercises

(1) Let $F = k\langle x_1, \ldots, x_d \rangle$. Show that $F = k \oplus \sum x_i F$; deduce that $H(F) = 1 + dtH(F)$ and hence again show that $H(F) = (1 - dt)^{-1}$. What is the corresponding formula when $x_i$ is of degree $i$?

(2) Define the differencing operator $\Delta$ by $\Delta f(x) = f(x+1) - f(x)$. Show that $\Delta\binom{x}{n} = \binom{x}{n-1}$ and generally, for any polynomial $f$ of degree $n$, $\Delta f$ has degree $n - 1$.

(3) Show that a polynomial over $\mathbf{Q}$ has integer values for all integer arguments iff it is an integral linear combination of binomial coefficients $\binom{x}{n}$. Show that this holds even for polynomials which take integer values for all integer arguments $\geqslant N$, for some $N$. (Hint: use Ex. (2) and induction.)

(4) (Golod) With the notation of Th. 3.5, show that if for some $\varepsilon$,

$$0 < 2\varepsilon < d \qquad r_n \leqslant \varepsilon^2(d-2\varepsilon)^{n-2} \tag{1}$$

then

$$(1 - dt + \sum r_n t^n)^{-1} \geqslant \frac{1-(d-2\varepsilon)t}{(1-(d-\varepsilon)t)^2}.$$

Show that this expression has positive coefficients, and deduce that $F/\mathfrak{a}$ is infinite-dimensional whenever $r_n$ satisfies (1). (Note that for fixed $n$, the estimate (1) for $r_n$ is greatest when $\varepsilon = d/n$.)

(5) (Golod) Let $F = k\langle x_1, \ldots, x_d\rangle$, where $d \geqslant 2$ and denote the augmentation ideal by $F_+$. Show that the general polynomial $g$ in $F_+$ of degree $v$ is a linear combination of $q = d + d^2 + \cdots + d^v$ monomials, and that $g^N$ can be written as a linear combination of $\binom{N+q-1}{q-1}$ forms, where these forms are independent of the coefficients of $g$, and have degrees between $N$ and $vN$.

Deduce the existence of a sequence $U', U'', \ldots$ of spaces and integers $N_1, N_2, \ldots$ such that $vN_v < N_{v+1}$ and the homogeneous components of $U^{(v)}$ are 0 except for degrees in the range $[N_v, vN_v]$. Show that if $\mathfrak{a}^{(v)}$ is the ideal of $F$ generated by $V^{(v)} = U' + \cdots + U^{(v)}$ and $r_n = \dim((V^{(v)})_n)$, then (i) $r_n$ satisfies (1) of Ex. (4) for $\varepsilon = d/n$ and (ii) for any $g \in F_+$ of degree $\leqslant v, g^{Nv} \in \mathfrak{a}^{(v)}$. Deduce that $F_+/\mathfrak{a}$, where $\mathfrak{a} = \cup \mathfrak{a}^{(v)}$, is a finitely generated nilalgebra, which is infinite-dimensional, and hence not nilpotent. (Hint: observe that $\binom{N+q-1}{q-1} < (N+q)^q$ and choose $N = N_v$ to satisfy $(N+q)^q \leqslant \varepsilon^2(d-2\varepsilon)^{N-2}$.)

(6) (Golod) Let $A = F/\mathfrak{a}$ be the algebra with nil but not nilpotent augmentation ideal $A_+ = F_+/\mathfrak{a}$ constructed in Ex. (5), with $k = \mathbf{F}_p$ as ground field, where $p$ is a given prime number. Show that every element of the form $1 + c$, $c \in A_+$ has multiplicative order a power of $p$. Deduce that the group $G$ generated by the $1 + x_i \pmod{\mathfrak{a}}$ is an infinite $p$-group. (Hint: if $G$ were finite, the degrees of its members would be bounded, by $M$ say. Now pick a monomial $u_1 \cdots u_{m+1} \neq 0$ in $A$ and consider the element $\prod(1 + u_i)$ of $G$.)

## 2.4 The exterior algebra on a module

In this section we shall describe an algebra which is useful in the study of subspaces of a vector space. But the definition can be formulated more generally. Let $U$ be a $K$-module, where $K$ is any commutative ring, and let $\mathbf{T}(U)$ be the tensor algebra on $U$, as graded algebra. In $\mathbf{T}(U)$ consider the ideal $\mathfrak{a}$ generated by all the elements $u^2, u \in U$. Since the elements $u^2$ are homogeneous in $\mathbf{T}(U)$, the ideal $\mathfrak{a}$ is graded and therefore the quotient $\mathbf{T}(U)/\mathfrak{a}$ is again graded. It is called the *exterior algebra* on $U$ and is written $\Lambda(U)$. The component of degree $r$ of $\Lambda(U)$ is denoted by $\Lambda^r(U)$; its elements are called *r-vectors* or *multivectors*. The multiplication in $\Lambda(U)$ is denoted by $x \wedge y$. Thus $\Lambda(U)$ is the $K$-algebra

generated by the module $U$ with the defining relations

$$u \wedge u = 0 \qquad \text{for all } u \in U. \tag{1}$$

For any $u, v \in U$ we have $uv + vu = (u+v)^2 - u^2 - v^2$ in $\mathbf{T}(U)$, hence (1) entails the relation $u \wedge v + v \wedge u = 0$, i.e.

$$u \wedge v = -v \wedge u \qquad \text{for all } u, v \in U. \tag{2}$$

By induction on the degree we find that $\Lambda(U)$ is anticommutative; in fact the definition shows that when $K$ is a field of characteristic not 2 then $\Lambda(U)$ is the free anticommutative algebra on the $K$-module $U$.

On every graded algebra we can define a linear mapping $\sigma$ which multiplies each homogeneous element of degree $r$ by $(-1)^r$. Since $(-1)^r \cdot (-1)^s = (-1)^{r+s}$, this is an algebra homomorphism; in fact it is an automorphism, since its square is the identity mapping. Any mapping whose square is 1 is called an *involution*, and $\sigma$ is called the *standard involution* on $\Lambda(U)$. By an *antiderivation* of $\Lambda(U)$ we understand a $(\sigma, 1)$-derivation of $\Lambda(U)$ into itself, i.e. a linear mapping $\delta$ such that

$$(a \wedge b)^\delta = a^\delta \wedge b + a^\sigma \wedge b^\delta \qquad \text{for all } a, b \in \Lambda(U). \tag{3}$$

Such a mapping is entirely determined by its effect on the generating set $U$ of $\Lambda(U)$, and a mapping $\delta: U \to \Lambda(U)$ defines an antiderivation of $\Lambda(U)$ provided that

$$u^\delta \wedge u - u \wedge u^\delta = 0 \qquad \text{for all } u \in U. \tag{4}$$

For we can always extend $\delta$ to an antiderivation of $\mathbf{T}(U)$, again written $\delta$, and this will induce an antiderivation of $\Lambda(U)$ provided that $\delta$ maps $\mathfrak{a}$, the kernel of the mapping $\mathbf{T}(U) \to \Lambda(U)$, into itself. Now by (4), $(u^2)^\delta = u^\delta u + u^\sigma u^\delta = u^\delta u - uu^\delta \in \mathfrak{a}$, and more generally, for any $a, b \in \mathbf{T}(U)$, $(au^2 b)^\delta = a^\delta u^2 b + a^\sigma (u^2)^\delta b + a^\sigma (u^\sigma)^2 b^\delta$, and this again lies in $\mathfrak{a}$. Thus $\delta$ maps $\mathfrak{a}$ into itself and so induces an antiderivation on $\Lambda(U)$.

To give an example, if $u^\delta \in U$, then (4) reduces to $2u \wedge u^\delta = 0$; on the other hand, if $u^\delta \in K$, i.e. if $\delta \in U^* = \mathrm{Hom}_K(U, K)$, then the condition (4) always holds. Thus any element of the dual module $U^*$ defines an antiderivation of degree $-1$ on $\Lambda(U)$; this is sometimes called *interior* multiplication.

The structure of $\Lambda(U)$ may be described as follows.

THEOREM 4.1 *If the $K$-module $U$ is generated by $e_1, \ldots, e_n$, then $\Lambda(U)$ is generated by the elements*

$$e_{i_1} \wedge \cdots \wedge e_{i_r} \qquad \text{where } i_1 < \cdots < i_r, r = 0, 1, \ldots. \tag{5}$$

*In particular, whenever $U$ is finitely generated, by $n$ elements, then $\Lambda(U)$ can be generated by $2^n$ elements. Moreover, when $U$ is a free $K$-module on $e_1, \ldots, e_n$ as basis, then the elements* (5) *form a basis of $\Lambda(U)$.*

*Proof.* From the definition of $\Lambda(U)$ as quotient of the tensor algebra on $U$ it follows that $\Lambda(U)$ is generated by the products (5), where the suffixes range from 1 to $n$ without restriction. By (2) we can permute adjacent factors by changing the sign. Thus we can bring the suffixes into ascending order and, if two are equal, the corresponding product vanishes by (1). This shows that $\Lambda(U)$ is spanned by the products (5) with strictly ascending suffix sets, as claimed.

Now assume that $U$ is a free $K$-module with $e_1, \ldots, e_n$ as basis. We must show that the products (5) are linearly independent. Since $\Lambda(U)$ is graded, every relation is a sum of homogeneous relations; let

$$\sum a_{i_1 \cdots i_r} e_{i_1} \wedge \cdots \wedge e_{i_r} = 0 \tag{6}$$

be a non-trivial relation of least degree. Take a dual basis $\lambda_1, \ldots, \lambda_n$ for the $e$'s, i.e. a family of linear functionals on $U$ such that $\langle \lambda_i, e_j \rangle = \delta_{ij}$, and denote the corresponding antiderivations on $\Lambda(U)$ again by $\lambda_1, \ldots, \lambda_n$. Suppose that (6) contains a non-zero term in which $i_1 = 1$. Applying $\lambda_1$ to (6), we obtain

$$\sum a_{1 i_2 \cdots i_r} e_{i_2} \wedge \cdots \wedge e_{i_r} = 0,$$

and this is a non-trivial relation in $\Lambda(U)$ of lower degree. This is a contradiction, and it shows that no non-trivial relation (6) exists with $i_1 = 1$. Similarly for $i_1 = 2, \ldots, n$, hence there are no non-trivial relations (6) and so the elements (5) form a basis, as claimed. ■

If $U$ is a free $K$-module of infinite rank, with a totally ordered basis $\{e_i\}$, then it follows in exactly the same way that the set of elements (5) where the $i_\rho$ now range over the whole index set, forms a basis of $\Lambda(U)$. All this applies in particular when $K$ is a field, so that every $K$-module is free. In that case we have the following criterion for linear dependence:

COROLLARY 4.2 *Let $U$ be a vector space over a field $k$. Then the elements $u_1, \ldots, u_r \in U$ are linearly dependent if and only if*

$$u_1 \wedge \cdots \wedge u_r = 0 \qquad \text{in } \Lambda(U).$$

*Proof.* If the $u$'s are linearly dependent, one of them can be expressed in terms of the rest, say $u_1 = \sum_2^r \alpha_i u_i$, hence

$$u_1 \wedge u_2 \wedge \cdots \wedge u_r = \sum_2^r \alpha_i u_i \wedge u_2 \wedge \cdots \wedge u_r,$$

and the right-hand side vanishes because each term has a repeated factor. Conversely, if $u_1, \ldots, u_r$ are linearly independent, they can be completed to a basis of $U$ and then $u_1 \wedge \cdots \wedge u_r \neq 0$ by Th. 4.1 (and the remark following it), because it forms part of a basis of $\Lambda(U)$. ■

The following universal property of $\Lambda(U)$ is clear from the definition.

PROPOSITION 4.3 *Let $U$ be a $K$-module and $A$ a $K$-algebra. Then any linear mapping $\alpha: U \to A$ such that $(u^\alpha)^2 = 0$ for all $u \in U$ can be extended uniquely to a homomorphism of $\Lambda(U)$ into $A$.* ■

Taking in particular $A = \Lambda(V)$, for any $K$-module $V$, we see that any linear mapping $f: U \to V$ induces a homomorphism $\Lambda(f): \Lambda(U) \to \Lambda(V)$. It is easily verified that $\Lambda(fg) = \Lambda(f)\Lambda(g)$ and $\Lambda(1) = 1$, so that $\Lambda$ is a functor from $K$-modules to graded $K$-algebras.

We next show how determinants can be quite naturally defined in terms of $\Lambda(U)$. Let $U$ be a free $K$-module of rank $n$ and $f: U \to U$ a linear mapping of $U$ into itself. In terms of a basis $e_1, \ldots, e_n$ of $U$, $f$ is given by an $n \times n$ matrix $\alpha = (\alpha_{ij})$, where

$$e_i f = \sum \alpha_{ij} e_j.$$

$f$ induces an endomorphism $\Lambda(f)$ of $\Lambda(U)$, in particular, $\Lambda^n(f)$ maps $\Lambda^n(U)$ into itself, and since $\Lambda^n(U)$ is free of rank 1, $\Lambda^n(f)$ is determined by a scalar factor. Let us find this scalar; we have

$$(e_i \wedge \cdots \wedge e_n)f = \sum \alpha_{1i_1}\alpha_{2i_2}\cdots\alpha_{ni_n} e_{i_1} \wedge \cdots \wedge e_{i_n}.$$

The sum on the right is over all suffix sequences $(i_1, \ldots, i_n)$, but $e_{i_1} \wedge \cdots \wedge e_{i_n}$ is zero unless $(i_1, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$, and then it is $\pm e_1 \wedge \cdots \wedge e_n$ with $+$ or $-$ according to whether this permutation is even or odd. Now recall the definition of the determinant of a matrix $\alpha$ (A.1, **7.1**):

$$\det \alpha = \sum \varepsilon(i_1, \ldots, i_n)\alpha_{1i_1}\cdots\alpha_{ni_n}.$$

A comparison shows that

$$(e_1 \wedge \cdots \wedge e_n)f = (\det \alpha)e_1 \wedge \cdots \wedge e_n. \tag{7}$$

In fact we can take (7) as the definition of $\det \alpha$; this associates the determinant with an endomorphism of $U$ rather than a matrix. Of course we have already seen in A.1, **4.6** that if $f$ is given by a matrix $A$ in one coordinate system and by $B$ in another, then $B = P^{-1}AP$, for some invertible matrix $P$, so that $\det A = \det B$.

From this definition it is particularly easy to prove the multiplication theorem for determinants. Writing $e_T = e_1 \wedge \cdots \wedge e_n$, we have for any linear mappings $f, g$ with matrices $\alpha, \beta$ respectively, $e_T(fg) = (e_T f)g$; hence by applying (7) to both sides,

$$\det(\alpha\beta)e_T = (\det \alpha)(\det \beta)e_T,$$

and so $\det(\alpha\beta) = (\det \alpha)(\det \beta)$.

More generally, any endomorphism $f$ of $U$ induces an endomorphism $\Lambda^r(f)$ of $\Lambda^r(U)$, for $1 \leqslant r \leqslant n$; the corresponding matrix $\alpha^{(r)}$ has $\binom{n}{r}$ rows and columns,

the entries being $r$th-order minors of the matrix $\alpha$. This matrix $\alpha^{(r)}$ is called the $r$th *compound matrix* of $\alpha$ and as before we see that

$$(\alpha\beta)^{(r)} = \alpha^{(r)}\beta^{(r)}. \tag{8}$$

This is an identity of degree $r$ in the entries of $\alpha$ and $\beta$, called the *Binet–Cauchy identity*; the case $r = n$ is just the multiplication theorem for determinants.

Let $U$ again be the free $K$-module with basis $e_1, \ldots, e_n$ and let us find the multiplication table for $\Lambda(U)$ in terms of the standard basis (5). For any subset $I$ of $\{1, 2, \ldots, n\}$ we denote by $e_I$ the product of the $e_i (i \in I)$ in ascending order; thus if $I = \{1, 2, \ldots, r\}$ then $e_I = e_1 \wedge \cdots \wedge e_r$. Given any subsets $H, K$ of $\{1, \ldots, n\}$, if $H \cap K \neq \varnothing$, then it is clear that $e_H \wedge e_K = 0$. For $H \cap K = \varnothing$ we have

$$e_H \wedge e_K = \varepsilon_{HK} e_{H \cup K}, \tag{9}$$

where $\varepsilon_{HK} = \pm 1$. Here the number of inversions necessary to arrange the suffixes of $e_H \wedge e_K$ in ascending order is just the number of pairs $i \in H$, $j \in K$ such that $i > j$. If this number is $v$, then $\varepsilon_{HK} = (-1)^v$.

If $H$ is any subset of $T = \{1, \ldots, n\}$ and $H'$ is the complementary subset, then $e_H \wedge e_{H'} = \varepsilon_{HH'} e_T$. Suppose that $|H| = r$; on applying a linear mapping $f$ with matrix $\alpha$, we find

$$e_H \cdot \Lambda^r(f) = \sum \alpha_{HI} e_I,$$

where $I$ runs over all $r$-element subsets of $T$ and $\alpha_{HI}$ is the minor corresponding to the rows indexed by $H$ and the columns indexed by $I$. Thus by applying $f$ to both sides of the equation

$$\varepsilon_{HH'} e_T = e_H \wedge e_{H'}$$

we obtain

$$\begin{aligned}(\det \alpha)\varepsilon_{HH'} e_T &= e_H \Lambda^r(f) \wedge e_{H'} \Lambda^{n-r}(f) \\ &= \sum \alpha_{HI} \alpha_{H'J} e_I \wedge e_J,\end{aligned}$$

where $I$ runs over all $r$-element sets and $J$ over all $(n-r)$-element sets in $T$. It is clear that

$$e_I \wedge e_J = \begin{cases} \varepsilon_{II'} e_T & \text{if } J = I', \\ 0 & \text{otherwise,} \end{cases}$$

where $I'$ is the complement of $I$. Hence we obtain the formula

$$\det \alpha = \sum_I \varepsilon_{II'} \varepsilon_{HH'} \alpha_{HI} \alpha_{H'I'} \tag{10}$$

where the summation is over all $r$-element sets $I$, and $I'$ is again the complement of $I$. This is called the *Laplace expansion* of $\det \alpha$ in terms of the $r$ rows indexed

by $H$. In terms of compound matrices this can be written briefly as

$$\alpha^{(r)}\alpha_{(n-r)} = (\det \alpha)\cdot I, \tag{11}$$

for an appropriately defined matrix of cofactors $\alpha_{(n-r)}$.

In a vector space $U$ over a field, any non-zero vector can be transformed into any other non-zero vector, i.e. the automorphisms of $U$ act transitively on the non-zero vectors. By contrast the automorphisms of $\Lambda(U)$ do not always act transitively on the set of all non-zero $r$-vectors for $r > 1$, e.g. the 2-vectors of the form $u \wedge v$ form a proper subset of $\Lambda^2(U)$ (at least when $\dim U > 3$), which is mapped into itself by all automorphisms. An element of $\Lambda(U)$ is said to be *decomposable* if it can be written as a product of vectors: $u_1 \wedge \cdots \wedge u_r$. These decomposable $r$-vectors were used to describe $r$-dimensional subspaces by H. G. Grassmann, who introduced exterior algebras in 1844; the exterior algebra on the dual space $U^*$ is also called the *Grassmann algebra* of $U$.

As an example let us derive the condition for a 2-vector to be decomposable.

PROPOSITION 4.4 *Let $U$ be a vector space over a field of characteristic not 2, with basis $e_1, \ldots, e_n$. Then any 2-vector in $\Lambda(U)$ can be written in the form*

$$w = \sum c_{ij} e_i \wedge e_j, \qquad \text{where } c_{ii} = 0, c_{ij} = -c_{ji}, \tag{12}$$

*and $w$ is decomposable if and only if, for any four suffixes $i,j,k,l$,*

$$c_{ij}c_{kl} + c_{ik}c_{lj} + c_{il}c_{jk} = 0. \tag{13}$$

*Proof.* By definition any 2-vector has the form $\sum c_{ij} e_i \wedge e_j$, but there are no terms in $e_i \wedge e_i$, so we may take $c_{ii} = 0$, and for fixed $i,j$ we have $\lambda e_i \wedge e_j + \mu e_j \wedge e_i = (\lambda - \mu)e_i \wedge e_j = ce_i \wedge e_j - ce_j \wedge e_i$, where $c = (\lambda - \mu)/2$. Thus every 2-vector can be expressed as in (12). Now let $w = p \wedge q$, where $p = \sum a_i e_i, q = \sum b_i e_i$; then $w = \sum (a_i b_j - a_j b_i) e_i \wedge e_j$. Consider the determinant

$$\begin{vmatrix} a_i & a_j & a_k & a_l \\ b_i & b_j & b_k & b_l \\ a_i & a_j & a_k & a_l \\ b_i & b_j & b_k & b_l \end{vmatrix}.$$

If we make a Laplace expansion by the first two rows and observe that the determinant vanishes, we obtain (13), bearing in mind that $a_i b_j - a_j b_i = c_{ij}$. Conversely, given $n^2$ elements $c_{ij}$ of $k$, where $c_{ii} = 0$ and $c_{ij} = -c_{ji}$, assume that (13) holds for all choices of suffixes. If all the $c_{ij}$ vanish, the 2-vector (12) is zero and hence decomposable. Otherwise let $c_{12} \neq 0$ say, and put $a_j = (c_{1j})$, $b_i = (c_{i2})$; then $p = \sum a_i e_i$ and $q = \sum b_i e_i$ are such that $p \wedge q$ has coordinates

$$\begin{aligned} c'_{ij} = a_i b_j - a_j b_i &= c_{1i}c_{j2} - c_{1j}c_{i2} \\ &= -c_{12}c_{ij}, \qquad \text{by (13).} \end{aligned}$$

Hence $w = -c_{12}^{-1} p \wedge q$ is decomposable. ■

We observe that in three dimensions no relations are needed: every 2-vector is then decomposable. In a four-dimensional space there is just one relation

$$c_{01}c_{23} + c_{02}c_{31} + c_{03}c_{12} = 0. \tag{14}$$

As the proof of Prop. 4.4 shows, in $n$ dimensions we need only impose the condition (13) for fixed $i,j$ such that $c_{ij} \neq 0$, and all pairs $k,l$; thus $\binom{n-2}{2}$ relations are needed and the rest are then a consequence.

Proposition 4.4 has an interesting application to lines in projective 3-space. The straight lines in three-dimensional projective space are two-dimensional subspaces of a four-dimensional space, and so can be described by decomposable 2-vectors. From Prop. 4.4 we see that each line is described by six homogeneous coordinates $c_{ij}(i,j = 0,1,2,3)$ satisfying (14), the *Plücker* or *line coordinates*. This equation represents a quadric $\Omega$ in five dimensions (since there are six coordinates), the *Klein quadric*; many properties of lines in space can be obtained very simply by regarding the lines as points of this quadric. To give an example, this interpretation shows that the lines in 3-space form a four-dimensional variety. A subvariety of codimension 1 is called a *line complex*; such a subvariety can be obtained by imposing certain algebraic relations on the line coordinates, and a theorem of Klein asserts that any line complex in projective 3-space can be defined by a single equation besides (14). For a proof we note that the coordinate ring of a line complex is obtained from the coordinate ring $A$ of $\Omega$ by dividing by a minimal prime ideal (minimal because the subvariety is of codimension 1, i.e. maximal). Now it can be shown that $A$ is a UFD (Ex. (8)), hence every minimal ideal is principal (A.2, Th. 9.2.6) and this means that the quotient of $A$ by a minimal prime ideal is obtained by imposing a single equation, as claimed (this simple proof is due to Nagata (1957), cf. also Cohn (1973)). For a discussion of the properties of lines using $\Omega$ see Semple and Roth [1949].

The decomposable $r$-vectors correspond to the $r$-dimensional subspaces of $U$; the coordinates can generally be taken to be antisymmetric in all suffixes (at least in characteristic 0). They are again called the *Plücker coordinates* of the subspace, and they can again be characterized by relations like (13). Their derivation is straightforward and we shall not stop to give it here (cf. Ex. (11)), but we note the following relation between the multivectors of different subspaces.

PROPOSITION 4.5 *Let $U$ be an $n$-dimensional $k$-space. To each $r$-dimensional subspace $V$ of $U$ $(1 \leqslant r \leqslant n)$ there corresponds a decomposable $r$-vector $v$, determined up to a non-zero element of $k$, such that*

$$x \in V \quad \text{if and only if} \quad x \wedge v = 0.$$

*Given subspaces $V, W$ with vectors $v, w$, we have $v \wedge w \neq 0$ if and only if $V \cap W = 0$, and when this is so, then $v \wedge w$ is the vector corresponding to $V + W$.*

*Proof.* Let $e_1,\dots,e_r$ be a basis for $V$; the corresponding $r$-vector is $v = e_1 \wedge \cdots \wedge e_r$. Any other basis of $V$ is related to the $e$'s by a linear transformation $\alpha$ and the change of basis changes $v$ by a factor $\det \alpha$, by (7). Now $x \in V$ iff $x$ is linearly dependent on $e_1,\dots,e_r$ and this is so iff $x \wedge v = 0$, by Cor. 4.2. Given subspaces $V, W$ with multivectors $v, w$, if $V \cap W \neq 0$, then for a suitable choice of basis $v$ and $w$ have a common factor and so $v \wedge w = 0$. Otherwise we obtain a basis for $V + W$ by taking a basis for $V$ and one for $W$, so $v \wedge w$ is a multivector for $V + W$. ■

## Exercises

(1) Show that the elements of even degree of $\Lambda(U)$ form a subalgebra.

(2) Show that any exterior algebra satisfies the identity $[[x, y], z] = 0$, where $[x, y] = xy - yx$.

(3) Find the first power of $t$ in $\left(1 - dt + \binom{d}{2}t^2\right)^{-1}$ which has a negative coefficient (there must be one, by Th. 3.5 and 4.1).

(4) In a $k$-module $U$ consider an equation $a = \sum \alpha_i u_i$, where $a, u_i \in U$ and $\alpha_i \in K$. Show that

$$u_1 \wedge \cdots \wedge u_{i-1} \wedge a \wedge u_{i+1} \wedge \cdots \wedge u_n = \alpha_i(u_1 \wedge \cdots \wedge u_n).$$

Taking $K$ to be a field and $u_1,\dots,u_n$ linearly independent, deduce Cramer's rule for the solution of the equation.

(5) Let $U$ be an $n$-dimensional vector space. Show that the decomposable multivector $w_i$ corresponding to a subspace $W_i (i = 1, 2)$ satisfies $w_2 = z \wedge w_1$ for some multivector $z$ iff $W_2 \supseteq W_1$.

(6) Let $U$ be a vector space with basis $e_1,\dots,e_n$. Show that if $u_1,\dots,u_n \in U$ satisfy $\sum u_i \wedge e_i = 0$, then $u_i = \sum a_{ij} e_j$, where $a_{ij} = a_{ji}$.

(7) Let $A$ be the coordinate ring of the quadric $x_0 x_3 = x_1 x_2$. Show that $A$ is an integral domain, but not a UFD.

(8) (Nagata) Let $A$ be the $\mathbf{C}$-algebra generated by $x_0,\dots,x_n (n \geqslant 4)$ with a single defining relation $\Phi = 0$, where $\Phi$ is a (homogeneous) non-degenerate quadratic form in the $x$'s. Show that by a change of coordinates $\Phi$ can be written $x_0 x_1 + \varphi$, where $\varphi$ is a non-degenerate quadratic form in $n - 1 \geqslant 3$ variables. Show that if $S$ is the multiplicative set generated by $x_0$, then the localization $A_S$ satisfies $A_S \cong \mathbf{C}[x_1,\dots,x_n]_S$. Use Nagata's theorem (Th. 9.3.7 of A.2, which for a Noetherian domain $A$ states that (i) if $A$ is a UFD, then so is $A_S$ and (ii) if $A_S$ is a UFD and $S$ consists of products of primes, then $A$ is a UFD) to deduce that $A$ is a UFD.

(9) Let $V$ be a vector space with basis $e_1,\dots,e_n$ over a field of characteristic 0. Show

that any $r$-vector can be written as $\sum a_{i_1 \cdots i_r} e_{i_1} \wedge \cdots \wedge e_{i_r}$, where $a_{i_1 \cdots i_r}$ is antisymmetric, and it is decomposable iff

$$a_{i_1 \cdots i_r} a_{j_1 \cdots j_r} - \sum a_{i_1 \cdots i_{\lambda-1} j_1 i_{\lambda+1} \cdots i_r} a_{i_\lambda j_2 \cdots j_r} = 0.$$

(10) Prove (7) by verifying that $\Lambda^n(f)$ satisfies the axioms for a determinant (cf. A.1, **7.1**: $\det A$ is the alternating multilinear function of the columns of $A$ which is 1 for $A = I$).

(11) Show that the determinant of the $r$th compound of an $n \times n$ matrix $\alpha$ is a power of $\det \alpha$: $\det \alpha^{(r)} = (\det \alpha)^m$. By comparing degrees show that $m = \binom{n-1}{r-1}$ (Sylvester–Franke theorem. Hint: use (11).)

(12) Show that the left-hand side of (14) is the Pfaffian of the skew-symmetric matrix $(c_{ij})$, up to sign.

## Further exercises on Chapter 2

(1) Let A be a graded $K$-algebra. Find the conditions for the mapping $a_n \mapsto \lambda_n a_n$ (where $\deg a_n = n$ and $\lambda_n \in K$) to be (i) an endomorphism and (ii) a derivation.

(2) Show that the centre of $K\langle X \rangle$ is $K$ whenever $|X| > 1$.

(3) Show that any homomorphism $K\langle X \rangle \to K$, for any ring $K$, maps nilpotent elements to nilpotent elements. If $K$ has no nilpotent elements apart from 0, the same is true for $K\langle X \rangle$. (Hint: take the free monoid on $X$ to be ordered lexicographically.)

(4) By a *filtered* $K$-algebra one understands a $K$-algebra $A$ with a sequence of submodules

$$\cdots \supseteq A_n \supseteq A_{n+1} \supseteq \cdots, \qquad A_i A_j \subseteq A_{i+j}, \qquad \cap A_n = 0, \qquad \cup A_n = A.$$

Given such an algebra, put $G_n = A_n / A_{n+1}$ and define a multiplication $G_i \times G_j \to G_{i+j}$ by taking for $\alpha \in G_i, \beta \in G_j$ any representative $a$ of $\alpha$, $b$ of $\beta$ and defining $\alpha\beta$ as the residue class of $ab$ (mod $A_{i+j+1}$). Verify that this definition of $\alpha\beta$ depends only on $\alpha, \beta$ and not on $a, b$ and that with this definition $(G_n)$ becomes a graded $K$-algebra (called the *graded ring associated with* $A$ and denoted by $\mathrm{gr}(A)$). Verify that for any ideal $\mathfrak{a}$ of a ring $R$, if $\cap \mathfrak{a}^n = 0$, then the powers of $\mathfrak{a}$ define a filtration and so give rise to a graded ring.

(5) Let $A$ be a filtered $K$-algebra and $\mathrm{gr}(A)$ its associated graded algebra. Show that if $\mathrm{gr}(A)$ is an integral domain, then so is $A$. Does the converse hold?

(6) Show that for any vector space $V$ over a field, each element of $\Lambda(V)$ is either invertible or nilpotent (a ring with this property is said to be *completely primary*.).

(7) Let $V$ be a vector space over a field. Show that if a $p$-vector $w$ in $\Lambda(V)$ is divisible by $u_1, \ldots, u_r$, where the $u_i$ are linearly independent, then $w$ is divisible by $u_1 \wedge \cdots \wedge u_r$.

Deduce that any $p$-vector $w$ can be written as $w = u_1 \wedge \cdots \wedge u_r \wedge v$, where $v$ has no linear factor.

(8) Let $F = k\langle x_1, \ldots, x_d \rangle$; given an equation $\sum x_i u_i = 0$, where $u_i \in F$, show that $u_i = 0$. Deduce that in any element $f \in F$ the right-hand cofactor of $x_i$ is uniquely determined (this is called the *transduction* with respect to $x_i$).

(9) Show that in $K\langle x, y \rangle$ the elements $xy^n (n = 1, 2, \ldots)$ generate a subalgebra which is free on these generators.

(10) Let $X$ be a finite set and denote by $K \ll X \gg$ the set of all sums 2.2 (1), allowing infinitely many non-zero coefficients. Show that the multiplication law 2.2 (2) defines a $K$-algebra structure on $K \ll X \gg$, containing $k\langle X \rangle$ as subalgebra. Show that a mapping $X \to K \ll X \gg$ extends to an endomorphism of $K \ll X \gg$ provided that the image of any $x \in X$ has zero constant term. (This algebra $K \ll X \gg$ is the free power series ring in $X$ over $K$, cf. A.2, **11.5**.)

(11) Let $K$ be an integral domain. Show that $f \in K \ll X \gg$ is invertible iff its constant term is invertible in $K$, whereas $f \in K\langle X \rangle$ is invertible iff $f$ reduces to its constant term and the latter is invertible in $K$.

(12) Let $V$ be a vector space over a field $k$. Show that there is a pairing $\langle\ ,\ \rangle : \Lambda^r(V^*) \times \Lambda^r(V) \to k$ such that

$$\langle f_1 \wedge \cdots \wedge f_r, x_1 \wedge \cdots \wedge x_r \rangle = \det(\langle f_i, x_j \rangle).$$

Hence find an injective linear mapping $\Lambda(V^*) \to (\Lambda(V))^*$ which is an isomorphism if $V$ is finite-dimensional.

(13) Using Ex. (12) show that if $V$ is $n$-dimensional, then every $(n-1)$-vector in $\Lambda(V)$ is decomposable. Show that (in characteristic not two) a 2-vector $p$ is decomposable iff $p \wedge p = 0$.

(14) Let $V$ be a left $k$-space and $V^*$ its dual. If $f \in \mathrm{End}(V)$ has the matrix $\alpha$ relative to a certain basis of $V$, then $f^*$ relative to the dual basis of $V^*$ again has the matrix $\alpha$ if $V^*$ is regarded as right $k$-space (and $f^*$ acts on the left), but it has the transposed matrix $\alpha^T$ if $V^*$ is regarded as left $k$-space (with $f^*$ acting on the right). Deduce that $\det \alpha = \det \alpha^T$.

(15) Let $V$ be a vector space and $a \in V$. Show that the mapping $x \mapsto a \wedge x$ of $V$ into $\Lambda(V)$ extends to an antiderivation of $\Lambda(V)$; obtain its expression in terms of multiplication by $a$.

(16) Show that the $r$th compound matrix $\alpha^{(r)}$ satisfies

$$\alpha_{11}^{(n-1)} \alpha_{ii}^{(n-1)} - \alpha_{1i}^{(n-1)} \alpha_{i1}^{(n-1)} = \alpha_{1i,1i}^{(n-2)} \cdot \det \alpha,$$

where suffixes indicate the rows and columns omitted.

(17) Show that a square matrix $A$ over a commutative ring $K$ is a zero-divisor iff $\det A$ is a zero-divisor in $K$.

(18) For any square matrix $A$ (over a commutative ring $K$) define $e_i(A)$ to be the $i$th elementary symmetric function of its eigenvalues. Show that $e_i(A) = \mathrm{tr}(A^{(i)})$, where $A^{(i)}$ is the $i$th compound, and that the characteristic polynomial of $A$ is $\sum(-1)^r e_r(A)\lambda^{n-r}$.

Deduce that if $K = \mathbf{C}$ and $A^H$ is the Hermitian conjugate, then

$$\det(A^H A + \lambda I) = \sum_0^n \mathrm{tr}(A^{(i)H} A^{(i)})\lambda^{n-i}.$$

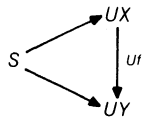
# 3
# Homological algebra

The present chapter serves as a concise introduction to homological algebra. Only the basic notions of category theory (treated in Volumes 1 and 2) are assumed. The definition of abelian categories and of functors between them is followed by an abstract description of module categories. A study of resolutions leads to the notion of homological dimension; derived functors are then defined and exemplified by the instances that are basic for rings: Ext and Tor.

## 3.1 Additive and abelian categories

We have met general categories in A.1 and 2, but most of the instances have been categories of modules or at least categories with similar properties. For a general study we shall therefore postulate the requisite properties; this will lead to additive categories, and more particularly, abelian categories. Later in **3.2** we shall see what further assumptions are needed to reach module categories.

We recall that an object $I$ in a category $\mathscr{A}$ is said to be *initial* if for each $\mathscr{A}$-object $X$ there is a unique morphism $I \to X$; dually, if there is a unique morphism $X \to I$ for each $X$, $I$ is called a *final* object. As we saw in A.2, **1.3**, an initial (or final) object, when it exists, is unique up to a unique isomorphism. Initial objects arise in the solution of universal problems. Thus let $\mathscr{A}$ be a concrete category, i.e. a category with a forgetful functor $U$ to Ens (the category of sets), which is faithful, and denote by $UX$ the underlying set of an $\mathscr{A}$-object $X$. We fix a set $S$ and consider the category $(S, U)$ whose objects are mappings $S \to UX$ $(X \in \mathrm{Ob}\,\mathscr{A})$ and whose morphisms are commutative triangles arising from an $\mathscr{A}$-map $f: X \to Y$ by applying $U$. This is the *comma category* based on $S$ and $U$. An initial object in this category is said to have the *universal property* for the set $S$. For example, in the category of groups, the free group on $S$ has the universal property for $S$.




Let $\mathscr{A}$ be a category and $(X_i)$ any family of $\mathscr{A}$-objects. Given $\mathscr{A}$-objects $P$,

$Y$, each family of maps $\pi_i: P \to X_i$ gives rise to a natural mapping

$$\mathscr{A}(Y, P) \to \prod \mathscr{A}(Y, X_i), \tag{1}$$

where $g \mapsto g\pi_i$. When (1) is bijective for each $Y$ we call $P$ a *product* of the $X_i$ with *natural projections* $\pi_i$. The product $P$ with its maps $\pi_i$ can also be described as the solution of a universal problem, for it is the final object in the category whose objects $(A, f_i)$ are families of maps $f_i: A \to X_i$ and whose morphisms $\varphi: (A, f_i) \to (B, g_i)$ are families of commutative triangles, thus $\varphi: A \to B$ satisfies $f_i = \varphi g_i$. It follows that the product, when it exists, is unique up to isomorphism. We shall denote it by $\prod X_i$; thus we have an isomorphism, natural in $Y$:

$$\mathscr{A}(Y, \prod X_i) \cong \prod \mathscr{A}(Y, X_i). \tag{2}$$

Here $\prod$ on the left denotes the product just defined, and on the right the usual Cartesian product of sets.

*Examples*

1. In Ens, $\prod$ reduces to the Cartesian product. Likewise in *Ab*, or more generally, in $\mathscr{M}_R$, $\prod$ is the direct product introduced in A.1, **2.3**, **9.6**.
2. In the category of all abelian torsion groups, $\prod X_i$ is the torsion subgroup of the ordinary direct product of the $X_i$.
3. In the category of all finite abelian groups the product does not exist; this is easily seen by taking an infinite family of non-trivial groups.
4. The product of the empty family (in any category where it exists) is a final object in the category. For here we have on the right of (1) the empty product, which by convention is a one-element set.

There is a dual construction, the coproduct: given any family $(X_i)$ of $\mathscr{A}$-objects, their *coproduct* is an $\mathscr{A}$-object $S$ with maps $\mu_i: X_i \to S$, called the *natural injections*, such that $(S, \mu_i)$ is a product of the $X_i$ in the dual category $\mathscr{A}^{op}$. For an explicit definition we need only reverse the arrows in the definition of the product. Thus $(S, \mu_i)$ is a coproduct if for any $\mathscr{A}$-object $Y$ the natural mapping

$$\mathscr{A}(S, Y) \to \prod \mathscr{A}(X_i, Y)$$

given by $f \mapsto \mu_i f$ is a bijection. As before, the coproduct, when it exists, is unique up to isomorphism; denoting it by $\coprod X_i$, we have the isomorphism

$$\mathscr{A}(\coprod X_i, Y) \cong \prod \mathscr{A}(X_i, Y).$$

For example, in Ens the coproduct is the disjoint union, while in Ab or $\mathscr{M}_R$ it is the direct sum (A.1, **10.3**). In particular, for a finite family of modules the product and coproduct are the same, except for the associated mappings (which

go in opposite directions). The connexion between these two concepts becomes clearer in additive categories.

DEFINITION A category $\mathscr{A}$ is said to be *additive* if

**Ad.**1 *each* $\mathscr{A}(X, Y)$ *is an abelian group with respect to an operation written* $+$,

**Ad.**2 *the composition* $\alpha, \beta \mapsto \alpha\beta$ *is biadditive, i.e.*

$$(\alpha_1 + \alpha_2)\beta = \alpha_1\beta + \alpha_2\beta, \qquad \alpha(\beta_1 + \beta_2) = \alpha\beta_1 + \alpha\beta_2, \tag{3}$$

**Ad.**3 *each finite family of objects has a product and a coproduct.*

By applying **Ad.**3 to the empty family of objects we see that each additive category has an initial and a final object. In **Ad.**3 it is enough to assume the existence for pairs of objects and for the empty family; the full strength can then be recovered by an easy induction argument. We also observe that the axioms **Ad.**1–3 are self-dual. Further we remark that in **Ad.**3 it is enough to postulate products (or only coproducts); the existence of the other sort results from the remarks following Th. 1.1 below.

For example, the category $\mathscr{M}_R$ of $R$-modules is additive, since $\mathrm{Hom}_R(M, N)$ has an abelian group structure for which (3) holds. On the other hand, Ens, **Rg** and Gp are not additive, for there is no way of defining an abelian group structure on the hom sets to satisfy (3).

It is clear that a category with a single object is just a monoid; if the category also satisfies **Ad.**1, 2, it is a ring. Similarly, in any additive category $\mathscr{A}$, the group $\mathscr{A}(X, X)$ is a ring for each $X \in \mathrm{Ob}\,\mathscr{A}$.

For finite families of objects in an additive category we can define a further type of product, which helps to clarify the connexion between products and coproducts. Let $\mathscr{A}$ be any additive category and $X_1, \ldots, X_n$ any $\mathscr{A}$-objects. The *biproduct* of the finite family $(X_i)$, written $\coprod X_i$, is an object $B$ with $2n$ maps $p_i: B \to X_i$, $q_i: X_i \to B$, such that

$$q_i p_j = \delta_{ij} 1_{X_i}, \qquad \sum p_i q_i = 1_B. \tag{4}$$

The three kinds of product are related by

THEOREM 1.1 *In an additive category* $\mathscr{A}$ *let* $(X_i)$ *be a finite family of objects. Given an* $\mathscr{A}$*-object* $B$ *and maps* $p_i: B \to X_i$, $q_i: X_i \to B$ *such that* $q_i p_j = \delta_{ij}$, *the following conditions are equivalent*:

(a) $(B, p_i)$ *is product of the* $X_i$,

(b) $(B, q_i)$ *is a coproduct of the* $X_i$,

(c) $(B, p_i, q_i)$ *is a biproduct of the* $X_i$.

*Proof*. (a)$\Rightarrow$(c). Let $(B, p_i)$ be a product and write $\varphi = \sum p_i q_i$; then $\varphi p_i: B \to X_i$ satisfies $\varphi p_i = p_i$, for $\varphi p_i = \sum_j p_j q_j p_i = p_i$. By uniqueness, $\varphi = 1_B$ and so $B$ is a biproduct.

(c)⇒(a). Given $f_i: A \to X_i$, we can define $f: A \to B$ by $f = \sum f_i q_i$. Then $f p_j = \sum_i f_i q_i p_j = f_j$, therefore $(B, p_i)$ is a product. We have thus shown that (a)⇔(c) and by duality, (b)⇔(c). ■

We observe that any finite product (or coproduct) in any category satisfying **Ad.**1, 2 can be completed to a biproduct in a unique way. Given a product $(B, p_i)$ of $X_1, \ldots, X_n$, fix $j$ and define $f_i: X_j \to X_i$ by $f_i = \delta_{ij}$. Then there exists $q_j: X_j \to B$ such that $q_j p_i = \delta_{ij}$. This holds for $j = 1, \ldots, n$, and now $(B, p_i, q_i)$ is a biproduct, by the theorem. Thus every product can be completed to a biproduct in a unique way, and the same holds for coproducts, by duality. Hence we obtain

COROLLARY 1.2 *Let $\mathscr{A}$ be a category satisfying* **Ad.**1, 2. *Then any finite family of $\mathscr{A}$-objects has a coproduct if and only if it has a product, and the two are isomorphic. In particular, the product and coproduct of any finite family are isomorphic in any additive category.* ■

Taking the empty family, we see that the initial and final object are isomorphic. An object that is both initial and final is called a *zero object*; thus every additive category has a zero object. By a *zero morphism* we understand a morphism which can be factored via a zero object. With this definition it is easily seen that in an additive category $\mathscr{A}$ the neutral element in each hom group is just the zero morphism.

We also note that on writing $p = (p_1, \ldots, p_n)$, $q = (q_1, \ldots, q_n)^T$, we can express (4) in the form

$$pq = 1_B, \qquad qp = \begin{pmatrix} 1_{X_i} & & 0 \\ & \ddots & \\ 0 & & 1_{X_i} \end{pmatrix}.$$

Our next task is a categorical description of kernels; we begin with monomorphisms and subobjects. In any category (not necessarily additive) a map $\alpha: X \to Y$ is said to be *monic* or a *monomorphism* if whenever $\lambda\alpha$, $\mu\alpha$ are defined,

$$\lambda\alpha = \mu\alpha \text{ implies } \lambda = \mu.$$

In an additive category the condition can of course be simplified to

$$\lambda\alpha = 0 \text{ implies } \lambda = 0.$$

In Ens the monomorphisms are just the injective mappings. More generally, in any concrete category injective morphisms are monic; the converse holds frequently but not always. By a *subobject* of an object $A$ we understand a pair $(X, \alpha)$ such that $\alpha: X \to A$ is monic. Two subobjects $(X, \alpha)$ and $(X', \alpha')$ are said to be *equivalent* if there is an isomorphism $\lambda: X \to X'$ such that $\alpha = \lambda\alpha'$. It is clear that this leads to the expected notion of subset in the category Ens, or a subgroup in Gp.

Dually a morphism $\alpha: A \to X$ is said to be *epic* or an *epimorphism* if it is left cancellable:

$$\alpha\lambda = \alpha\mu \text{ implies } \lambda = \mu.$$

In an additive category this can again be shortened to $\alpha\lambda = 0 \Rightarrow \lambda = 0$. If $\alpha: A \to X$ is epic, the pair $(X, \alpha)$ is called a *quotient object* of $A$, and equivalence is defined as before. A quotient object in Ens is just a quotient set; in Gp it is a quotient group (by a normal subgroup).

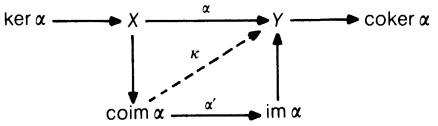
Let $\mathscr{A}$ be an additive category; given a map $\alpha: X \to Y$, we shall define the kernel of $\alpha$ as a certain subobject of $X$. Consider all maps $\lambda: A \to X$ such that $\lambda\alpha = 0$; for fixed $\alpha$ we obtain a category by taking these maps $\lambda$ as objects and as morphisms between $\lambda$ and $\lambda'$ maps $\varphi$ from the source of $\lambda$ to that of $\lambda'$ such that $\lambda = \varphi\lambda'$, with the obvious composition rule (obtained by composing maps in $\mathscr{A}$). A final object in this category, if one exists, is called a *kernel* of $\alpha$. Thus a kernel of $\alpha$ is a map $\lambda: A \to X$ such that $\lambda\alpha = 0$ and any other map $\lambda': A' \to X$ satisfying $\lambda'\alpha = 0$ can be factored uniquely by $\lambda$, i.e. we have $\lambda' = f\lambda$ for a unique map $f$. This can be expressed more briefly by saying that the kernel is the largest subobject 'killed' (i.e. mapped to 0) by $\alpha$. The kernel need not exist, but if it does, it is unique up to equivalence, and in fact is a subobject of $X$. For let $\lambda$ be the kernel and assume that $f\lambda = 0$; then by the uniqueness of the factorization, $f = 0$. The kernel of $\alpha$ will be denoted by $\ker\alpha$.

Dually the *cokernel* of $\alpha: X \to Y$ is an initial object in the category of all maps $\mu: Y \to C$ such that $\alpha\mu = 0$. This is a quotient object of $Y$, unique up to equivalence if it exists; it will be denoted by $\operatorname{coker}\alpha$.

Given any map $\alpha: X \to Y$, assume that $\ker\alpha$, $\operatorname{coker}\alpha$ exist. Then we can define two further objects, the *image* of $\alpha$, a subobject of $Y$, and the *coimage* of $\alpha$, a quotient object of $X$:

$$\operatorname{im}\alpha = \ker\operatorname{coker}\alpha, \qquad \operatorname{coim}\alpha = \operatorname{coker}\ker\alpha.$$

Again they need not exist, but if they do they are unique up to equivalence. Further we have the following diagram:




Here $\operatorname{coim}\alpha$ is the largest quotient of $X$ killing $\ker\alpha$, hence there is a map $\kappa: \operatorname{coim}\alpha \to Y$ (shown dotted) such that the upper triangle in the figure commutes. It follows that $(\operatorname{coim}\alpha)\kappa(\operatorname{coker}\alpha) = 0$; but $\operatorname{coim}\alpha$ is epic, so $\kappa(\operatorname{coker}\alpha) = 0$, and since $\operatorname{im}\alpha$ is the largest subobject of $Y$ killed by $\operatorname{coker}\alpha$, there is a unique map $\alpha': \operatorname{coim}\alpha \to \operatorname{im}\alpha$ to make the lower triangle commute. We now have a commutative square. If we had proceeded in dual fashion, starting from $\operatorname{im}\alpha$ and going via a map $X \to \operatorname{im}\alpha$, we would have obtained another map

$\alpha''$:coim $\alpha \to$ im $\alpha$ to make the square commute. Now $\alpha = (\text{coim}\,\alpha)\alpha'(\text{im}\,\alpha) = (\text{coim}\,\alpha)\alpha''(\text{im}\,\alpha)$, but coim $\alpha$ is epic and im $\alpha$ is monic, hence $\alpha' = \alpha''$, so the maps coincide and there is complete symmetry. In important cases $\alpha'$ is an isomorphisms, and this suggests the following

DEFINITION An *abelian category* is an additive category $\mathscr{A}$ such that

**Ab.1** *every map in* $\mathscr{A}$ *has a kernel and cokernel,*

**Ab.2** *the induced map* coin $\alpha \to$ im $\alpha$ *is an isomorphism.*

These axioms, like the others, are self-dual. An example of an abelian category is Ab, the category of abelian groups, or more generally, the category $\mathscr{M}_R$ of right $R$-modules, for any ring $R$ (cf. A.1, **10.2**). By contrast the category Gp is not abelian (it is not even additive), and there are additive categories that are not abelian, e.g. the category of topological abelian groups and continuous homomorphisms; here **Ab.2** need not hold, because there are continuous homomorphisms that are bijective but have no continuous inverse.

In an abelian category monic and epic maps have a simple description:

PROPOSITION 1.3 *In an abelian category a map* $\alpha$ *is monic if and only if* $\ker \alpha = 0$ *and epic if and only if* $\text{coker}\,\alpha = 0$. *If* $\ker \alpha = \text{coker}\,\alpha = 0$, *then* $\alpha$ *is an isomorphism.*

*Proof.* By definition of $\ker \alpha$ we have $\lambda\alpha = 0$ iff $\lambda = \lambda'(\ker \alpha)$ for some $\lambda'$. Hence $\lambda = 0$ holds for all such maps $\lambda$ iff $\ker \alpha = 0$. This proves the first assertion; the second follows by duality. If $\alpha: X \to Y$ is such that $\ker \alpha = \text{coker}\,\alpha = 0$, then im $\alpha = Y$, coim $\alpha = X$ and $\alpha = \alpha'$ is an isomorphism. ■

We observe that this result often fails to hold in more general categories, e.g. in Rg the inclusion map $\mathbf{Z} \to \mathbf{Q}$ is both epic and monic but is clearly not an isomorphism.

A sequence of objects and maps in an abelian category

$$\cdots \to A_{n-1} \xrightarrow{\alpha_n} A_n \xrightarrow{\alpha_{n+1}} A_{n+1} \to \cdots$$

is called a *complex* if $\alpha_n \alpha_{n+1} = 0$ for all $n$. This means that im $\alpha_n$ is a subobject of $\ker \alpha_{n+1}$ for all $n$. If we have equality at $A_n$:im $\alpha_n = \ker \alpha_{n+1}$, the sequence is said to be *exact* at $A_n$. If the sequence is exact at each object, we speak of an *exact sequence*. It is clear that this generalizes the usage introduced for modules in A.1, **10.2**.

The simplest cases of exact sequences are $0 \to A \to 0$, which means that $A = 0$ ($A$ is the zero object), and $0 \to A \to B \to 0$, which means that $A \cong B$, by Prop. 1.3. The first non-trivial case is that of a *short exact sequence*:

$$0 \to A' \xrightarrow{\lambda} A \xrightarrow{\mu} A'' \to 0. \tag{5}$$

In the case of modules this indicates that $A'$ is isomorphic to a submodule of $A$, with quotient isomorphic to $A''$; thus $A$ is an extension of $A'$ by $A''$. In an abelian category we take (5) as the definition of an extension; thus we call $A$ an *extension* of $A'$ by $A''$ when there is an exact sequence (5). A monomorphism $\lambda$ clearly satisfies $\lambda = \ker \operatorname{coker} \lambda$, hence in a short exact sequence (5), $\lambda = \ker \mu$ and dually, $\mu = \operatorname{coker} \lambda$. The next case is that of an exact sequence with four non-zero terms; this arises for example when we analyse a general map $\alpha: X \to Y$ and obtain the exact sequence in the top line of the diagram on p. 73.

As for modules, the case of split exact sequences is important:

PROPOSITION 1.4 *Given maps $\alpha: X \to Y, \beta: Y \to X$ in an abelian category such that $\alpha\beta = 1_X$, the canonical composite* $\ker \beta \to Y \to \operatorname{coker} \alpha$ *is an isomorphism.*

*Proof.* Write $\alpha' = \operatorname{coker} \alpha, \beta' = \ker \beta$, so that $\alpha\alpha' = 0 = \beta'\beta$ ; we have to show that $\beta'\alpha'$ is an isomorphism. Since $\beta$ is epic, $\beta = \operatorname{coker} \beta'$; if $\beta'\alpha' f = 0$ for some $f$, then there exists $g$ such that $\alpha' f = \beta g$, by the definition of $\beta'$ as $\ker \beta$. Hence $g = \alpha\beta g = \alpha\alpha' f = 0$, so $\alpha' f = \beta g = 0$, but $\alpha'$ is epic and hence $f = 0$. This shows that $\beta'\alpha'$ is epic; by duality it is monic and so is an isomorphism. ■

When maps $\alpha, \beta$ are related as in Prop. 1.4, $\alpha$ is called a *section* and $\beta$ a *retraction.*

COROLLARY 1.5 *For a short exact sequence* (5) *in an abelian category the following conditions are equivalent:*

(a) $\lambda$ *is a section,*

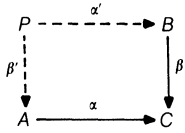
(b) $\mu$ *is a retraction,*

(c) $A \cong A' \coprod A''$ *for suitable maps* $\alpha: A \to A'$, $\beta: A'' \to A$.

*Proof.* (a)⇒(c) By hypothesis there is a map $\alpha$ such that $\lambda\alpha = 1$. Put $\nu: \ker \alpha \to A$ for the canonical inclusion. Since $\mu = \operatorname{coker} \lambda$, it follows from Prop. 1.4 that $\nu\mu$ is an isomorphism and on writing $\beta = (\nu\mu)^{-1}\nu: A'' \to A$, we find that $\beta\mu = 1$. We claim that $A$ is a biproduct of $A', A''$ relative to the maps $\lambda, \beta$; $\alpha, \mu$. Clearly $\lambda\mu = 0$, $\beta\alpha = 0$, so it only remains to show that $\alpha\lambda + \mu\beta = 1$. Write $f = \alpha\lambda + \mu\beta - 1_A$; then $f\mu = \mu\beta\mu - \mu = 0$, hence $f = f'\lambda$, and $f' = f'\lambda\alpha = f\alpha = \alpha\lambda\alpha + \mu\beta\alpha - \alpha = \alpha - \alpha = 0$; it follows that $f = 0$ as claimed. Thus $A \cong A' \coprod A''$; the converse is clear, hence (a)⇔(c), and now (b)⇔(c) follows by duality. ■

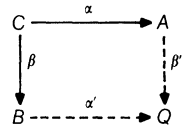
A short exact sequence is said to be *split exact* when the equivalent conditions of this Cor. are satisfied.

We shall need another very basic construction which can be described in quite arbitrary categories. Given two maps $\alpha: A \to C$, $\beta: B \to C$, with the same target, as shown overleaf, we consider the different ways of completing them to a commutative square. We can regard these commutative squares as



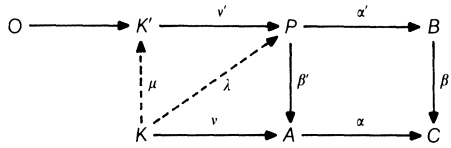

objects of another category, a morphism being a map between the new vertices so as to obtain a wedge with commutative faces. A final object in this category is called a *pullback* of $\alpha, \beta$. Explicitly, a pullback consists of maps $\alpha' : P \to B$, $\beta' : P \to A$ such that $\alpha'\beta = \beta'\alpha$ and given any other pair $\alpha'', \beta''$ such that $\alpha''\beta = \beta''\alpha$ there is a unique map $\lambda$ such that $\alpha'' = \lambda\alpha'$, $\beta'' = \lambda\beta'$. Borrowing from the terminology for rings, we may describe the pullback of $\alpha, \beta$ as a *least common left multiple* of $\alpha, \beta$. Like all universal constructions it is unique up to isomorphism, when it exists. In an abelian category the pullback always exists: let $A \prod B$ be the product of the sources of $\alpha, \beta$ with projections $p, q$ on $A, B$ respectively. Then it is easily verified that $\ker(p\alpha - q\beta)$ is a pullback of $\alpha, \beta$.

Dually, a *pushout* of two maps with the same source, $\alpha : C \to A$, $\beta : C \to B$ can be described as a least common right multiple of $\alpha, \beta$; the explicit definition is left to the reader. For example, if $A, B$ are groups and $C$ is a subgroup of $A$ and $B$, with inclusion homomorphisms, then their pushout is the 'free product of $A$ and $B$ amalgamating $C$'. Again pushouts always exist in an abelian category, by duality.




We note that the product of $A$ and $B$ is a special case of the pullback, obtained by taking $C$ to be a final object, and the coproduct is a special case of the pushout, obtained by taking $C$ to be an initial object. Sometimes the pullback is called 'fibre product' and the pushout the 'fibre sum'. We record a fact about pullbacks which is used later.

PROPOSITION 1.6 *Let $\mathscr{A}$ be an additive category. Given a pullback diagram as shown, if* $\ker \alpha'$ *exists, then* $\ker \alpha$ *exists and* $\ker \alpha \cong \ker \alpha'$. *A dual result holds in pushouts.*




*Proof.* Write $\ker \alpha' = (K', \nu')$; we shall show that $(K', \nu'\beta')$ is a kernel of $\alpha$. In the first place $\nu'\beta'\alpha = \nu'\alpha'\beta = 0$; secondly, if $\nu : K \to A$ is such that $\nu\alpha = 0$, then

the triangle $ABC$ can be completed by $K$ to a commutative square, with $0: K \to B$, hence there is a unique map $\lambda: K \to P$ such that $\lambda\beta' = \nu$, $\lambda\alpha' = 0$. Since $\nu' = \ker \alpha'$, there is a unique map $\mu: K \to K'$ such that $\mu\nu' = \lambda$, hence $\mu\nu'\beta' = \nu$ and this shows that $\nu$ can be factored uniquely by $\nu'\beta'$; therefore $\ker \alpha = (K', \nu'\beta')$, as claimed. ■

In particular we see that in a pullback in an abelian category, $\alpha'$ is monic iff $\alpha$ is monic. Consider a commutative square, as in the above diagram. This corresponds to a complex

$$0 \to P \xrightarrow{\lambda} A \coprod B \xrightarrow{\mu} C \to 0, \tag{6}$$

where $\lambda = (\beta' i, \alpha' j)$, $\mu = p\alpha - q\beta$ and $i, j, p, q$ are the natural injections and projections of the biproduct $A \coprod B$. The square is a pullback iff $P = \ker(p\alpha - q\beta)$, i.e. (6) is exact at $P$ and $A \coprod B$; it is a pushout iff $C = \operatorname{coker}(\beta' i, \alpha' j)$, i.e. (6) is exact at $A \coprod B$ and $C$. It follows that a pullback is also a pushout whenever $\mu$ is epic. Suppose now that $\alpha$ is epic and let $\nu$ be such that $\mu\nu = 0$. Then $\alpha\nu = ip\alpha\nu = i(p\alpha - q\beta)\nu = i\mu\nu = 0$, and hence $\nu = 0$; this means that $\mu$ is epic. Thus if in a pullback $\alpha$ is epic, then we have a pushout and so by Prop. 1.6, $\alpha'$ is also epic. This proves

COROLLARY 1.7 *Given a pullback diagram as in Prop.* 1.6 *in an abelian category, if* $\alpha$ *is epic, then so is* $\alpha'$. ■

## Exercises

(1) Show that Ens has an initial and a final object, but not a zero object.

(2) Show that in Rg the inclusion $\mathbf{Z} \to \mathbf{Q}$ is monic and epic but is not an isomorphism.

(3) Show that in a concrete category every monomorphism is injective.

(4) Show that any epimorphism of groups is surjective. (Hint: if $\alpha$ with target $G$ is not surjective, examine maps from $G$ to the group of all permutations on $G$.)

(5) Show that the pushout of two maps $\alpha: C \to A$, $\beta: C \to B$, one of which is the zero map, is $A \oplus B$.

(6) Show that $A \xrightarrow{\lambda} B \xrightarrow{\mu} C$ is exact iff the compositions $\operatorname{im}\lambda \to B \to \operatorname{coim}\mu$ and $\ker\mu \to B \to \operatorname{coker}\lambda$ are both zero.

(7) Show that in an abelian category a map $\alpha$ is monic iff $\alpha = \ker \operatorname{coker} \alpha$, and epic iff $\alpha = \operatorname{coker} \ker \alpha$.

(8) Let $\mathscr{A}$ be an abelian category; a subcategory $\mathscr{A}'$ is said to be *abelian* if with any

morphism $\alpha$, $\ker \alpha$ and $\operatorname{coker} \alpha$ (formed in $\mathscr{A}$) lie in $\mathscr{A}'$. Verify that $\mathscr{A}'$ is again an abelian category.

(9) ($3 \times 3$ lemma in abelian categories) Given three short exact sequences whose second and third terms, topped and tailed by 0's, form columns of exact sequences, so as to form a commutative diagram, show that there is just one way of filling in arrows between the first terms so as to make the diagram commutative, and the first column is then exact.

(10) (Windmill lemma) Given two rows of short exact sequences with a common middle term, written as a row and column, say, form the pullback of the NW-square, the pushout of the SE-square and factorize the SW- and NE-squares through their images. Show that the resulting diagram is commutative, with exact rows and columns. Deduce the second isomorphism theorem for abelian categories: $A_1/(A_1 \cap A_2) \cong (A_1 + A_2)/A_2$.

(11) Let $R$ be any ring. Given $R$-modules and homomorphisms $\alpha: A \to C$, $\beta: B \to C$, show that the pullback of $\alpha$ and $\beta$ is the submodule of $A \oplus B$ given by $\{(x, y) | x\alpha = y\beta\}$. Given $\alpha: C \to A$, $\beta: C \to B$, show that their pushout is $A \oplus B/K$, where $K = \{(z\alpha, z\beta) | z \in C\}$.

(12) Show that the pushout of two $K$-algebras $A, B$ relative to the natural maps $K \to A$, $K \to B$ is their tensor product over $K$.

## 3.2 Functors on abelian categories

Whenever we consider functors between additive categories we shall assume that they are additive; here $F: \mathscr{A} \to \mathscr{B}$ is called *additive* if $(\alpha + \alpha')^F = \alpha^F + \alpha'^F$, whenever $\alpha + \alpha'$ is defined; thus the mapping $\mathscr{A}(X, Y) \to \mathscr{B}(X^F, Y^F)$ is a group homomorphism. For example, in any additive category $\mathscr{A}$ the hom functors $h^A: X \mapsto \mathscr{A}(A, X)$ and $h_A: X \mapsto \mathscr{A}(X, A)$ are additive functors from $\mathscr{A}$ to Ab; on the other hand, the functor $X \mapsto \operatorname{Hom}(X^*, X)$ between vector spaces, where $X^*$ is the dual of $X$, is not additive. To give another example, an additive category with a single object is just a ring; now an additive functor between one-object categories is nothing other than a ring homomorphism, or an antihomomorphism in the case of a contravariant functor. Henceforth all functors are assumed to be additive unless otherwise stated.

If we apply an additive functor to a biproduct $(B, p_i, q_i)$, the defining equations between the $p$'s and $q$'s are preserved, hence the result is again a biproduct. By Th. 1.1 we obtain

PROPOSITION 2.1 *Any additive functor acting on an abelian category preserves finite products, coproducts and biproducts.* ■

Clearly any functor takes zero maps to zero maps and hence transforms a complex into a complex. We shall be particularly interested in functors that preserve exactness. A functor $T$ is said to be *exact* if it transforms each exact

sequence

$$A \xrightarrow{\lambda} B \xrightarrow{\mu} C \tag{1}$$

into an exact sequence

$$A^T \xrightarrow{\lambda^T} B^T \xrightarrow{\mu^T} C^T. \tag{2}$$

For example, an equivalence between categories is an exact functor. We recall (A.2, **1.3**) that two categories $\mathscr{A}$, $\mathscr{B}$ are *equivalent* if there are two functors $T:\mathscr{A} \to \mathscr{B}$, $S:\mathscr{B} \to \mathscr{A}$ such that $TS$ is naturally isomorphic to the identity functor on $\mathscr{A}$, and similarly $ST$ is naturally isomorphic to the identity on $\mathscr{B}$. Any functor $T:\mathscr{A} \to \mathscr{B}$ defines for each pair $X, Y$ of $\mathscr{A}$-objects a mapping

$$\mathscr{A}(X, Y) \to \mathscr{B}(X^T, Y^Y). \tag{3}$$

The functor $T$ is called *faithful* if (3) is injective and *full* if (3) is surjective. For an equivalence functor $T$, (3) is a bijection, so in this case $T$ is full and faithful. Moreover, an equivalence functor $T$ is *dense* in the sense that every $\mathscr{B}$-object is isomorphic to one of the form $X^T$, for some $X \in \mathrm{Ob}\ \mathscr{A}$. As we saw in A.2, Prop. 1.3.1, a functor $T$ is an equivalence iff it is full, faithful and dense.

All this holds in quite arbitrary categories; when $\mathscr{A}$, $\mathscr{B}$ are additive (and by assumption $T$ is an additive functor), (3) is clearly a group homomorphism, and it follows easily from this that any equivalence functor is again exact.

However, most functors satisfy only a weaker condition. We define a functor to be *left exact* if it preserves kernels and *right exact* if it preserves cokernels. First we have a restatement of this condition:

PROPOSITION 2.2 *A functor between abelian categories $T:\mathscr{A} \to \mathscr{B}$ is left exact if and only if the exactness of*

$$0 \to A \xrightarrow{\lambda} B \xrightarrow{\mu} C \tag{4}$$

*implies that of*

$$A^T \xrightarrow{\lambda^T} B^T \xrightarrow{\mu^T} C^T. \tag{5}$$

*Similarly $T$ is right exact if and only if it preserves exactness when the* 0 *in* (4) *is at the other end* (*i.e. when coker $\mu = 0$*).

*Proof*. The exactness of (4) is expressed by the equation $\lambda = \ker \mu$. If $T$ preserves kernels, it follows that $\lambda^T = \ker \mu^T$ and so (5) is exact. Conversely, if (5) is exact, by applying $T$ to the exact sequence $0 \to 0 \to A \to B$, we find that the sequence $0 \to A^T \to B^T$ is exact, as well as (5), so $\lambda^T = \ker \mu^T$, and this shows $T$ to be left exact. Similarly for right exactness. ■

COROLLARY 2.3 *A functor $T$ between abelian categories is exact if and only if it is left and right exact.*

*Proof.* Clearly an exact functor is left and right exact; conversely, if $T$ is left and right exact, it preserves kernels and cokernels, hence images and coimages. By hypothesis, $\operatorname{im}\lambda = \ker\mu$ in (1), hence $\operatorname{im}\lambda^T = (\operatorname{im}\lambda)^T = (\ker\mu)^T = \ker\mu^T$, so $T$ is indeed exact. ■

We note that (1) is exact iff the sequence

$$0 \to \operatorname{im}\lambda \to B \to \operatorname{coim}\mu \to 0$$

is exact. Thus if $T$ transforms short exact sequences to short exact sequences, then it is exact. The converse is clear, so we have

COROLLARY 2.4 *A functor between abelian categories is exact if and only if it preserves the exactness of short exact sequences.* ■

So far all functors were tacitly assumed to be covariant. If $T:\mathscr{A} \to \mathscr{B}$ is a contravariant functor, we shall call $T$ *left exact* if the covariant functor $\mathrm{op}\cdot T:\mathscr{A}^{\mathrm{op}} \to \mathscr{B}$ is left exact. *Right exact* contravariant functors are defined correspondingly, by the right exactness of $\mathrm{op}\cdot T$. The reason for this form of the definition (rather than using $T\cdot\mathrm{op}:\mathscr{A} \to \mathscr{B}^{\mathrm{op}}$) is to be found in

THEOREM 2.5 *For any abelian functor $\mathscr{A}$, the bifunctor $\mathscr{A}(A, B)$ is left exact in each argument, i.e. $h^A$, $h_B$ are both left exact.*

*Proof.* For any $\mu:B \to B''$ in $\mathscr{A}$ the kernel of the induced mapping $\mathscr{A}(A,\mu):\mathscr{A}(A,B) \to \mathscr{A}(A,B'')$ is the set of morphisms killed by $\mu$, i.e. the maps that factor uniquely through $\ker\mu$. Thus

$$\ker\mathscr{A}(A,\mu) = \mathscr{A}(A, \ker\mu),$$

hence $h^A$ is left exact, as claimed. Similarly, for any $\lambda:A' \to A$ in $\mathscr{A}$,

$$\ker\mathscr{A}(\lambda, B) = \mathscr{A}(\operatorname{coker}\lambda, B),$$

therefore $h_B$ is left exact. ■

Let $\mathscr{A}$ be any category. A functor $F$ from $\mathscr{A}$ to sets is said to be *representable* if there is an $\mathscr{A}$-object $P$ such that $X^F = \mathscr{A}(P, X)$; in other words, $F$ is then naturally isomorphic to $h^P$ and one also says that $F$ is *represented* by $P$. When the category $\mathscr{A}$ is abelian, we can similarly define the representability of a functor from $\mathscr{A}$ to abelian groups. A contravariant functor $G$ from $\mathscr{A}$ is called *representable* if there is an $\mathscr{A}$-object $Q$ such that $Y^G = \mathscr{A}(Y, Q)$, thus $G$ is naturally isomorphic to $h_Q$. For example, the dual of a vector space is represen-

table, almost by definition: $V^* = \mathrm{Hom}_k(V, k)$. To give another example, consider $\mathbf{U}(R)$, the group of units of a ring $R$. It can be shown that this functor is representable by the infinite cyclic group $\mathbf{C}$, thus $\mathbf{U}(R) \cong \mathrm{Mon}(\mathbf{C}, R)$, where Mon is the category of monoids and $R$ is considered as multiplicative monoid.

Sometimes we shall need a criterion for a functor to preserve inexact sequences; a sequence $\xrightarrow{\lambda} \xrightarrow{\mu}$ is called *inexact* if it is not exact, i.e. if $\mathrm{im}\,\lambda \neq \ker \mu$.

PROPOSITION 2.6 *A functor $T$ between abelian categories preserves inexact sequences if and only if it is faithful.*

*Proof.* Suppose first that $T$ preserves inexact sequences; we must show that $T$ is faithful, i.e. $\alpha \neq 0$ implies $\alpha^T \neq 0$. Given $\alpha: A \to B$, where $\alpha \neq 0$, the sequence $A \xrightarrow{1} A \xrightarrow{\alpha} B$ is inexact, hence it remains so on applying $T$, i.e. $\ker \alpha^T \neq A^T$ and so $\alpha^T \neq 0$.

Conversely assume that $T$ is faithful and consider the sequence (2), p. 79. If this is exact, then $(\lambda\mu)^T = \lambda^T\mu^T = 0$, hence $\lambda\mu = 0$. Now let $\ker \mu = (B', i)$ and consider the composition $B' \xrightarrow{i} B \xrightarrow{\mu} C$. This is zero, hence so is $B'^T \xrightarrow{i^T} B^T \xrightarrow{\mu^T} C^T$ and it gives rise to a map $(\ker \mu)^T \to \ker \mu^T$. Similarly we have a map $\mathrm{coker}\,\lambda^T \to (\mathrm{coker}\,\lambda)^T$, and the sequence

$$(\ker \mu)^T \to \ker \mu^T \to B^T \to \mathrm{coker}\,\lambda^T \to (\mathrm{coker}\,\lambda)^T$$

is exact at $B^T$; hence the composition $(\ker \mu)^T \to B^T \to (\mathrm{coker}\,\lambda)^T$ is zero, and it follows that the sequence (1) p. 79 is exact at $B$. ■

There is a useful test for exactness in the case of adjoint functors. Given two functors $T: \mathscr{A} \to \mathscr{B}$, $S: \mathscr{B} \to \mathscr{A}$, we call $S, T$ an *adjoint pair*, or more precisely, $S$ a *left adjoint* and $T$ a *right adjoint* if for any $\mathscr{A}$-object $X$ and $\mathscr{B}$-object $Y$,

$$\mathscr{A}(Y^S, X) \cong \mathscr{B}(Y, X^T), \tag{6}$$

where (in the case of additive categories) $\cong$ is an isomorphism of abelian groups which is natural in $X$ and $Y$. The notion of an adjoint pair can of course be defined in quite general categories; then (6) is merely a bijection of sets (still natural in $X$ and $Y$). For example, if $U: \mathrm{Gp} \to \mathrm{Ens}$ is the forgetful functor from groups, associating with each group its underlying set, then

$$\mathrm{Gp}(F_X, G) \cong \mathrm{Ens}(X, G^U),$$

where $F_X$ is the free group on $X$. Generally nearly every universal construction arises as the left adjoint of a forgetful functor. To give another example, if $i: \mathrm{Ab} \to \mathrm{Gp}$ is the inclusion functor and $\mathrm{ab}: \mathrm{Gp} \to \mathrm{Ab}$ is abelianization (i.e. passing

from a group $G$ to $G^{ab} = G/G'$, cf. A.1, **9.5** or **2.2** above), then

$$\mathrm{Ab}(G^{ab}, A) \cong \mathrm{Gp}(G, iA).$$

The typical construction described by a right adjoint singles out a subset by some closure operation; see e.g. Ex. (8).

Returning to the general case of an adjoint pair (6), we observe that either of $S, T$ determines the other up to natural isomorphism, for if we had

$$\mathscr{B}(Y, X^T) \cong \mathscr{B}(Y, X^{T'}), \tag{7}$$

let us first take $Y = X^T$ and denote by $\alpha: X^T \to X^{T'}$ the map on the right of (7) corresponding to the identity map on the left; next take $Y = X^{T'}$ and let $\beta: X^{T'} \to X^T$ be the map corresponding to the identity map on the right. Then $\alpha\beta = 1_{X^T}$, $\beta\alpha = 1_{X^{T'}}$, so $\alpha$ is a natural isomorphism.

We also note that the hom functor as a bifunctor is faithful. Taking for example, $h^A: X \mapsto \mathscr{A}(A, X)$, we have for $\alpha: X \to Y$, $h^\alpha: \lambda \mapsto \lambda\alpha$, thus $h^\alpha$ is right multiplication by $\alpha$, and choosing $\lambda = 1_X$ we find that $\lambda\alpha = 0$ for all $\lambda$ implies $\alpha = 0$; similarly for $h_B$. With these preparations we have

THEOREM 2.7 *Let $S$ and $T$ be a pair of adjoint functors between abelian categories $\mathscr{A}, \mathscr{B}$. Then the left adjoint $S$ is right exact and the right adjoint $T$ is left exact. Moreover, if $\mathscr{A}$, $\mathscr{B}$ have arbitrary products and coproducts, then $T$ preserves products and $S$ preserves coproducts.*

*Proof.* Let us apply (6) to a short exact sequence

$$0 \to X' \xrightarrow{\lambda} X \xrightarrow{\mu} X''.$$

We obtain a commutative diagram of complexes

$$\begin{array}{ccccccc} 0 & \longrightarrow & \mathscr{A}(Y^S, X') & \longrightarrow & \mathscr{A}(Y^S, X) & \longrightarrow & \mathscr{A}(Y^S, X'') \\ & & \downarrow \cong & & \downarrow \cong & & \downarrow \cong \\ 0 & \longrightarrow & \mathscr{B}(Y, X'^T) & \longrightarrow & \mathscr{B}(Y, X^T) & \longrightarrow & \mathscr{B}(Y, X''^T) \end{array} \tag{8}$$

By Th. 2.5 the top row is exact, hence so is the bottom row, and this arises by applying the functor $h^Y$ to the sequence

$$0 \to X'^T \to X^T \to X''^T. \tag{9}$$

But $h^Y$, when $Y$ is allowed to vary, is faithful and so preserves inexact sequences; since the bottom row in (8) is exact, so is (9). This proves $T$ to be left exact, and it preserves products, by (6). A dual argument shows $S$ to be right exact and to preserve coproducts. ■

Let $\mathscr{C}$ be an abelian category and $I$ a partially ordered set, regarded as a small

category. We denote by $\mathscr{C}^I$ the functor category whose objects are functors from $I$ to $\mathscr{C}$ with natural transformations as morphisms; thus the objects are families of $\mathscr{C}$-objects indexed by $I$, with families of $\mathscr{C}$-maps as morphisms. Explicitly a $\mathscr{C}^I$-object $(A_i, \alpha_{ij})$ consists of $\mathscr{C}$-objects $A_i$ with $\mathscr{C}$-maps $\alpha_{ij}: A_i \to A_j$ for $i \leqslant j$, such that

$$\alpha_{ii} = 1 \qquad \alpha_{ij}\alpha_{jk} = \alpha_{ik} \qquad (i \leqslant j \leqslant k). \tag{10}$$

These conditions (10) are called the *coherence conditions.* A morphism $f: (A_i, \alpha_{ij}) \to (B_i, \beta_{ij})$ is a family of maps $f_i: A_i \to B_i$ such that $\alpha_{ij} f_j = f_i \beta_{ij}$ for $i \leqslant j$.

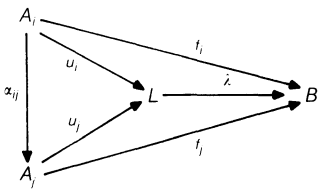
We have the *diagonal functor*

$$\Delta: \mathscr{C} \to \mathscr{C}^I, \tag{11}$$

which, with each $\mathscr{C}$-object $A$, associates the constant family $(A_i, \alpha_{ij})$ with $A_i = A$, $\alpha_{ij} = 1$. The adjoint functors of $\Delta$ play an important role. The *direct limit* (also called *inductive limit* or *colimit*) $\varinjlim$ is defined as the left adjoint of $\Delta$:

$$\mathscr{C}(\varinjlim(A_i, \alpha_{ij}), B) = \mathscr{C}^I((A_i, \alpha_{ij}), \Delta(B)). \tag{12}$$

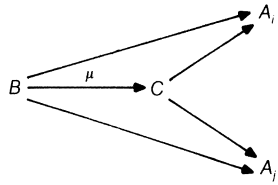
If $L = \lim(A_i)$ exists, there are maps $u_i: A_i \to L$ satisfying $u_i = \alpha_{ij} u_j$ for $i \leqslant j$, such that any map $(f_i)$ from $(A_i, \alpha_{ij})$ to $\Delta(B)$ can be factored uniquely by $(u_i)$, thus $f_i = u_i \lambda$ for all $i \in I$, for a unique $\lambda: L \to B$.




For example, when $I$ is totally unordered, the direct limit reduces to the coproduct. If $I$ consists of three points $i, j, k$ with $i < j$, $i < k$, we obtain the pushout. Let us describe direct limits for modules. The construction is simplified if we assume $I$ to be a directed ordered set (i.e. given $i, j \in I$, there exists $k \geqslant i, j$); in that case the family $(A_i, \alpha_{ij})$ is also called a *direct family.* The direct limit $L$ is the direct sum of the $A_i$ modulo the submodule generated by the elements $x - x\alpha_{ij}$ $(x \in A_i)$ for $i \leqslant j$. For an example from field theory let $F$ be a field and $(E_i)$ the family of all extensions of $F$ of finite degree, with inclusions as mappings; then it is clear that we have a direct family. Its direct limit is a field $\Omega$ containing $F$, which is algebraic over $F$ and algebraically closed, and hence is the algebraic closure of $F$ (cf. A.2, **3.3** and **5.6** below).

Similarly the *inverse limit* (also *projective limit* or simply *limit*) $\varprojlim$ is defined as the right adjoint of $\Delta$:

$$\mathscr{C}(B, \varprojlim(A_i, \alpha_{ij})) = \mathscr{C}^I(\Delta(B), (A_i, \alpha_{ij})).$$




The inverse limit $C$ has maps $v_i: C \to A_i$ such that $v_i\alpha_{ij} = v_j$ and any map $(f_i)$ from $\Delta(B)$ to $(A_i, \alpha_{ij})$ can be factored uniquely by $(v_i)$, thus $f_i = \mu v_i$ for all $i \in I$, for a unique $\mu: B \to C$.

When $I$ is totally unordered, this reduces to the product of the $A_i$. For a triple $i, j, k$ with $i < k, j < k$ it becomes the pullback. To describe the construction for modules we take $I$ inversely directed and refer to $(A_i, \alpha_{ij})$ as an *inverse family*. The inverse limit of such a family of modules is obtained by forming the product $\prod A_i$ and taking the submodule of all elements $(x_i)$ such that $x_j\alpha_{ji} = x_i$.

To illustrate the notion of inverse limit (and for later use, in **5.6** below) consider a free group $F$ of rank greater than 1. Let us write $(N_i)$ for the family of all normal subgroups of finite index in $G$. Then $G_i = F/N_i$ is a finite group and for $N_i \subseteq N_j$ we have a natural homomorphism $\varphi_{ij}: G_i \to G_j$, and these homomorphisms are coherent. Since the intersection of any two $N_i$ is again of finite index, we have an inverse family and we can form the inverse limit $G = \varprojlim(G_i)$. This group $G$ is called a *profinite* group (as *projective* limit of *finite* groups). Since the natural homomorphisms $F \to G_i$ are compatible with the $\varphi_{ij}$, we have a canonical homomorphism $\gamma: F \to G$. As we shall see in **4.5**, $\cap N_i = 1$, and it follows easily from this that $\gamma$ is injective. However, it is not surjective, for $G$ as inverse limit is uncountable, whereas $F$ is countable whenever its rank is at most countable. A similar construction is possible for abelian groups, thus e.g. $\mathbf{Z}$ can be embedded in a profinite group, or even in a pro-$p$-group (cf. Ex. (10)).

The last example illustrates another important point, namely the lack of duality in general module categories. As we have seen, the notion of an abelian category can be developed in an entirely self-dual manner. However, the category of all modules over a ring is not self-dual, except for very special rings, so in order to describe module categories axiomatically one will need axioms whose duals may not hold. We shall not carry out the full axiomatization (which can be found in most books on category theory) but merely list one axiom holding in all module categories but not always for their duals. This is Grothendieck's

AB5 Axiom. *Given a chain of subobjects $(A_i)$ and any subobject $B$ of an object, we have*

$$(\cup A_i) \cap B = \cup (A_i \cap B). \tag{14}$$

The following equivalent form will be more convenient for us:

*I. The functor $\varinjlim$ is exact.*

Since $\varinjlim$ is in any case right exact, as left adjoint, this requires that for any families of exact sequences $0 \to A_i \to B_i$ the sequence

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

be again exact. Like (14) this condition is easily verified in any category of modules. The dual states

$I^0$. *The functor* $\varprojlim$ *is exact.*

By duality we need only verify that the exactness of $A_i \to B_i \to 0$ entails that of $\varprojlim(A_i) \to \varprojlim(B_i) \to 0$. In most module categories this does not hold. For example taking $A_i = \mathbf{Z}$ and $B_i$ the family of finite images, we have $\varprojlim(A_i) = \mathbf{Z}$, but the mapping $\mathbf{Z} \to \hat{\mathbf{Z}} = \varprojlim(B_i)$ is far from surjective, since the limit $\hat{\mathbf{Z}}$ is uncountable.

In A.2, **4.5** we have already met projective and injective modules. The counterpart in abelian categories is of importance, because it can be used to remedy the lack of exactness of the hom functor (Th. 2.5).

DEFINITION Let $\mathscr{A}$ be an abelian category. An $\mathscr{A}$-object $P$ is called *projective* if the covariant functor $h^P = \mathscr{A}(P, -)$ is exact; an $\mathscr{A}$-object $I$ is called *injective* if the contravariant hom functor $h_I = \mathscr{A}(-, I)$ is exact.

An alternative description of projective objects is given in

THEOREM 2.8 *Let $P$ be an object in an abelian category $\mathscr{A}$. Then the following conditions are equivalent:*

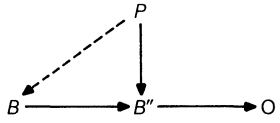
(a) *$P$ is projective,*

(b) *every short exact sequence*

$$0 \to A \xrightarrow{\lambda} B \xrightarrow{\mu} P \to 0 \tag{15}$$

*with $P$ in third place splits,*

(c) *given a diagram with exact row as shown, there exists a map $P \to B$ to make the triangle commutative.*



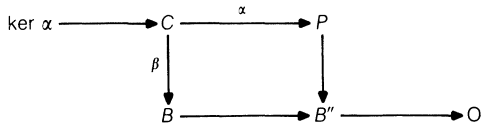

Condition (c) may be expressed by saying: every map from $P$ to a quotient of $B$ may be lifted to $B$. We note that the statement of this theorem is quite similar to that of Th. 4.5.4 of A.2 for modules, but we shall not be able to use the proof given there (which depended on the existence of free modules).

*Proof.* (a)⇒(b) Given a short exact sequence (15), we find by (a) that the sequence of abelian groups

$$0 \to \mathscr{A}(P, A) \to \mathscr{A}(P, B) \to \mathscr{A}(P, P) \to 0$$

is exact. Now $1_P \in \mathscr{A}(P, P)$ and by exactness there exists $\beta \in \mathscr{A}(P, B)$ such that $\beta\mu = 1_P$; hence (15) splits.

(b)⇒(c). If we form the pullback of the given diagram, we obtain




By Cor. 1.7, $\alpha$ is epic, hence by (b) the top row splits, so there is a map $f: P \to C$ and $f\beta: P \to B$ is the required map.

(c)⇒(a). Given a short exact sequence, we form

$$0 \to \mathscr{A}(P, B') \to \mathscr{A}(P, B) \to \mathscr{A}(P, B'') \to 0. \tag{16}$$

By the left exactness of hom this can fail to be exact only at $\mathscr{A}(P, B'')$. But by (c) every map $P \to B''$ lifts to a map $P \to B$, and this means that (16) is also exact at $\mathscr{A}(P, B'')$. ■

Of course there is a dual characterization of injectives:

THEOREM 2.8$^0$ *Let $I$ be an object in an abelian category $\mathscr{A}$. Then the following conditions are equivalent:*

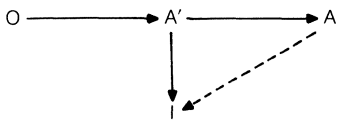
(a) *$I$ is injective,*

(b) *every short exact sequence*

$$0 \to I \to B \to C \to 0$$

*with $I$ in first place splits,*

(c) *given a diagram with exact row as shown, there is a map $A \to I$ such that the resulting triangle is commutative.*




Here (c) may be expressed by saying: every map from a subobject of $A$ to $I$ can be extended to $A$.

The proof is dual to that of Th. 2.8 and so may be omitted. ■

Although the notions of projective and injective object are dual, they can have very different appearance in actual categories and we shall return to the question for module categories in **3.3** and Ch. 6.

### Exercises

(1) Show that a functor between additive categories is additive iff it preserves finite products.

(2) Show that a subcategory of an abelian category is abelian iff the inclusion functor is exact.

(3) Show that for a faithful functor $T$, $C \neq 0$ implies $C^T \neq 0$. In the case of an exact functor this condition is sufficient as well as necessary.

(4) Let $T:\mathscr{A} \to \mathscr{B}$, $S:\mathscr{B} \to \mathscr{A}$ be a pair of functors giving an equivalence of categories. Show that $S, T$ is an adjoint pair as well as $T, S$.

(5) Show that a functor between additive categories forming part of an adjoint pair is necessarily additive. (Hint: verify that it preserves 0 and finite products).

(6) Show that for any abelian category the following are equivalent: (a) every object is projective, (b) every object is injective, (c) every short exact sequence splits.

(7) Let $S, T$ be a pair of adjoint functors. Show that if $S$ is left exact, then $T$ preserves injectives; if $T$ is right exact then $S$ preserves projectives.

(8) For any group $G$ denote by $\mathbf{Z}G$ the group ring of $G$ over $\mathbf{Z}$. Show that $G \mapsto \mathbf{Z}G$ is a functor from Gp to Rg whose right adjoint is the functor $R \to \mathbf{U}(R)$, where $\mathbf{U}(R)$ is the group of units of $R$.

(9) Show that a functor $\mathscr{A} \to \mathscr{B}$ is full and faithful iff $\mathscr{A}$ is equivalent to a full subcategory of $\mathscr{B}$.

(10) Let $p$ be a prime number. Verify that $\cap p^n\mathbf{Z} = 0$ and deduce that there is a natural injection $\mathbf{Z} \to \varprojlim(\mathbf{Z}/p^n)$.

## 3.3 The category $\mathscr{M}_R$.

We have already seen that the category $\mathscr{M}_R$ of all right-$R$-modules, for any ring $R$, is abelian. Frequently $R$ will be a $K$-algebra (associative, with 1), where $K$ is some commutative ring; in that case the hom sets $\mathrm{Hom}_R(M, N)$ are $K$-modules and not merely abelian groups. We shall say that we have a *$K$-linear* category in that case. A functor $F$ between $K$-linear categories is required to be (not merely additive but also) *$K$-linear*:

$$(\alpha + \alpha')^F = \alpha^F + \alpha'^F \qquad (\lambda\alpha)^F = \lambda \cdot \alpha^F \qquad (\lambda \in K).$$

As a rule $K$ will be an arbitrary commutative ring, fixed in any given context, and all rings will be $K$-algebras. The case of abstract rings is included by taking $K = \mathbf{Z}$.

We recall that a right $R$-module structure on $M$ can be described by saying that we have a homomorphism

$$f: R \to \mathrm{End}_K(M). \tag{1}$$

Similarly, a left $R$-module structure on $M$ corresponds to an antihomomorphism (1), i.e. a homomorphism $R^0 \to \mathrm{End}_K(M)$ from the opposite ring; in detail this is an additive mapping $f$ such that $(xy)f = yf \cdot xf$. This remark is often used to avoid having to pass to the opposite ring. Thus if we have a homomorphism $R^0 \to \mathrm{End}_K(M)$, we shall regard $M$ as a left $R$-module rather than a right $R^0$-module.

Let $R, T$ be any rings and ${}_T\mathscr{M}_R$ the category of $(T, R)$-bimodules; clearly this is a subcategory of $\mathscr{M}_R$. The following lemma on the transport of the ring action is often useful:

LEMMA 3.1 *Let $R, S, T$ be any rings (or $K$-algebras) and $F: \mathscr{M}_R \to \mathscr{M}_S$ a covariant functor. Then $F$ induces a functor $F': {}_T\mathscr{M}_R \to {}_T\mathscr{M}_S$. Similarly, a contravariant functor $G: \mathscr{M}_R \to {}_S\mathscr{M}$ induces a functor $G': {}_T\mathscr{M}_R \to {}_S\mathscr{M}_T$.*

*Proof.* Let $M$ be a $(T, R)$-bimodule. We know that $M^F$ is an $S$-module; further, given $t \in T$, we can define the action of $t$ on $M^F$ as $t^F$. Since $t$ defines an endomorphism $M_R$, $t^F$ defines an endomorphism of $(M^F)_S$, i.e. an element of $\mathrm{End}_S(M^F)$, and so

$$(xa)t^F = (xt^F)a \qquad \text{for any } x \in M^F,\ a \in S.$$

We claim that the rule $t \mapsto t^F$ defines an antihomomorphism of $T$ into $\mathrm{End}_S(M)$. For if $t, t' \in T$, then $(tt')^F = t'^F t^F$, because $M$ is a *left* $T$-module. Hence $M^F$ is indeed a $(T, S)$-bimodule. Moreover, a homomorphism $\alpha$ between $(T, R)$-bimodules may be characterized as an $R$-homomorphism centralizing $T$, hence $\alpha^F$ is an $S$-homomorphism centralizing $T$, i.e. a homomorphism between $(T, S)$-bimodules. The second part, referring to $G$, is proved similarly; since $G$ is contravariant, it defines a *homomorphism* of $T$ this time, which means that $M^G$ is an $(S, T)$-bimodule. ■

As an example, important in what follows, let us consider the hom functor. Let $M$ be an $(S, R)$-bimodule and $N$ a $(T, R)$-bimodule; we shall express this briefly by saying that we are in the situation $({}_SM_R, {}_TN_R)$. Consider $H = \mathrm{Hom}_R(M, N)$; when we consider $M, N$ as right $R$-modules, $H$ is just an abelian group (or $K$-module). But the left $T$-module structure on $N$ induces a left $T$-module structure on $H$, while the left $S$-module structure on $M$ induces a right $S$-module structure on $H$; here the side was reversed because Hom is

contravariant in $M$. Thus we see that $H$ is a left $T$-, right $S$-module; in fact it is a $(T, S)$-bimodule. To show this explicitly, let us write $(f, x)$ for the effect of $f \in H$ on $x \in M$. Then by definition, $(f, x)r = (f, xr)$, $(fs, x) = (f, sx)$, $(tf, x) = t(f, x)$ for $r \in R$, $s \in S$, $t \in T$. Hence $((tf)s, x) = (tf, sx) = t(f, sx) = t(fs, x) = (t(fs), x)$, i.e. $(tf)s = t(fs)$, as claimed.

A second functor of great importance is the tensor product. We recall from A.2, **4.7** that for a $K$-algebra $R$ and modules $({}_{U_R}, {}_RV)$ there is a $K$-module $U \otimes_R V$ with a mapping

$$\lambda: U \times V \to U \otimes_R V,$$

which is universal for $K$-bilinear mappings $f$ from $U \times V$ to $K$-modules which are *R-balanced*, i.e. such that

$$(xr, y)f = (x, ry)f \qquad \text{for all } x \in U, y \in V, r \in R.$$

We remark that (assuming the tensor product over the commutative ring $K$ as known), $U \otimes_R V$ may also be obtained as the homomorphic image of $U \otimes_K V$ by the relations $xr \otimes y - x \otimes ry$. Further we recall the equations of adjoint associativity which follow from the definition of the tensor product. For the situation $({}_QU_R, {}_RV_S, {}_TW_S)$ we have the natural isomorphism of $(T, Q)$-bimodules (adjoint associativity)

$$\mathrm{Hom}_S(U \otimes_R V, W) \cong \mathrm{Hom}_R(U, \mathrm{Hom}_S(V, W)). \tag{2}$$

This may be expressed by saying that $- \otimes_R V$ is the left adjoint of the functor $h^V = \mathrm{Hom}_S(V, -)$. By symmetry the same holds for $U \otimes_R -$ in the situation $({}_SU_R, {}_RV_T, {}_SW_Q)$, using the isomorphism

$$\mathrm{Hom}_S(U \otimes_R V, W) \cong \mathrm{Hom}_R(V, \mathrm{Hom}_S(U, W)). \tag{3}$$

By Th. 2.7 we conclude

PROPOSITION 3.2 *For a left R-module V over any ring R, the tensor product functor* $- \otimes_R V$ *is right exact and preserves direct sums; similarly for right R-modules.* ■

We further recall the associative law for tensor products; in the situation $({}_{U_R}, {}_RV_S, {}_SW)$ we have

$$U \otimes_R (V \otimes_S W) \cong (U \otimes_R V) \otimes_S W. \tag{4}$$

We also recall the identity

$$U \otimes_R R \cong U \qquad \text{for } U_R; \tag{5}$$

it corresponds to the well known identity for the hom functor

$$\mathrm{Hom}_R(R, U) \cong U. \tag{6}$$

We have already met projective and injective objects in the category of

modules in A.2, **4.5**. In particular we see from the characterization given there that the projective $R$-modules are precisely the direct summands of free modules. There is no such explicit description of injective modules (but cf. **6.6** below); for the moment we note that by Th. $2.8^0$ a module $M$ is injective iff every short exact sequence with $M$ as first term splits, i.e. iff $M$ is a direct summand in every module containing it as submodule. This leads to the following criterion. An extension of modules $M \subseteq N$ is said to be *essential* if $M$ has a non-zero inersection with every non-zero submodule of $N$.

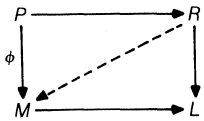
PROPOSITION 3.3 *An R-module is injective if and only if it has no proper essential extension.*

*Proof.* Suppose that $M$ is injective. If $M$ is contained as a submodule in $N$, then it is a direct summand, so if $N \neq M$, the extension is not essential, thus $M$ has no proper essential extensions. Conversely, assume that $M$ has no proper essential extensions, and let $L$ be any module containing $M$ as submodule. The family of all submodules of $L$ meeting $M$ in 0 is clearly inductive and so (by Zorn's lemma) has a maximal member, $L_0$ say. Consider $\bar{L} = L/L_0$; since $M \cap L_0 = 0$, $M$ maps isomorphically to a submodule $\bar{M}$ of $\bar{L}$, and by the maximality of $L_0$, $\bar{L}$ is an essential extension of $\bar{M} \cong M$. Hence it cannot be proper, so $\bar{L} = \bar{M}$, i.e. $M + L_0 = L$ and $M \cap L_0 = 0$, so $M$ is a direct summand in $L$. It follows that $M$ is injective, as we had to show. ■

We have already met Baer's criterion for injectivity in A.2, Th. 4.5.7; here is another very short proof, due to P. Freyd.

THEOREM 3.4 (Baer's criterion) *Let R be a ring. A left R-module M is injective if and only if every homomorphism from a left ideal of R into M can be extended to a homomorphism from R into M.*

*Proof.* The necessity is clear; to prove the sufficiency of the condition we show that when it holds, $M$ has no proper essential extension. Let $M \subset L$ be a proper extension, fix $u \in L \setminus M$ and consider the pullback diagram shown, where $R \to L$ is the map $r \mapsto ru$.




By Prop. 1.6, the map $P \to R$ is monic, so $P$ is isomorphic to a left ideal of $R$ and by hypothesis the map $P \to M$ extends to a homomorphism $R \to M$. If $1 \mapsto v$ in this homomorphism, we have $x \mapsto xv = xu$ for all $x \in P$, and $x \in P$ if $xu \in M$, hence $R(u - v)$ is a submodule of $L$ such that $R(u - v) \cap M = 0$,

but $R(u-v) \neq 0$, because $v \in M$, $u \notin M$, so $u \neq v$. Thus $L$ is not an essential extension and since $L$ was arbitrary, it follows by Prop. 3.3 that $M$ is injective. ■

With every homomorphism of rings there are several transfer functors associated which are often useful. Given any rings $R, S$ and a homomorphism $f: R \to S$, any $S$-module $U$ may be defined as an $R$-module by putting

$$x \cdot a = xa^f \qquad \text{for } x \in U,\ a \in R.$$

This $R$-action on $U$ is said to be defined by *pullback along* $f$ (not to be confused with the pullback diagram in **3.2**), and the resulting $R$-module is written ${}^fU$. The correspondence $U \mapsto {}^fU$ is a functor from $\mathscr{M}_S$ to $\mathscr{M}_R$ rather like the forgetful functor. We shall want to go in the opposite direction and construct an adjoint; thus we are given an $R$-module $A$ and we ask for an associated $S$-module. There are two constructions, arising as the left adjoint and the right adjoint of the functor ${}^fU$; they are known as the *change-of-rings* constructions.

PROPOSITION 3.5 *Let $R, S$ be rings and $f: R \to S$ a homomorphism. Given a right $R$-module $A$, there is an $S$-module $A_f = A \otimes_R S$ left adjoint to ${}^fU$:*

$$\mathrm{Hom}_S(A_f, U) \cong \mathrm{Hom}_R(A, {}^fU), \tag{7}$$

*and an $S$-module $A^f = \mathrm{Hom}_R(S, A)$ right adjoint to ${}^fU$:*

$$\mathrm{Hom}_S(U, A^f) \cong \mathrm{Hom}_R({}^fU, A). \tag{8}$$

*Moreover, there is a map $\alpha: A \to A_f$ which induces a homomorphism of $R$-modules (from $A$ to ${}^f(A_f)$) and a map $\beta: A^f \to A$ inducing a homomorphism of $R$-modules (from ${}^f(A^f)$ to $A$). If the $R$-module structure on $A$ arose by pullback along $f$ from an $S$-module $V$, then $V$ is a direct summand of $A_f$ and of $A^f$, as $R$-module.*

*Proof.* The proof is a simple verification, using (2), (5) and (6):

(i) $\mathrm{Hom}_S(A \otimes_R S, U) \cong \mathrm{Hom}_R(A, \mathrm{Hom}_S(S, U) \cong \mathrm{Hom}_R(A, U)$,
(ii) $\mathrm{Hom}_S(U, \mathrm{Hom}_R(S, A) \cong \mathrm{Hom}_R(U \otimes_S S, A) \cong \mathrm{Hom}_R(U, A)$.

Now $\alpha: A \to A \otimes S$ is just the map $a \mapsto a \otimes 1$, and $\beta: \mathrm{Hom}_R(S, A) \to A$ is the map $\varphi \mapsto 1\varphi$.

If $A = {}^fV$, we put $U = V$ in (7), (8) and consider the image of the identity map on the right of (7), (8): this provides maps $A_f \to V$, $V \to A^f$ which together with $\alpha, \beta$ respectively define a splitting of $A_f$, $A^f$ respectively. ■

The modules $A_f$ and $A^f$ are called the *induced* and *coinduced extensions* of $A$ along $f$. Here the variance refers to $S$, not $A$ (in fact both are covariant in $A$). They are also sometimes called *relatively projective* and *relatively injective*, on account of the following property:

COROLLARY 3.6 *If $A$ is projective as $R$-module, then $A_f$ is projective as $S$-module; if $A$ is injective as $R$-module, then $A^f$ is injective as $S$-module.*

The first part follows because $\mathrm{Hom}_S(A_f, -)$ is exact whenever $\mathrm{Hom}_R(A, -)$ is exact, by (7); similarly the second part follows by (8). ■

An abelian category is said to possess *enough projectives* if every object can be written as a quotient of a projective object. For example, the category $\mathscr{M}_R$ of right modules over any ring $R$ has enough projectives, because every module is a homomorphic image of a free (hence projective) module, by A.2, Th. 4.4.3. Dually, an abelian category is said to have *enough injectives* if every object is a subobject of an injective object. For example, $\mathbf{Z}$ as $\mathbf{Z}$-module is contained in $\mathbf{Q}$, which is injective. Let us show that $\mathscr{M}_R$ has enough injectives.

PROPOSITION 3.7 *Let $R$ be any ring. Then $\mathscr{M}_R$ has enough injectives, i.e. every $R$-module can be embedded in an injective module.*

*Proof.* We first take the special case $R = \mathbf{Z}$. Every abelian group $A$ can be written as a quotient of a free abelian group: $A \cong F/N$. Now $F$ is a direct sum of copies of $\mathbf{Z}$ and by embedding $\mathbf{Z}$ in $\mathbf{Q}$ we can embed the abelian group $F$ in a vector space over $\mathbf{Q}$, $G$ say. Clearly $G$ is divisible as $\mathbf{Z}$-module and hence so is $G/N$, and it contains $F/N \cong A$ as a submodule. But for a $\mathbf{Z}$-module 'divisible' is the same as 'injective' (by A.2, Prop. 4.5.8), so the $\mathbf{Z}$-module $A$ has been embedded in an injective $\mathbf{Z}$-module.

We now take the general case. Given any ring $R$, there is a natural homomorphism $f: \mathbf{Z} \to R$, obtained by mapping $n \mapsto n \cdot 1$, and we can consider any $R$-module $M$ as $\mathbf{Z}$-module by pullback among $f$. By what has been proved, $M$ can be embedded in an injective $\mathbf{Z}$-module $I$, hence the coinduced extension $M^f = \mathrm{Hom}_{\mathbf{Z}}(R, M)$ is a submodule of $I^f$, by the left exactness of Hom. By Prop. 3.5, $M$ is a direct summand of $M^f$, hence it is an $R$-submodule of $I^f$, and $I^f$ is injective by Cor. 3.6. ■

It is possible to go beyond Prop. 3.7 and describe the 'least' injective module containing a given module.

THEOREM 3.8 *Let $R$ be any ring. Given $R$-modules $M$, $E$ where $M \subseteq E$, the following conditions are equivalent:*

(a) *$E$ is a maximal essential extension of $M$,*
(b) *$E$ is a minimal injective module containing $M$.*

*Such an extension $E$ exists for any $R$-module $M$, and if $E'$ is another extension of $M$ satisfying* (a) *and* (b), *then there is an isomorphism from $E$ to $E'$ leaving $M$ elementwise fixed.*

*Proof*. (a)⇒(b). If $E$ is a maximal essential extension of $M$, then any essential extension $F$ of $E$ is an essential extension of $M$, for any non-zero submodule of $F$ meets $E$ and hence $M$ non-trivially. By maximality we have $F = E$, so $E$ has no proper extensions and is therefore injective, by Prop. 3.3; but any submodule of $E$ containing $M$ has $E$ as essential extension and so cannot be injective unless it is the whole of $E$, again by Prop. 3.3. Hence $E$ is a minimal injective module containing $M$.

(b)⇒(a). Assume that $E$ is a minimal injective module containing $M$ and let $F$ be any essential extension of $M$; we claim that $F$ can be embedded in $E$. For the inclusion of $M$ in $E$ extends to a homomorphism $f: F \to E$ because $E$ is injective, and $M \cap \ker f = 0$, hence $\ker f = 0$, because $F$ is an essential extension of $M$. Thus $F$ is embedded in $E$. If $F$ is a maximal essential extension of $M$, then as we have just seen, we can take $F$ to be a submodule of $E$ and by the first part of the proof $F$ is injective. Therefore $F$ is a direct summand of $E$, and so, by the minimality of $E$, we have $E = F$, as we had to show.

We can always construct such an $E$ by taking an injective module $I$ containing $M$ (Prop. 3.7) and inside $I$ taking a maximal essential extension of $M$, using Zorn's lemma. Finally, if $E$, $E'$ are both modules satisfying (a) and (b), then the identity mapping on $M$ extends to a homomorphism $\alpha: E \to E'$ by the injectivity of $E'$. The kernel of $\alpha$ meets $M$ in 0, hence $\ker \alpha = 0$ by (a), so $\operatorname{im} \alpha$ is an injective submodule of $E'$ and $\operatorname{im} \alpha = E'$ by (b). This shows $\alpha$ to be an isomorphism. ■

The module $E$ in Th. 3.8, first constructed by Eckmann and Schopf in 1953, is called the *injective hull* of $M$. Although $E$ is determined up to isomorphism by $M$, this isomorphism is not unique and the correspondence of $E$ to $M$ is not a functor. Later, in Ch. 6, we shall find that over certain rings there is a dual notion of projective cover for every finitely generated module.

## Exercises

(1) Show that the finitely generated right $R$-modules and homomorphisms form a full subcategory of $\mathscr{M}_R$. Is it abelian? What about the full subcategory of cylic $R$-modules?

(2) Prove the rules (4)–(6) in detail.

(3) Let $(A_i)$ be a family of objects in an abelian category. Show that if $\prod A_i$ exists then it is injective iff each $A_i$ is injective; likewise, if $\coprod A_i$ exists, then it is projective iff each $A_i$ is projective. (Warning: exact sequences need not be preserved under products and coproducts in general abelian categories.)

(4) An object $P$ in a category $\mathscr{A}$ is called a *generator* of $\mathscr{A}$ if $h^P = \mathscr{A}(P, -)$ is faithful. Show that a generator in $\mathscr{M}_R$ is faithful as $R$-module, i.e. any non-zero element of $R$ defines a non-zero action.

(5) Show that in an abelian category with arbitrary coproducts an object $P$ is a generator iff every object is a quotient of a copower of $P$. Deduce that an abelian category with arbitrary coproducts and a projective generator has enough projectives.

(6) Dualize Ex. (4) and (5) to show that an abelian category with arbitrary products and an injective *cogenerator* has enough injectives.

(7) Let $f: R \to S$ be a homomorphism. Show that for any modules $A_R$, ${}_SU$, we have $A_f \otimes_S U \cong A \otimes_R {}^fU$.

(8) Show that for $f: R \to S$ as in Ex. (7) and modules $U_R, {}_RV$ we have ${}^f(U_f) \otimes_R V \cong U \otimes_R {}^f(V_f)$. Show further that when $R, S$ are commutative, then $U_f \otimes_S V_f \cong (U \otimes_R V)_f$.

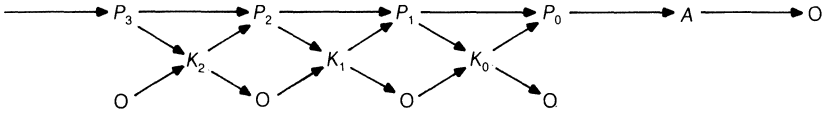
(9) Show that if $M$ is a finitely generated module over a Noetherian ring $R$, then $M^* = \mathrm{Hom}_R(M, R)$ is again finitely generated.

(10) Show that $R$ is a generator of $\mathscr{M}_R$. More generally, show that $M$ is a generator of $\mathscr{M}_R$ iff $R$ is a direct summand of ${}^nM$ for some $n \geqslant 1$.

### 3.4 Homological dimension

For any abelian group $A$ we can write down a presentation $A \cong F/K$, where $F$ is a free abelian group and $K$ as subgroup of $F$ is also free. For modules over a ring we still have such a presentation, but now $K$ need no longer be free, nor even projective. If we take the view that projective modules are particularly simple (justified in the sequel), we may next take a presentation of $K$ and hope that the process terminates, i.e. that $K$ is in some sense closer to being projective than $A$. Our first objective is to assign a numerical value to this lack of projectivity.

Given any $R$-module $A$, we take a projective module $P_0$ mapping onto $A$. The kernel $K_0$ need not be projective, but we can take again a projective $P_1$ mapping onto $K_0$. This map has kernel $K_1$ and we can continue the process, giving rise to a commutative diagram with exact row as follows:




As a rule one omits the kernels and just writes the exact sequence

$$\cdots \to P_3 \to P_2 \to P_1 \to P_0 \to A \to 0. \tag{1}$$

This is called a *projective resolution* of $A$. For example, let $R = k[x, y]$ be the polynomial ring in $x, y$ over a field $k$, and consider $k$ as $R$-module, by pullback

along the natural homomorphism $R \to R/(x, y) \cong k$. We resolve $k$ by mapping $R$ to $R/(x, y)$; the kernel is the ideal $(x, y)$ and we next take the map $R^2 \to (x, y)$ defined by $(a, b) \mapsto ax - by$. Its kernel is the set $(cy, cx)$, $c \in R$, and this is isomorphic to $R$. Thus we have obtained a resolution

$$0 \to R \to R^2 \to R \to k \to 0. \tag{2}$$

As a second example, let $R = \mathbf{Z}/4$, $A = \mathbf{Z}/2$. Clearly $A$ is not projective, but we have homomorphism $R \to A$ consisting of multiplication by 2, with kernel $2\mathbf{Z}/4 \cong A$, hence we obtain an infinite resolution

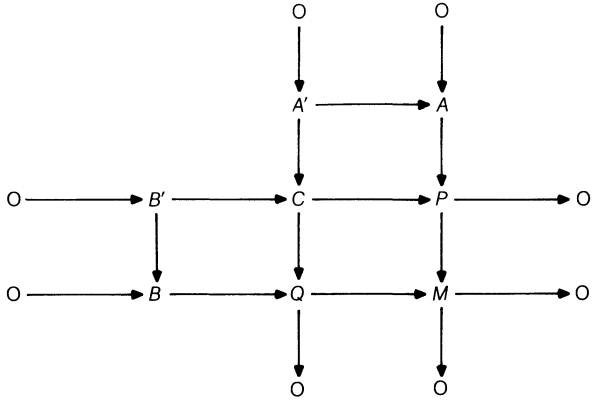
$$\cdots \to R \xrightarrow{2} R \xrightarrow{2} A \to 0. \tag{3}$$

It is clear that a projective resolution, possibly infinite, exists for any module, because $\mathscr{M}_R$ has enough projectives, and the resolution terminates when we reach a projective kernel. In order to compare different resolutions of a given module we need Schanuel's lemma. It is useful to have this in an extended form:

PROPOSITION 4.1 *Let $R$ be any ring and $M$ an $R$-module. Given two short exact sequences $0 \to A \to P \to M \to 0$ and $0 \to B \to Q \to M \to 0$, where $P$ is projective, we have an exact sequence*

$$0 \to A \to P \oplus B \to Q \to 0. \tag{4}$$

*Proof.* If we form the pullback of $P \to M$, $Q \to M$ and use Cor. 1.7, we obtain an exact commutative diagram




Moreover, $A' \cong A$, $B' \cong B$, by Prop. 1.6. Since $P$ is projective, the middle sequence splits and $C \cong P \oplus B$. Now the middle vertical is the desired exact sequence (4). ■

If $Q$ as well as $P$ is projective, (4) splits and we obtain

LEMMA 4.2 (Schanuel's lemma) *Given two short exact sequences as in Prop* 4.1, *if* $P$ *and* $Q$ *are projective, then* $P \oplus B \cong Q \oplus A$. ■

This result suggests the following definition. Two modules $M, N$ are called *projectively equivalent* if there exist projectives $P$, $Q$ such that

$$P \oplus M \cong Q \oplus N. \tag{5}$$

It is clear that this is in fact an equivalence relation; we shall denote the class of $M$ by $[M]$ and note that $[M] = 0$ iff $M$ is projective.

On the set of all equivalence classes we can define an operation as follows. Given $M$, we resolve it by a projective:

$$0 \to A \to P \to M \to 0,$$

and write $\pi(M) = [A]$. By Schanuel's lemma the class $[A]$ depends only on $M$, not on the resolution chosen. If we replace $M$ by $M \oplus Q$, where $Q$ is projective, we have a resolution

$$0 \to A \to P \oplus Q \to M \oplus Q \to 0,$$

and this shows that $\pi(M)$ depends in fact only on the class $[M]$ and not on $M$ itself; $\pi$ is sometimes called the *loop functor*.

We can now define the *homological* (or *projective*) dimension of a module $M$ as

$$\mathrm{hd}(M) = \min\{n | \pi^{n+1}(M) = 0\} \qquad \text{where } \pi^0(M) = [M].$$

This depends only on the class of $M$ and the definition shows that $\mathrm{hd}(M) \leqslant n$ iff $M$ has a projective resolution of length $\leqslant n$ (i.e. of the form (1) with $P_i = 0$ for $i > n$).

The *global dimension* of a ring $R$ is defined as

$$\mathrm{gl\cdot dim\cdot}(R) = \sup\{\mathrm{hd}(M) | \text{all } M_R\}.$$

If necessary we shall distinguish the *right* and *left* global dimensions, formed from right or left modules. In general these numbers may be distinct, although they coincide for Noetherian rings (cf. **3.6** below). The rings of global dimension 0 are just the semisimple rings, for they are the rings for which every module is projective (cf. A.2, Th. 5.3.7). As an example of a ring of infinite global dimension we have the ring $\mathbf{Z}/4$, as the resolution (3) shows.

There is an analogous development using injective resolutions, based on the fact that $\mathscr{M}_R$ also has enough injectives (Prop. 3.7). Given $\mathscr{M}_R$, we form an injective resolution

$$0 \to M \to I_0 \to I_1 \to \cdots, \tag{5}$$

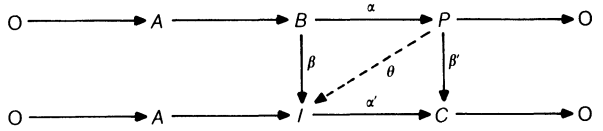
by embedding $M$ in an injective module $I_0$, then embedding the cokernel in $I_1$ and so on. The dual of Schanuel's lemma shows that in a short resolution

$0 \to M \to I \to L \to 0$, the 'injective class' of $L$ (defined like the projective equivalence class) depends only on that of $M$, and so may be written $\iota(M)$. The least integer $n$ such that $\iota^{n+1}(M) = 0$ is called the *cohomological* (or *injective*) *dimension* of $M$, written $\mathrm{cd}(M)$. As before we can define the corresponding global dimension of $R$, but as we shall see in **3.6** below, this agrees with the global dimension as defined in terms of homological dimension. In the special case of global dimension 0 this is already clear from the definition of a semisimple ring as a ring over which any short exact sequence splits.

To illustrate these ideas we shall examine the class of rings of global dimension 1. We shall need a lemma which may be regarded as the dual of Baer's injectivity criterion (Th. 3.4).

LEMMA 4.3 *An R-module P is projective if and only if every homomorphism from P to a quotient of an injective module I can be lifted to I itself.*

*Proof.* The condition is necessary by Th. 2.8. Conversely, assume that it holds. Given a short exact sequence $0 \to A \to B \to P \to 0$, we embed $B$ in an injective module $I$ and form the pushout:




Since $\beta$ is monic, the pushout is a pullback, by the dual of the argument following Prop. 1.6. Now by hypothesis there is a map $\theta: P \to I$ such that $\beta' = \theta\alpha'$, therefore by the pullback property there is a map $\lambda: P \to B$ such that $\lambda\alpha = 1$, $\lambda\beta = \theta$. Thus the given sequence splits and this shows $P$ to be projective. ■

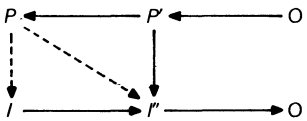
THEOREM 4.4 *For any ring R the following conditions are equivalent:*

(a) *every quotient of an injective right R-module is injective,*

(b) *every submodule of a projective right R-module is projective,*

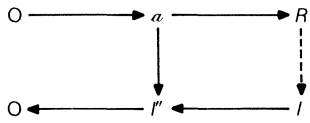
(c) *every right ideal of R is projective.*

*Proof.* (a)⇒(b) Given the diagram shown, where $P$ is projective, we have to fill in the map $P' \to I$ to produce a commutative triangle.




By Lemma 4.3 we may assume that $I$ is injective, and then $I''$ is injective, by (a). We can therefore fill in $P \to I''$, and then $P \to I$ (because $P$ is projective). Now the composition $P' \to P \to I$ provides the required map.

(b)⇒(c) is trivial and the proof of (c)⇒(a) is dual to the first part, using the diagram shown. ■




A ring satisfying the conditions of this theorem is said to be *right hereditary*. For example, any principal ideal domain (commutative or not) is right (and also left) hereditary. From (b) it is clear that the right hereditary rings are just the rings of global dimension at most 1, and (a) shows that the same class is obtained by computing the global dimension from injective resolutions.

In the commutative case hereditary integral domains have another more illuminating description. We recall from A.2, **9.5** that an ideal $\mathfrak{a}$ in a commutative integral domain $R$ is called *invertible* if there is an $R$-submodule $\mathfrak{b}$ of its field of fractions $K$ such that $\mathfrak{ab} = R$. In Prop. 9.5.1 of A.2 we saw that an ideal is invertible iff it is non-zero projective; in particular such an ideal must be finitely generated. In fact a commutative hereditary domain is precisely a Dedekind domain, as the description of the latter in A.2, **9.5** shows.

## Exercises

(1) Given two projective resolutions $(P_i)$, $(P'_i)$ of finite length of a module $M$, show that $P_0 \oplus P'_1 \oplus P_2 \oplus \cdots \cong P'_0 \oplus P_1 \oplus P'_2 \oplus \cdots$.

(2) Find the global dimension of $\mathbf{Z}/n$. (Hint: take first the case of a prime power.)

(3) Let $R$ be a right hereditary Noetherian ring and $M$ a finitely generated submodule of $R^I$, as right $R$-module, for some set $I$. Show that $M$ is a direct sum of modules isomorphic to right ideals of $R$, and hence is projective. (Hint: among the direct summands of $M$ isomorphic to a direct sum of right ideals pick a maximal one.)

(4) Show that if $R$ is as in Ex. (3) and $M$ is a finitely generated $R$-module, then $M = M_0 \oplus \mathfrak{a}_1 \oplus \cdots \oplus \mathfrak{a}_r$, where the $\mathfrak{a}_i$ are right ideals of $R$ and $M_0 = \cap\{\ker f \mid f : M \to R\}$.

(5) Let $R$ be a ring such that every right ideal is free as right $R$-module. Show that every submodule of a free right $R$-module is free (a ring with IBN in which every right ideal is free is called a *right fir*, cf. **10.9**). Show that if $R$ is also commutative, it must be a principal ideal domain.

(6) Show that a left fir (defined by symmetry as in Ex. (5)) which is right Noetherian is also a right fir, in fact right principal (cf. **10.9**).

(7) Show that if $R$ and $S$ are right hereditary and $U$ is any $(R,S)$-bimodule, then the triangular matrix ring $\begin{pmatrix} R & U \\ 0 & S \end{pmatrix}$ is again right hereditary.

(8) Show that $\mathbf{Q}$ is not projective, as $\mathbf{Z}$-module. Deduce that the triangular matrix ring $\begin{pmatrix} \mathbf{Z} & \mathbf{Q} \\ 0 & \mathbf{Q} \end{pmatrix}$ is right but not left hereditary.

(9) Show that over a commutative integral domain $R$, every quotient of a divisible module is divisible. Deduce that $R$ is a Dedekind domain iff every divisible $R$-module is injective.

(10) In the ring $\mathbf{R}[\sin\theta, \cos\theta]$ show that the ideal generated by $\sin\theta$ and $1-\cos\theta$ is projective but not principal, and hence is not a direct summand. (This ring is a Dedekind domain, cf. A.2. **9.5**)

## 3.5 Derived functors

We have seen that the hom functor and tensor product are only left and right exact respectively, and we shall now describe a method of measuring this lack of exactness. The method is quite general and applies to any functor which is left or right exact. The basic idea of the construction is as follows. Let $F$ be a functor which is right exact covariant, say. Given any module $A$, we take a projective resolution

$$\cdots \to X_n \to \cdots \to X_1 \to X_0 \to A \to 0,$$

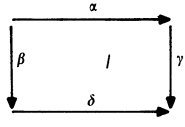
and apply $F$:

$$\cdots \to FX_n \to \cdots \to FX_1 \to FX_0 \to FA \to 0.$$

In general this is no longer exact, but it is still a complex. From any such complex one can form homology groups $H_n(A)$ described below, which measure the lack of exactness of $F$; taking the $X_i$ projective ensures that these groups depend only on $A$ and $F$, but not on the choice of the resolution $X$.

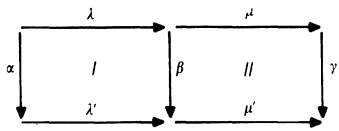
Before we can enter on the actual construction, we need some properties of commutative diagrams. These are true in any abelian category, but we shall only consider the case of modules, where they can be verified by diagram-chasing. As a matter of fact, most of the results then follow for general abelian categories, because every small abelian category has an exact embedding into a module category (cf. Mitchell [1975] and Ex. (12) of Ch. 6), but we shall not make use of this fact.

Given any commutative square $I$:



we define the *image ratio* of $I$ as $i(I) = (\operatorname{im}\gamma \cap \operatorname{im}\delta)/\operatorname{im}(\beta\delta)$ and the *kernel ratio* of $I$ as $k(I) = \ker(\beta\delta)/(\ker\alpha + \ker\beta)$. We note that if $\gamma$ or $\delta$ is monic then $i(I) = 0$, and if $\alpha$ or $\beta$ is epic then $k(I) = 0$. The ratios of two adjacent squares are related by

LEMMA 5.1 (Two-square lemma) *Given a commutative diagram with exact rows:*




*we have* $i(I) \cong k(II)$.

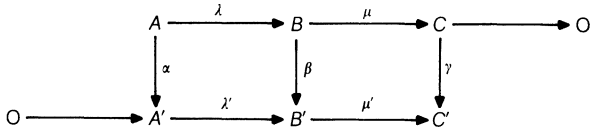
*Proof.* We must show that

$$\frac{\operatorname{im}\lambda' \cap \operatorname{im}\beta}{\operatorname{im}(\lambda\beta)} \cong \frac{\ker(\beta\mu')}{\ker\beta + \ker\mu}.$$

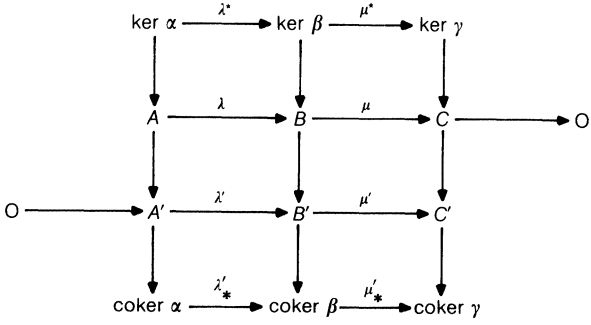
Clearly $\operatorname{im}\beta \cap \operatorname{im}\lambda' = \operatorname{im}\beta \cap \ker\mu' = \{x\beta \mid x\beta\mu' = 0\} = (\ker\beta\mu')\beta$, and $\operatorname{im}\lambda\beta = (\operatorname{im}\lambda)\beta = (\ker\mu)\beta = (\ker\beta + \ker\mu)\beta$. Now both $\ker\beta\mu'$ and $\ker\beta + \ker\mu$ contain $\ker\beta$, hence by the third isomorphism theorem,

$$k(II) \cong \frac{\ker\beta\mu'}{\ker\beta + \ker\mu} \cong \frac{(\ker\beta\mu')\beta}{(\ker\beta + \ker\mu)\beta} \cong i(I). \quad \blacksquare$$

LEMMA 5.2 *Given the commutative diagram with exact rows,*




*the following diagram is commutative with exact rows and columns:*

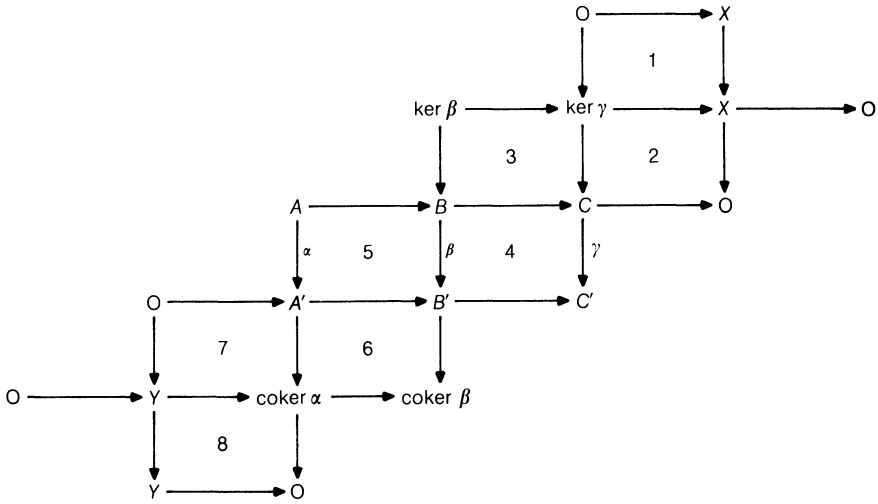


*Here $\lambda^*, \mu^*$ are the maps induced between the kernels and $\lambda'_*, \mu'_*$ are the maps induced between the cokernels. Moreover, if $\lambda$ is monic, then so is $\lambda^*$ and if $\mu'$ is epic, so is $\mu'_*$.*

The proof, by diagram chasing, is straightforward and may be left to the reader. ■

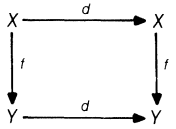
LEMMA 5.3 (Snake lemma) *Given the diagram in the hypothesis of lemma 5.2, there exists a homomorphism $\Delta: \ker\gamma \to \operatorname{coker}\alpha$ such that the sequence* $\ker\alpha \xrightarrow{\lambda^*} \ker\beta \xrightarrow{\mu^*} \ker\gamma \xrightarrow{\Delta} \operatorname{coker}\alpha \xrightarrow{\lambda'_*} \operatorname{coker}\beta \xrightarrow{\mu'_*} \operatorname{coker}\gamma$ *is exact.*

*Proof.* (J. Lambek) We have to prove exactness at $\ker\gamma$ and at $\operatorname{coker}\alpha$, for a suitable $\Delta$, and for this it is enough to show that $\operatorname{coker}\mu^* \cong \ker\lambda'_*$. Writing $X = \operatorname{coker}\mu^*$, $Y = \ker\lambda'_*$, we have the following commutative diagram with exact rows and columns:




By the 2-square lemma, $X \cong i(1) \cong k(2) \cong i(3) \cong k(4) \cong i(5) \cong k(6) \cong i(7) \cong k(8) \cong Y$. ■

We can now return to the task of constructing the homology of a complex. It will be convenient to treat a special case first, that of a differential module. By a *differential module* $X$ we understand an $R$-module $X$ with an endomorphism $d$ of square zero: $d^2 = 0$. If we regard these modules as objects of a category $\mathscr{D}_R$, the maps of the category are taken to be homomorphisms preserving the structure, i.e. the $R$-homomorphisms $f: X \to Y$ such that the square shown commutes;




these maps are traditionally called *chain-maps*. The condition $d^2 = 0$ means that $\operatorname{im} d \subseteq \ker d$. We shall write $\operatorname{im} d = B$, $\ker d = C$ and call the elements of $B$ *boundaries* and those of $C$ *cycles*. Finally $H = C/B$ is called the *homology group* of $X$. (For an excellent concise indication of the geometrical background, see Mac Lane [1963]).

In general $B$ and $C$ and hence $H = H(X)$ are merely abelian groups, but when $R$ is a $K$-algebra, they are $K$-modules. We have

THEOREM 5.4 *For any $K$-algebra $R$, $H: \mathscr{D}_R \to \mathscr{M}_K$ is a covariant linear functor from differential modules to $K$-modules.*

The proof is a straightforward verification, which may be left to the reader. ■

We observe that a complex may be regarded as a special case of a differential module. To see this we need only replace modules by graded modules (regarding $R$ as a graded ring concentrated in degree 0). The type of complex we encounter will generally be a graded module with an antiderivation, also called a *differential*, of degree $r = 1$ or $-1$. In the case $r = -1$ one speaks of a *chain complex*, in the case $r = 1$ of a *cochain complex*. Thus given a chain complex

$$\cdots \to X_n \xrightarrow{d_n} X_{n-1} \xrightarrow{d_{n-1}} \cdots \to X_1 \xrightarrow{d_1} X_0 \to 0, \tag{1}$$

we can regard this as a graded differential module and the homology group $H(X)$ is again a graded module; in detail we have

$$H_n(X) = \ker d_n / \operatorname{im} d_{n+1} \qquad (n = 0, 1, \ldots).$$

Here $d_0$ is taken to be the zero map; thus $H_0(X) = X_0/\operatorname{im} d_1$. It is clear that the complex given by (1) is exact precisely when $H(X) = 0$, so that $H(X)$ may be taken as a measure of the lack of exactness of $X$. We note further if in (1) each $X_n$ for $n \geqslant 1$ is projective, then (1) is a projective resolution of the $R$-module $A$ iff

$$H_n(X) = \begin{cases} A & \text{for } n = 0, \\ 0 & \text{for } n \neq 0. \end{cases} \tag{2}$$

A complex $(X_n)$ satisfying (2) is said to be *acyclic* over $A$.

We shall state the next few results for differential modules rather than complexes ( = graded differential modules), for the sake of simplicity. The change to complexes can easily be made by the reader.

By Th. 5.4 an isomorphism of differential modules induces an isomorphism of homology groups, but since structure is lost in passing to homology, one would expect a much wider class of mappings to induce isomorphisms. The appropriate notion is suggested by the topological background:

DEFINITION Two chain maps $f, g: X \to Y$ of differential modules are said to be *homotopic*, $f \simeq g$, if there is an $R$-homomorphism $s: X \to Y$, called a *homotopy*, such that

$$s \cdot d_Y + d_X \cdot s = f - g. \tag{3}$$

It is clear that this relation between chain maps is an equivalence; moreover it has the desired property:

PROPOSITION 5.5 *Homotopic chain maps induce the same homology map, i.e. if* $f \simeq g$, *then* $H(f) = H(g)$.

*Proof.* By linearity it is enough to show that if $f \simeq 0$, then $H(f) = 0$. If $c \in C(X)$, then $cd = 0$ and since $f = sd + ds$, it follows that $cf = csd + cds = csd \in B(Y)$. Thus $C(X)f \subseteq B(Y)$ and so $H(f) = 0$. ■

It is clear how this definition and proposition have to be modified if $X, Y$ are graded, say both are chain complexes. For $f, g$ to be homotopic they must be of the same degree $r$ say, and then the homotopy $s$ will have degree $r + 1$.

A chain map between differential modules $f: X \to Y$ is said to be a *chain equivalence* or *homotopy equivalence* if there is a second chain map $f': Y \to X$ such that $ff' \simeq 1_X$, $f'f \simeq 1_Y$. This is easily seen to be an equivalence relation between chain maps and now Prop. 5.5 yields

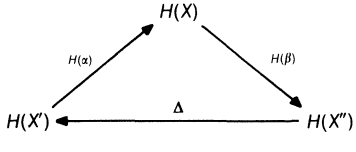
COROLLARY 5.6 *If* $f: X \to Y$ *is a chain equivalence, then* $H(f)$ *is an isomorphism.* ■

We now come to a basic property of short exact sequences of differential modules:

THEOREM 5.7 *Given a short exact sequence of differential modules:*

**X**: $$0 \to X' \xrightarrow{\alpha} X \xrightarrow{\beta} X'' \to 0,$$

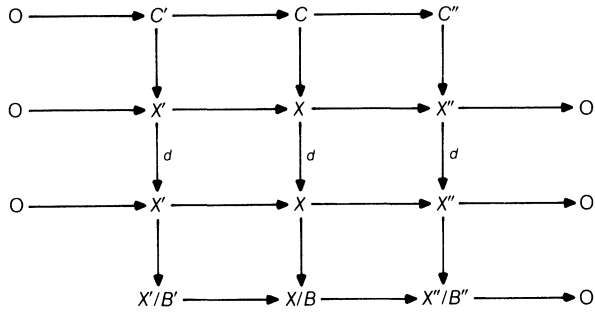
*there exists a homomorphism* $\Delta: H(X'') \to H(X')$ *natural in* **X**, *such that the triangle*



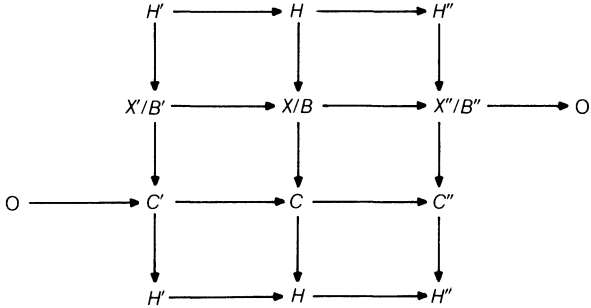
*is exact.*

$\Delta$ is known as the *connecting homomorphism.*

*Proof.* The exact sequence **X** may be written




where $C$ and $X/B$ are the kernel and cokernel of $d$ respectively. By Lemma 5.2 the whole diagram is commutative, with exact rows and columns. Now the map $d: X \to X$ induces a map $X/B \to C$, because $Xd \subseteq C$ and $Bd = 0$, and this map has both kernel and cokernel equal to $H = C/B$. Hence we have a commutative diagram




which has exact rows and columns, by Lemma 5.2. Further, by the snake lemma, there is a homomorphism $\Delta: H'' \to H'$ which makes the homology triangle exact, and which from its derivation is natural in **X**. ■

The important case of this theorem is that where $X', X, X''$ are in fact graded modules, usually chain or cochain complexes, with maps of degree zero between them. In the case of complexes, say, the exact triangle takes on the form of an infinite sequence

$$\cdots \to H_n(X') \to H_n(X) \to H_n(X'') \to H_{n-1}(X') \to \cdots \to H_0(X'') \to 0,$$

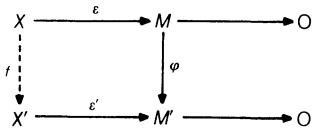
which is called the *exact homology sequence* associated with the short exact sequence **X**.

Any $R$-module may be trivially regarded as a chain complex concentrated in degree 0, i.e. $M_0 = M$, $M_n = 0$ for $n \neq 0$, and $d = 0$. In the sequel all chain complexes will be zero in negative dimension, i.e. $X_n = 0$ for $n < 0$. The complex $X$ is said to be *over* $M$ if there is an exact sequence $X \to M \to 0$ (regarding $M$ as a trivial chain complex in the way described above). In full this sequence reads

$$\cdots \to X_n \to X_{n-1} \to \cdots \to X_1 \to X_0 \to M \to 0. \tag{4}$$

If this is exact, it is called a *resolution*, or also an *acyclic complex*; then $H_n(X) = 0$ for $n > 0$, $H_0(X) \cong M$. If each $X_n$ is projective, we have a *projective resolution*. An important property of projective resolutions is that they are universal among resolutions of $M$. This follows from the more general

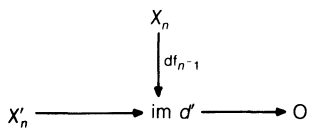
THEOREM 5.8 (Comparison theorem) *Given two complexes*




*where $X$ is projective, $X'$ is a resolution of $M'$ and $\varphi$ is a homomorphism, then there exists a chain map $f: X \to X'$ such that the resulting diagram commutes, and $f$ is unique up to homotopy.*

We shall also say that $f$ is *over* $\varphi$ or that $f$ *lifts* $\varphi$.

*Proof.* We have to construct $f_n: X_n \to X'_n$ such that $f_n d' = d f_{n-1}$ $(n \geqslant 1)$ and $f_0 \varepsilon' = \varepsilon \varphi$. We construct these maps by induction, using the fact that $X_n$ is projective and $\operatorname{im}(d f_{n-1}) \subseteq \operatorname{im} d'$ (by the exactness of $X'$). At the $n$th stage (when $f_{n-1}$ has been constructed) we have the diagram shown




and this can be completed by a map $f_n: X_n \to X'_n$ because $X_n$ is projective; this still applies for $n = 0$, taking $\varepsilon, \varphi$ in place of $d, f_{n-1}$. To show that the map $f$ so constructed is unique, suppose that $g$ is another chain map lifting $\varphi$, then $f - g$ lifts 0. Writing $h = f - g$, we have to find $s_n: X_n \to X'_{n+1}$ such that

$$s_n d' + d s_{n-1} = h_n, \qquad s_0 d' = h_0.$$

The construction needed is quite similar to the one just carried out and may be left to the reader. ■

In particular, if we have two projective resolutions of $M$, we can lift the identity on M to a map of complexes which is unique up to homotopy, hence we obtain

COROLLARY 5.9 *Any two projective resolutions of $M$ are chain equivalent and hence give rise to isomorphic homology groups.* ■

We now have all the means at our disposal for constructing derived functors.

THEOREM 5.10 *Let $F$ be a covariant right exact functor on $\mathscr{M}_R$. Then there exist functors $F_n$ $(n = 0, 1, \ldots)$ such that*

(i) $F_0 M \cong FM$,
(ii) $F_n P = 0$ *for* $n > 0$ *if $P$ is projective,*
(iii) *to each short exact sequence of $R$-modules*

**A**: $$0 \to A' \to A \to A'' \to 0$$

*there corresponds a long exact sequence*

$$\cdots \to F_n A' \to F_n A \to F_n A'' \xrightarrow{\Delta} F_{n-1} A' \to F_{n-1} A \to \cdots \to F_0 A' \to F_0 A \to F_0 A'' \to 0, \tag{5}$$

*where the connecting homomorphism $\Delta$ is natural in* **A**. *Moreover, $F_n$ is determined up to natural isomorphism by* (i)–(iii).

*Proof*. We begin by proving the uniqueness. Given any short exact sequence

$$0 \to K \to P \to A \to 0, \tag{6}$$

where $P$ is projective, we have for any $F_n$ satisfying (i)–(iii), the exact sequence

$$0 \to F_1 A \to FK \to FP \to FA \to 0;$$

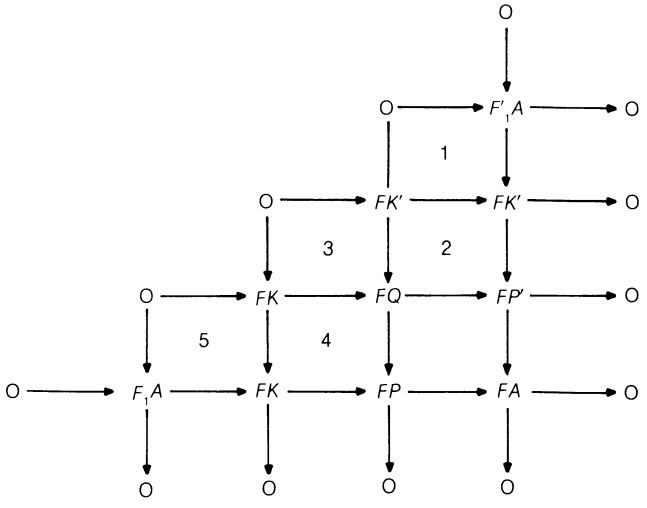
thus any exact sequence (6) with $P$ projective determines $F_1 A$. If $F_1 A$ is obtained from (6) and $F'_1 A$ from

$$0 \to K' \to P' \to A \to 0,$$

where $P'$ is again projective, we construct the pullback $Q$ of $P \to A$ and $P' \to A$ and apply $F$. We thus obtain a commutative diagram as on p. 107.

The row and column meeting in $FQ$ are split exact, because they arose by applying $F$ to split exact rows and columns (split because $P$ and $P'$ are projective). The remaining rows and columns are also exact, and by the 2-square lemma we have $F'_1 A \cong i(1) \cong k(2) \cong i(3) \cong k(4) \cong i(5) \cong F_1 A$. This shows that $F_1 A$ is determined up to isomorphism by (i)–(iii). For $n > 1$ the short exact sequence (6) yields the long exact sequence

$$\cdots \to F_n P \to F_n A \to F_{n-1} K \to F_{n-1} P \to \cdots$$




and $F_nP = F_{n-1}P = 0$ by (ii). Thus in terms of the loop operator $\pi$ introduced in **3.4** we have the formula

$$F_nA \cong F_{n-1}(\pi(A)); \tag{7}$$

this makes sense since $F_n$ is constant on projective equivalence classes for $n > 0$, by (ii). Now the uniqueness follows by induction on $n$.

It remains to prove the existence of $F_n$; here we may use any projective resolution for $A$, say $X \to A \to 0$. Applying $F$, we get a complex $FX \to FA \to 0$. In detail this reads

$$\cdots \to FX_n \to FX_{n-1} \to \cdots \to FX_1 \to FX_0 \to FA \to 0.$$

We put $F_nA = H_n(FX)$ and assert that this satisfies (i)–(iii).

(i) We have $X_1 \to X_0 \to A \to 0$, and by the right exactness of $F$ obtain the exact sequence

$$FX_1 \to FX_0 \to FA \to 0;$$

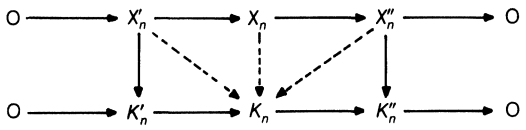
hence $FA \cong FX_0/\operatorname{im} FX_1 \cong H_0(FX)$.

(ii) If $A$ is projective, we can take as our resolution

$$0 \to A \to A \to 0;$$

applying $F$, we find $0 \to FA \to FA \to 0$, hence $F_nA = 0$ for $n > 0$.

(iii) Given a short exact sequence **A** as in (iii), we take projective resolutions $X'$, $X''$ of $A'$ and $A''$ and construct a projective resolution $X$ of $A$ by induction on $n$. If the kernels at the $(n-1)$st stage are $K'_n$, $K_n$, $K''_n$, we have




where $X_n = X'_n \oplus X''_n$. Since $X''_n$ is projective, we have a map $X''_n \to K_n$, while the map $X'_n \to K_n$ arises by composition. By definition of $X_n$ as a direct sum (i.e. product) we obtain a map $X_n \to K_n$ to make the squares commute, and a simple diagram chase shows that this map is epic (this also follows from the 5-lemma). By induction on $n$ we thus have a resolution $X$ of $A$. Now the row

$$0 \to X' \to X \to X'' \to 0$$

is split exact, by definition; applying $F$ we obtain the exact sequence of complexes

$$0 \to FX' \to FX \to FX'' \to 0,$$

and Th. 5.7 provides us with the exact homology sequence, which is the required exact sequence. ■

The functors $F_n$ constructed in Th. 5.10 are called the (*left*) *derived functors* of $F$. The same result, appropriately modified, gives a construction of right derived functors of a left exact covariant functor, using injective resolutions. Any given module $A$ can be embedded in an injective module $I_0$ by Prop. 3.7; by embedding the cokernel in an injective module $I_1$ and continuing in this fashion, we obtain an injective resolution

$$0 \to A \to I_0 \to I_1 \to \cdots. \tag{8}$$

If $F$ is left exact covariant, we obtain a series of functors $F^n$ such that

(i) $F^0 A \cong FA$,
(ii) $F^n I = 0$ for $n > 0$ when $I$ is injective,
(iii) for each short exact sequence **A** as in Th. 5.10 there is a corresponding long exact sequence

$$0 \to F^0 A' \to F^0 A \to F^0 A'' \to \cdots \to F^{n-1} A'' \xrightarrow{\Delta} F^n A' \to F^n A \to F^n A'' \tag{9}$$

where the connecting homomorphism $\Delta$ is natural in **A**, and $F^n$ is determined up to isomorphism by (i)–(iii).

The proof is precisely analogous to that of Th. 5.10. We note that here the index appears as a superscript, and its value increases along the sequence, wheras in (5) the index is a subscript, which decreases as we go along the sequence.

For a contravariant functor $F$ the sides are reversed; if $F$ is left exact, we define $F^n$ by means of a projective resolution and obtain the long exact sequence

(9), while for a right exact contravariant functor $F$ we define $F_n$ by an injective resolution and obtain the long exact sequence (5). To sum up, a projective resolution is needed for a right exact covariant or left exact contravariant functor, and an injective resolution for a left exact covariant or right exact contravariant functor.

Of course the construction of Th. 5.10 can be carried out for any functor, not necessarily right exact. In general we obtain in this way the *left derived* functor of $F$; similarly we can form the right derived functor (using an injective resolution) and together they form a long exact sequence extending in both directions (cf. Ex. (9)).

## Exercises

(1) Show that if $I$ is a pullback or pushout square, then $i(I) = 0$ and $k(I) = 0$.

(2) Show that the category of chain complexes is an abelian category.

(3) Let $M$ be a finitely presented $R$-module, i.e. with a resolution

$$0 \to G \to F \to M \to 0,$$

where $F$ is free and both $F$ and $G$ are finitely generated. Given any short exact sequence

$$0 \to A \to B \to M \to 0,$$

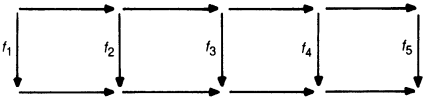
where $B$ is finitely generated, show that $A$ is finitely generated. (Hint: use Prop. 4.1.)

(4) Verify that homotopy is an equivalence between chain maps.

(5) Show that if $f \simeq g$ is a homotopy of chain maps, $f, g: X \to Y$ and $h: Y \to Z$, then $fh \simeq gh$; likewise $ef \simeq eg$ for $e: U \to X$.

(6) Prove Th. 5.4.

(7) (5-lemma) Given a commutative diagram with exact rows




if $f_1, f_2, f_4, f_5$ are isomorphisms, then so is $f_3$. More precisely, if $f_1$ is epic and $f_2, f_4$ are monic, then $f_3$ is monic, and dually.

(8) Verify that the isomorphisms $F_n \to F'_n$ between two functors satisfying (i)–(iii) of Th. 5.10 are compatible with $\Delta$.

(9) For any covariant functor $F$, $F_n M$ $(n \geqslant 0)$ is defined as in the proof of Th. 5.10 in

terms of a projective resolution of $M$. Show that there is a natural transformation $F_0M \to FM$ and that $F_0$ is right exact; if now $F^nM$ is defined similarly in terms of an injective resolution, there is a natural transformation $FM \to F^0M$ and $F^0$ is left exact. Hence obtain an exact sequence of derived functors ($F_n$, $F^n$ are called the *left* and *right derived* functors of $F$, respectively).

(10) Let $R$ be a ring with IBN and $M$ an $R$-module with a finite resolution $(F_i)$ by free modules of finite rank. Show that the integer $\chi(M) = \sum(-1)^i rk F_i$ depends only on $M$ and not on the resolution; it is called the *Euler characteristic*. Given a short exact sequence **A** of modules with a finite resolution by free modules of finite rank, show that $\chi(A) = \chi(A') + \chi(A'')$. Define $\chi$ for a complex $C$ and show that $\chi(H(C)) = \chi(C)$.

## 3.6 Ext and Tor, and the global dimension of a ring

The most important functors to which the construction of **3.5** has been applied are $A \otimes B$ and $\mathrm{Hom}(A, B)$. These are *bifunctors*, i.e. functors of two arguments, and it is possible to form derived functors in two ways, by resolving either the first or the second argument. We shall find that the results in the two cases are the same; this is based on a general criterion which we shall now derive.

Let $F(A, B)$ be any bifunctor, covariant right exact in each argument, say. Then $F$ is said to be *balanced* if $F(A, -)$ is an exact functor whenever $A$ is projective, and $F(-, B)$ is exact whenever $B$ is projective. The same definition applies if $F$ is contravariant left exact in either argument, while for a covariant left exact or a contravariant right exact argument, 'projective' is replaced by 'injective'.

THEOREM 6.1 *Let $F$ be a bifunctor, covariant right exact in each argument, and denote by $F'_n$, $F''_n$ the derived functors obtained by resolving the first and second argument of $F$ respectively. If $F$ is balanced, then*

$$F'_n(A, B) \cong F''_n(A, B), \tag{1}$$

*by an isomorphism which is natural in $A$ and $B$.*

*Proof.* By the uniqueness of derived functors (Th. 5.10) we need only take $F'_n(A, B)$ for fixed $A$ and verify that it satisfies the conditions of Th. 5.10 as a functor in $B$.

(i) $F'_0(A, B) \cong F(A, B)$ by definition.

(ii) If $B$ is projective, then $F(-, B)$ is exact. We apply this functor to a projective resolution of $A$:

$$X \to A \to 0, \tag{2}$$

and obtain

$$F(X, B) \to F(A, B) \to 0.$$

This is still a resolution, by the exactness of $F(-, B)$. Hence $F'_n(A, B) = 0$ for $n > 0$, by the definition of $F'_n(-, B)$.

(iii) Given a short exact sequence

$$0 \to B' \to B \to B'' \to 0, \tag{3}$$

and a projective resolution (2) of $A$, we apply $F(X, -)$ to (3). Since $F(X, -)$ is exact, we obtain an exact sequence of complexes:

$$0 \to F(X, B') \to F(X, B) \to F(X, B'') \to 0.$$

From this the long exact sequence is obtained by applying Th. 5.7. ■

A similar argument applies when there is a change of side or variance.

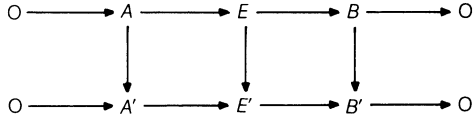
Let us apply these results to $\mathrm{Hom}(A, B)$. This is left exact covariant in $B$ and left exact contravariant in $A$; moreover it is balanced. By Th. 6.1 the derived functor may be obtained either by a projective resolution of $A$ or by an injective resolution of $B$. It is written

$$\mathrm{Ext}^n_R(A, B) \qquad \text{or} \qquad \mathrm{Ext}^n(A, B).$$

To account for the name we shall briefly indicate an interpretation of $\mathrm{Ext}^1$. Given two $R$-modules $A$, $B$ over a ring $R$, an *extension* of $A$ by $B$ is a module $E$ together with a short exact sequence

$$0 \to A \to E \to B \to 0. \tag{4}$$

We can form the category Ex, whose objects are short exact sequences, with the obvious maps between them: a morphism is a triple of homomorphisms making the diagram




commutative. We observe that if the maps $A \to A'$ and $B \to B'$ are isomorphisms, then so is $E \to E'$, by the 5-lemma, and we then have an isomorphism of extensions.

From the extension (4) we can form the exact homology sequence

$$0 \to \mathrm{Hom}(B, A) \to \mathrm{Hom}(B, E) \to \mathrm{Hom}(B, B) \xrightarrow{\Delta} \mathrm{Ext}^1(B, A) \to \cdots$$

Consider the image in $\mathrm{Ext}^1(B, A)$ of the identity map $j$ on $B$: $j\Delta$. This is called the *obstruction* or the *characteristic class* of the extension. Clearly it depends only on the isomorphism type of the extension (4). Moreover, it is zero iff (4) splits; for $j\Delta = 0$ iff $j$ is induced by a homomorphism $B \to E$, and that is just the condition for (4) to split, by Cor. 1.5.

We could also apply $\mathrm{Hom}(-, A)$ to the sequence (4) and get

$$0 \to \mathrm{Hom}(B, A) \to \mathrm{Hom}(E, A) \to \mathrm{Hom}(A, A) \xrightarrow{\Delta} \mathrm{Ext}^1(B, A) \to \cdots$$

This will give the same obstruction; in fact we have a bijection from the set of isomorphism classes of extensions of $A$ by $B$ to $\mathrm{Ext}^1(B, A)$. We shall return to this topic in **4.1**.

The homological dimension of a module was defined in **3.4** in terms of the loop functor $\pi$; we now show how to express it in terms of Ext. For simplicity we shall not distinguish between the class $[\pi A]$ and a module in it.

PROPOSITION 6.2 *For any R-module A over any ring R the following conditions are equivalent*:

(a) $\mathrm{hd}\, A \leqslant n$,

(b) $\mathrm{Ext}^k(A, -) = 0$ *for all* $k > n$,

(c) $\mathrm{Ext}^{n+1}(A, -) = 0$.

*Proof*. For any $k > 0$ and any $R$-module $B$ we have, by (7) of **3.5** and its dual,

$$\mathrm{Ext}^k(\pi A, B) = \mathrm{Ext}^{k+1}(A, B) = \mathrm{Ext}^k(A, \iota B) \qquad \text{for } k > 0. \tag{5}$$

Now (a) states that $\pi^n A$ is projective; so in that case we have for any $k > n$,

$$\mathrm{Ext}^k(A, -) = \mathrm{Ext}^{k-1}(\pi A, -) = \cdots = \mathrm{Ext}^{k-n}(\pi^n A, -) = 0,$$

and (b) follows. Clearly (b)$\Rightarrow$(c), so assume (c). Then

$$\mathrm{Ext}^1(\pi^n A, -) = \mathrm{Ext}^2(\pi^{n-1} A, -) = \cdots = \mathrm{Ext}^{n+1}(A, -) = 0;$$

this means that $\mathrm{Hom}(\pi^n A, -)$ is exact, i.e. $\pi^n A$ is projective, so (a) holds. ■

In the dual situation we can assert a little more:

PROPOSITION 6.3 *For any R-module B over any ring R the following conditions are equivalent*:

(a) $\mathrm{cd}\, B \leqslant n$,

(b) $\mathrm{Ext}^k(-, B) = 0$ *for all* $k > n$,

(c) $\mathrm{Ext}^{n+1}(-, B) = 0$,

(d) $\mathrm{Ext}^{n+1}(C, B) = 0$ *for all cyclic modules C*.

The proof that (a)–(c) are equivalent is entirely analogous to that of Prop. 6.2 and it is clear that (c) implies (d). Conversely, assume (d) for right modules say; then for any right ideal $\mathfrak{a}$ of $R$,

$$\mathrm{Ext}^1(R/\mathfrak{a}, \iota^n B) = \mathrm{Ext}^{n+1}(R/\mathfrak{a}, B) = 0,$$

hence the short exact sequence

$$0 \to \mathfrak{a} \to R \to R/\mathfrak{a} \to 0$$

leads to the exact sequence

$$0 \to \mathrm{Hom}(R/\mathfrak{a}, I^n B) \to \mathrm{Hom}(R, I^n B) \to \mathrm{Hom}(\mathfrak{a}, I^n B) \to 0.$$

This shows that any homomorphism $\mathfrak{a} \to I^n B$ is obtained by restriction from a homomorphism $R \to I^n B$. By Baer's criterion (Th. 3.4) it follows that $I^n B$ is injective, i.e. (a). ■

From Prop. 6.2 we see that the global homological dimension of $R$ may be defined as $\sup\{n | \mathrm{Ext}^n \neq 0\}$, while Prop. 6.3 shows that this determines the global cohomological dimension. Hence we have

COROLLARY 6.4 *For any ring $R$, the global homological and the global cohomological dimensions are equal, and may be defined as*

$$\sup\{n | \mathrm{Ext}^n(C, B) \neq 0 \text{ for cyclic } C \text{ and any } B\}$$

*i.e.* $\sup(\mathrm{hd}\, C)$ *for cylic modules $C$.* ■

Of course it must be borne in mind that the global dimension defined here refers to right $R$-modules, and in general it will be necessary to distinguish between right and left global dimensions. As we shall soon see, for Noetherian rings these numbers coincide, but we shall also meet more general examples where they differ (cf. Ex. 4.8).

We now turn to the tensor product. The functor $A \otimes_R B$ is covariant right exact in each argument, and is balanced. We therefore have a unique derived functor, sometimes called *torsion product*, written

$$\mathrm{Tor}_n^R(A, B) \qquad \text{or} \qquad \mathrm{Tor}_n(A, B).$$

As an example consider the case $R = \mathbf{Z}$. Here $\mathrm{Tor}_n = 0$ for $n > 1$, because $\mathbf{Z}$ is hereditary and so all projective resolutions have length at most 1. Write $\mathbf{C}_k$ for the cyclic group of order $k$; we have an exact sequence

$$0 \to \mathbf{Z} \xrightarrow{k} \mathbf{Z} \to \mathbf{C}_k \to 0,$$

where $k$ indicates multiplication by $k$. Tensoring up with $\mathbf{C}_k$ we get

$$0 \to \mathrm{Tor}_1(\mathbf{C}_k, \mathbf{C}_k) \to \mathbf{C}_k \otimes \mathbf{Z} \to \mathbf{C}_k \otimes \mathbf{Z} \to \mathbf{C}_k \otimes \mathbf{C}_k \to 0.$$

If $c$ is a generator of $\mathbf{C}_k$, then under the induced map $1 \otimes k$ we have $c \otimes 1 \mapsto c \otimes k = ck \otimes 1 = 0$. Therefore $\ker(1 \otimes k) = \mathbf{C}_k$ and we find

$$\mathrm{Tor}_1^{\mathbf{Z}}(\mathbf{C}_k, \mathbf{C}_k) \cong \mathbf{C}_k. \tag{6}$$

Since Tor, like $\otimes$, preserves direct sums, (6) together with the equation $\mathrm{Tor}_1(\mathbf{C}_h, \mathbf{C}_k) = 0$ for coprime $h, k$, is enough to determine Tor for any finitely generated abelian group.

A right $R$-module $A$ is said to be *flat* if $A \otimes -$ is an exact functor, or equivalently, if $\mathrm{Tor}_1^R(A, -) = 0$. For example, any projective module is flat. A corresponding definition applies to left $R$-modules.

Using Tor, we can define another dimension for modules, the *weak dimension*. This is defined, for a right $R$-module $A$, as

$$\mathrm{wd}\, A = \sup\{n | \mathrm{Tor}_n^R(A, -) \neq 0\}.$$

In terms of the loop functor $\pi$ we can also define $\mathrm{wd}\, A$ as the least integer $k$ such that $\pi^k A$ is flat. Now the *weak global dimension* of $R$ is defined as

$$\mathrm{w{\cdot}gl{\cdot}dim}(R) = \sup\{n | \mathrm{Tor}_n^R \neq 0\}.$$

From the definition (and the symmetry of Tor) it is clear that this function is left–right symmetric. To compare it to the global dimension, let us consider the situation $({}_A A_R, {}_Z B_R, {}_Z C)$ and define a natural transformation

$$A \otimes_R \mathrm{Hom}_Z(B, C) \to \mathrm{Hom}_Z(\mathrm{Hom}_R(A, B), C) \tag{7}$$

by mapping the element $a \otimes f$, where $a \in A$, $f \in \mathrm{Hom}_Z(B, C)$, to

$$\varphi(a, f): \theta \mapsto (\theta a) g, \qquad \text{where } \theta \in \mathrm{Hom}_R(A, B).$$

This map is clearly $R$-balanced and biadditive, and hence defines a homomorphism (7), which is easily seen to be natural. Moreover, for $A = R$ it is an isomorphism, hence it is an isomorphism for any finitely generated projective $R$-module $A$. We now fix $C$ to be $\mathbf{Z}$-injective, i.e. divisible and consider the two sides of (7) as a functor in $A$. Both sides are covariant right exact; if we apply them to a projective resolution (2) of $A$ we obtain a natural transformation

$$\mathrm{Tor}_n^R(A, \mathrm{Hom}_Z(B, C)) \to \mathrm{Hom}_Z(\mathrm{Ext}_R^n(A, B), C). \tag{8}$$

Suppose now that $R$ is right Noetherian and $A$ is finitely generated. Then each term $X_n$ of the resolution (2) may be taken to be finitely generated and so (8) will be an isomorphism. Thus we have proved

PROPOSITION 6.5 *Let $R$ be a right Noetherian ring and $C$ a divisible abelian group. Then for any right $R$-modules $A, B$ such that $A$ is finitely generated, we have*

$$\mathrm{Tor}_n^R(A, \mathrm{Hom}_Z(B, C)) \cong \mathrm{Hom}_Z(\mathrm{Ext}_R^n(A, B), C) \qquad \textit{for all } n \geqslant 0, \tag{9}$$

*by a natural isomorphism.* ■

We shall use this result to compare weak and homological dimensions. In the first place, any $R$-module $A$ satisfies

$$\mathrm{hd}\, A \geqslant \mathrm{wd}\, A, \tag{10}$$

because every projective module is flat. It follows that for any ring $R$,

$$\mathrm{w\cdot gl\cdot dim}(R) \leqslant \mathrm{r\cdot gl\cdot dim}(R), \mathrm{l\cdot gl\cdot dim}(R). \tag{11}$$

Now assume that $R$ is right Noetherian and $A$ is finitely generated over $R$. Choose $n$ such that $n \leqslant \mathrm{hd}\, A$; then $\mathrm{Ext}^n_R(A, B) \neq 0$ for some $R$-module $B$, and moreover we can find a divisible group $C$ into which $\mathrm{Ext}^n_R(A, B)$ has a non-zero homomorphism (indeed an embedding, by Prop. 3.7). Thus for suitable $B, C$ the right-hand side of (9) is non-zero; looking at the left-hand side, we deduce that $\mathrm{wd}\, A \geqslant n$. Hence equality must hold in (10) and we have proved

THEOREM 6.6 *If $R$ is a right Noetherian ring, then for any finitely generated right $R$-module $A$,* $\mathrm{wd}\, A = \mathrm{hd}\, A$. ■

COROLLARY 6.7 *For any right Noetherian ring $R$,*

$$\mathrm{r\cdot gl\cdot dim}(R) = \mathrm{w\cdot gl\cdot dim}(R) \leqslant \mathrm{l\cdot gl\cdot dim}(R). \tag{12}$$

For by Cor. 6.4 the right global dimension is the supremum of the projectve dimensions of the cyclic right $R$-modules, and by Th. 6.6 this cannot exceed the weak global dimension of $R$; by (10) it cannot be less, and so the equality in (11) follows. The inequality follows similarly from (10), bearing in mind that the weak global dimension is left–right symmetric. ■

If $R$ is left and right Noetherian, we obtain by symmetry,

COROLLARY 6.8 *In a Noetherian ring $R$,*

$$\mathrm{r\cdot gl\cdot dim}(R) = \mathrm{l\cdot gl\cdot dim}(R). \quad ■$$

For our last result in this section we shall need a change-of-rings theorem which is also generally useful. Let $f: R \to S$ be a homomorphism of rings; we saw in **3.3** that every $S$-module $U$ can be considered as $R$-module ${}^fU$ by pullback along $f$; in particular $S$ itself becomes an $R$-bimodule in this way. Further, every $R$-module $A$ gives rise to an induced extension $A_f = A \otimes_R S$ and a coinduced extension $A^f = \mathrm{Hom}_R(S, A)$.

THEOREM 6.9 *Let $R, S$ be any rings and $f: R \to S$ a homomorphism. If $S$ is projective as right $R$-module, then there is a natural isomorphism*

$$\mathrm{Ext}^n_S(U, A^f) \cong \mathrm{Ext}^n_R({}^fU, A) \qquad (U_S, A_R). \tag{13}$$

*If $S$ is flat as left $R$-module, then there is natural isomorphism*

$$\mathrm{Ext}^n_S(A_f, U) \cong \mathrm{Ext}^n_R(A, {}^fU) \qquad (U_S, A_R), \tag{14}$$

*and*

$$\mathrm{Tor}^S_n(A_f, U) \cong \mathrm{Tor}^R_n(A, {}^fU) \qquad (A_R, {}_SU). \tag{15}$$

*Proof.* Let us take an injective resolution $0 \to A \to I$ and apply the functor $\mathrm{Hom}_R(S, -)$, which is exact because $S_R$ is projective. We obtain an exact sequence

$$0 \to A^f \to I^f,$$

and here the terms $I_n^f$ are injective as coinduced modules, so this is an injective resolution of $A^f$. If we now apply the hom functor to this resolution and bear in mind that by (3) and (4) of **3.3**,

$$\mathrm{Hom}_S(U, I_n^f) \cong \mathrm{Hom}_R(U, I_n),$$

we obtain (13). Similarly, if ${}_RS$ is flat, then $- \otimes_R S$ is exact. Hence if $P \to A \to 0$ is a projective resolution, then so is $P_f \to A_f \to 0$. We now apply the hom functor and use the fact that

$$\mathrm{Hom}_S((P_n)_f, U) \cong \mathrm{Hom}_R(P_n, {}^fI)$$

to obtain (14); similarly we have

$$(P_n)_f \otimes_S U \cong P_n \otimes_R {}^fU,$$

and applying the tensor product functor we obtain (15). ■

We conclude this section by finding an estimate for the global dimension of a tensor ring:

THEOREM 6.10 (Yu. V. Roganov (1975)) *Let $E$ be a $K$-algebra with $\mathrm{r\cdot gl\cdot dim}(E) = n$, $U$ an $E$-bimodule and $R = T_E(U)$ the tensor $E$-ring on $U$ with canonical map $f: E \to R$.*

*(i) If ${}_EU$ is flat, then*

$$n \leqslant \mathrm{r\cdot gl\cdot dim}(R) \leqslant n + 1, \tag{16}$$

*and $\mathrm{r\cdot gl\cdot dim}(R) = n + 1$ if and only if $\mathrm{hd}(A \otimes_E U) = n$ for some right $E$-module $A$.*

*(ii) If $U_E$ is projective, then (16) holds, and $\mathrm{r\cdot gl\cdot dim}(R) = n + 1$ if and only if $\mathrm{cd}(\mathrm{Hom}_E(U, B)) = n$ for some right $E$-module $B$.*

*(iii) If ${}_EU$ is flat and $\mathrm{w\cdot gl\cdot dim}(E) = m$, then*

$$m \leqslant w\cdot \mathrm{gl\cdot dim}(R) \leqslant m + 1, \tag{17}$$

*and $\mathrm{w\cdot gl\cdot dim}(R) = m + 1$ if and only if $\mathrm{wd}(A \otimes_E U) = m$ for some $A_E$.*

*Proof*.* Throughout this proof all tensor products are understood to be over $E$, unless otherwise stated.

Let $\Omega = R \otimes U \otimes R$ be the universal derivation module for $R$, and consider

*I am indebted to W. Dicks for this proof.

the sequence (14) of **2.2**. Since it is split exact, it remains so on tensoring (over $R$) with any right $R$-module $M$. Recalling that $M$ qua $E$-module is just ${}^fM$, we thus obtain

$$0 \to {}^fM \otimes U \otimes R \to {}^fM \otimes R \to M \to 0,$$

which can also be written

$$0 \to ({}^fM \otimes U)_f \to ({}^fM)_f \to M \to 0.$$

Hence we obtain the exact cohomology sequence

$$\cdots \to \mathrm{Ext}_R^\nu(M,N) \to \mathrm{Ext}_R^\nu(({}^fM)_f, N) \to \mathrm{Ext}_R^\nu(({}^fM \otimes U)_f, N) \to \mathrm{Ext}_R^{\nu+1}(M,N) \to \cdots$$

for any right $R$-module $N$. Since ${}_EU$ is flat, so is ${}_ER$ and so by Th. 6.9 this simplifies to

$$\cdots \to \mathrm{Ext}_R^\nu(M,N) \to \mathrm{Ext}_E^\nu({}^fM, {}^fN) \to \mathrm{Ext}_E^\nu({}^fM \otimes U, {}^fN) \to \mathrm{Ext}_R^{\nu+1}(M,N) \to \cdots \tag{18}$$

It follows that $\mathrm{r\cdot gl\cdot dim}(R) \leqslant n+1$. Moreover, by the definition of $n$, $\mathrm{Ext}_E^{n+1}({}^fM, {}^fN) = 0$, so we have a surjection

$$\mathrm{Ext}_E^n({}^fM \otimes U, {}^fN) \to \mathrm{Ext}_R^{n+1}(M,N) \to 0. \tag{19}$$

We next show that $\mathrm{r\cdot gl\cdot dim}(R) \geqslant n$. Choose right $E$-modules $A, B$ such that $\mathrm{Ext}_E^n(A,B) \neq 0$ and consider $A, B$ as right $R$-modules with trivial $U$-action, i.e. $AU = BU = 0$. The $E$-module structure is then recovered by pullback along $f$. Taking $M = A$, $N = B$ in (18) and observing that $\mathrm{Hom}_R(A,B) \to \mathrm{Hom}_E(A,B)$ is then an isomorphism, we conclude by exactness that $\mathrm{Hom}_E(A,B) \to \mathrm{Hom}_E(A \otimes U, B)$ is then the zero map. The same applies if we resolve $B$, hence by (18) we have the exact sequence

$$0 \to \mathrm{Ext}_E^{n-1}(A \otimes U, B) \to \mathrm{Ext}_R^n(A,B) \to \mathrm{Ext}_E^n(A,B) \to 0. \tag{20}$$

It follows that $\mathrm{Ext}_R^n(A,B) \neq 0$, and so $\mathrm{r\cdot gl\cdot dim}(R) \geqslant n$; this proves (16). Now if $\mathrm{hd}(A \otimes U) = n$ for some $A_E$, then by (20) with $n$ replaced by $n+1$, we have $\mathrm{r\cdot gl\cdot dim}(R) = n+1$, while if $\mathrm{hd}(A \otimes U) < n$ for all $A_E$, then $\mathrm{hd}({}^fM \otimes U) < n$ and by (19), $\mathrm{r\cdot gl\cdot dim}(R) \leqslant n$.

The proof of (ii) is similar. For any right $R$-module $M$ we have

$$\begin{aligned} \mathrm{Hom}_R(R \otimes U \otimes R, M) &\cong \mathrm{Hom}_E(R, \mathrm{Hom}_R(U \otimes R, M)), \\ &\cong \mathrm{Hom}_E(R, \mathrm{Hom}_E(U, {}^fM)), \\ &\cong \mathrm{Hom}_E(U, {}^fM)^f. \end{aligned}$$

In particular, $\mathrm{Hom}_R(R \otimes R, M) = ({}^fM)^f$; hence if we apply $\mathrm{Hom}_R(-, M)$ to (14) of **2.2** we obtain an exact sequence

$$0 \to \mathrm{Hom}_E(U, {}^fM)^f \to ({}^fM)^f \to M \to 0,$$

and using Th. 6.9 we thus find the exact sequence

$$\cdots \to \mathrm{Ext}_R^{\nu}(M,N) \to \mathrm{Ext}_E^{\nu}(M,N) \to \mathrm{Ext}_E^{\nu}(\mathrm{Hom}_E(U,{}^fM),{}^fN) \to \mathrm{Ext}_R^{\nu+1}(M,N)$$

We see again that $\mathrm{r\cdot gl\cdot dim}(R) \leqslant n+1$, and the surjection (19) is replaced by

$$\mathrm{Ext}_E^n(\mathrm{Hom}_E(U,M),N) \to \mathrm{Ext}_R^{n+1}(M,N) \to 0.$$

The argument as before gives the analogue of (20):

$$0 \to \mathrm{Ext}_E^{n-1}(\mathrm{Hom}_E(U,A)B) \to \mathrm{Ext}_R^n(A,B) \to \mathrm{Ext}_E^n(A,B) \to 0,$$

and the rest follows as before.

To prove (iii) we apply $\otimes N$ and obtain the exact homology sequence

$$\cdots \to \mathrm{Tor}_{\nu+1}^R(M,N) \to \mathrm{Tor}_{\nu}^E({}^fM \otimes U,{}^fN) \to \mathrm{Tor}_{\nu}^E({}^fM,{}^fN) \to \mathrm{Tor}_{\nu}^R(M,N) \to \cdots .$$

Hence $\mathrm{w\cdot gl\cdot dim}(R) \leqslant m+1$. We obtain an exact sequence

$$0 \to \mathrm{Tor}_{m+1}^R(M,N) \to \mathrm{Tor}_m^E({}^fM \otimes U,{}^fN), \tag{21}$$

and instead of (20) we have

$$0 \to \mathrm{Tor}_m^E(A,B) \to \mathrm{Tor}_m^R(A,B) \to \mathrm{Tor}_{m-1}^E(A \otimes U,B) \to 0. \tag{22}$$

It follows that $\mathrm{Tor}_m^R(A,B) \neq 0$, so $\mathrm{w\cdot gl\cdot dim}(R) \geqslant m$. If $\mathrm{wd}(A \otimes U) = m$ for some $A_E$, then replacing $m$ by $m+1$ in (21) we find $\mathrm{w\cdot gl\cdot dim}(R) = m+1$; if $\mathrm{wd}(A \otimes U) < m$ for all $A_E$, then $\mathrm{wd}({}^fM \otimes U) < m$ and by (22) we have $\mathrm{w\cdot gl\cdot dim}(R) \leqslant m$. ■

Taking $U = E$, we obtain the Hilbert syzygy theorem:

COROLLARY 6.11 *For any ring $E$, if $x$ is a central indeterminate, then* $\mathrm{r\cdot gl\cdot dim}(E[x]) = \mathrm{r\cdot gl\cdot dim}(E) + 1$.

For $U^n \cong E$, as right $E$-module; if the generator of $U$ is written $x$, then $T(U) \cong E[x]$, as is easily verified. The rest is clear from (i), since $A \otimes U \cong A$. ■

In particular, since any field is semisimple, we have

$$\mathrm{gl\cdot dim}(k[x_1,\ldots,x_n]) = n$$

for any field $k$, even skew.

This means that when we resolve a module $M$ over the polynomial ring $k[x_1,\ldots,x_n]$ by a free resolution $F = (F_i)$ say, then the module of relations in $F_0$ is not generally free and any generating set for these relations has further relations. These relations are known as *syzygies*, from the Greek 'suzugos': yoked or paired (usually applied to the conjunction of two heavenly bodies). By the above result, any free resolution of at most $n$ steps leads to a projective module; Hilbert's theorem states slightly more than this, namely the free resolution $F$

can be chosen so as to terminate at the $n$th step (with a free module). This sharpening is also a consequence of the rather deeper theorem of Quillen and Suslin (cf. Lam [1978]), which states that every finitely generated projective module over a polynomial ring $k[x_1, \ldots, x_n]$ over a field $k$ is free.

When $K$ is any commutative ring and $U$ is free as $K$-module, then by taking $E = K$ we obtain

COROLLARY 6.12 *For any commutative ring $K$ and any set $X$,*

$$\mathrm{r\cdot gl\cdot dim}(K\langle X\rangle) = \mathrm{l\cdot gl\cdot dim}(K\langle X\rangle) = \mathrm{gl\cdot dim}(K) + 1. \quad \blacksquare$$

## Exercises

(1) Show that if $\mathrm{Ext}^n_R(A, C) = 0$ for all cyclic modules $C$, then the same holds for all finitely generated modules $C$. Do the same for $\mathrm{Ext}^n_R(C, A)$ and $\mathrm{Tor}^R_n(A, C)$.

(2) Verify that two extensions of $R$-modules $A$ by $B$ are isomorphic iff they correspond to the same element of $\mathrm{Ext}^1_R(B, A)$.

(3) Show that Tor preserves direct sums in both arguments, while Ext preserves direct products in the second argument and converts direct sums in the first argument to direct products.

(4) Verify that the two ways of defining the characteristic class of an extension agree.

(5) Let $F$ be a bifunctor from $R$-modules which is covariant right exact and preserves direct sums. Show that $F$ is balanced iff $F(R, -)$ and $F(-, R)$ are exact. Similarly if $F$ is contravariant left exact and converts direct sums to direct products.

(6) Given a situation $(A_R, {}_RB_S)$, where $A_R$ and $B_S$ are flat, show that $(A \otimes B)_S$ is flat. Deduce that if an $E$-bimodule $U$ is flat as right $E$-module, then so is $\mathbf{T}_E(U)$. Do the same for 'projective' in place of 'flat'.

(7) For any abelian group $A$ denote by $tA$ its torsion subgroup. Show that $\mathrm{Tor}^{\mathbf{Z}}_1(\mathbf{Q}/\mathbf{Z}, A) \cong tA$.

(8) Show that $\mathrm{Tor}^{\mathbf{Z}}_1(\mathbf{C}_h, \mathbf{C}_k) \cong \mathbf{C}_d$, where $d$ is the highest common factor of $h$ and $k$.

(9) Show that for any abelian groups $A$, $B$, $\mathrm{Tor}_1(A, B) \cong \mathrm{Tor}_1(tA, B)$; hence calculate $\mathrm{Tor}_1(A, B)$ for any two finitely generated abelian groups.

(10) For any abelian group $A$ show that $\mathrm{Ext}^1(\mathbf{C}_n, A) \cong A/nA$, by applying $\mathrm{Hom}(-, A)$ to a suitable short exact sequence. Deduce that any extension of finite abelian groups $A$ by $B$ splits if $A$, $B$, are of coprime orders.

(11) A finitely generated $R$-module $M$ is called *stably free* if integers $r$, $s$ exist such that

$M \oplus R^r \cong R^s$. Given that every finitely generated projective over a polynomial ring is stably free, derive Hilbert's form of the syzygy theorem from Cor. 6.11.

(12) Let $k$ be a field and $R$ the $k$-algebra generated by (disjoint) finite sets $X_1, \ldots, X_r$ with the defining relations $xy = yx$ precisely when $x, y$ lie in different sets. Show that $\mathrm{gl}\cdot\dim(R) = r$. Find the global dimension of $R$ when the relation $xy = yx$ holds precisely for $x, y$ in the same set.

(13) Show that (7) is not always an isomorphism. (Hint: Use Th. 6.6 and Ex. (8) of **3.4**.)

## Further exercises on Chapter 3

(1) Let $K$ be a commutative ring; describe products and coproducts in the category of $K$-algebras.

(2) Let $\mathscr{C}$ be a small category admitting products. If for some objects $X, Y$, $\mathscr{C}(X, Y)$ has two distinct members, show that $\mathscr{C}(X, Y^I)$ has at least $2^I$ members, for any set $I$. Deduce that any $\mathscr{C}(X, Y)$ has at most one morphism (and so $\mathscr{C}$ is a preordered set).

(3) Show that every map $\alpha: A \to B$ in an abelian category gives rise to an exact sequence

$$0 \to \ker \alpha \to A \xrightarrow{\alpha} B \to \operatorname{coker} \alpha \to 0. \tag{1}$$

Given $f: A \to B$ and $g: B \to A$ such that $fg = 1$, show that $f$ is monic and $g$ epic. By applying the windmill lemma (Ex. (10) of **3.1**) to the sequences obtained from (1) for $f, g$ deduce that $A$ is isomorphic to a summand in a coproduct representation of $B$.

(4) Prove Yoneda's lemma for additive categories: If $F: \mathscr{A} \to Ab$ is given and for $p \in X^F$ a natural transformation $\dot{p}: h^X \to F$ is defined by the rule that for $\alpha \in \mathscr{A}(X, Y)$, $\alpha \mapsto p\alpha^F$ maps $Yh^X$ to $Y^F$, verify the naturality and show that the resulting map $X^F \to \mathrm{Nat}(h^X, F)$ to the set of all natural transformations is an isomorphism.

(5) If $G: \mathscr{A} \to Ab$ is a contravariant functor, show that $X^G \cong \mathrm{Nat}(h_X, G)$ where $h_X: Y \to \mathscr{A}(Y, X)$. Deduce that

$$\mathscr{A}(X, Y) \cong \mathrm{Nat}(h^Y, h^X) \cong \mathrm{Nat}(h_X, h_Y).$$

(6) Use Yoneda's lemma to show that the left adjoint of a functor, if it exists, is unique up to isomorphism.

(7) Show that a functor is exact iff it is right exact and preserves monics or left exact and preserves epics.

(8) Show that any left exact functor preserves pullbacks.

(9) Show that in any category $\mathscr{A}$, the product of two $\mathscr{A}$-objects $X, Y$ is the object representing the functor $A \mapsto \mathscr{A}(A, X) \times \mathscr{A}(A, Y)$ (if it exists). Similarly their coproduct is the object representing $A \mapsto \mathscr{A}(X, A) \times \mathscr{A}(Y, A)$.

(10) Let $X$ be an object in an abelian category. Assuming that the equivalence classes of subobjects of $X$ form a set, partially ordered by inclusion, show that this set is a modular lattice.

(11) In an additive category consider the sequence

$$P \xrightarrow{\lambda} A \coprod B \xrightarrow{\mu} Q.$$

If $i, j, p, q$ are the natural injections and projections of the biproduct $A \coprod B$, then the square formed by the maps $\lambda p, -\lambda q, i\mu, j\mu$ commutes if $\lambda\mu = 0$, is a pullback if $\lambda = \ker \mu$ and is a pushout if $\mu = \operatorname{coker} \lambda$.

(12) Given a pullback in an additive category (with notation as in **3.1**) show that $\alpha'$ is monic iff $\alpha$ is (for abelian categories this follows from Prop. 1.6, but not in general).

(13) Let $P$ be a pullback of $\alpha: A \to C$, $\beta: B \to C$ in the category of rings. Show that if $P$ is an integral domain, then one of $\alpha$, $\beta$, say $\alpha$ is injective, and $P$ is isomorphic to a subring of $B$.

(14) Show that in any abelian category the intersection of two subobjects of a given object can be defined as a pullback and describe the dual concept.

(15) Given a $3 \times 3$ 'matrix' of short exact sequences between modules $U_{ij} (i, j = 1, 2, 3)$ forming a commutative diagram as in the $3 \times 3$ lemma, show that the kernel of the composite map $U_{22} \to U_{33}$ is $\operatorname{im}(U_{12}) + \operatorname{im}(U_{21})$. Deduce that for any two short exact sequences of modules $U_i$, $V_i$ the kernel of the map $U_2 \otimes V_2 \to U_3 \otimes V_3$ is $\operatorname{im}(U_1 \otimes V_2) + \operatorname{im}(U_2 \otimes V_1)$.

(16) Let $\mathscr{A}, \mathscr{B}$ be abelian categories with direct sums and $F, G$ two right exact functors from $\mathscr{A}$ to $\mathscr{B}$, which preserve coproducts. Show that if there is a natural transformation $t: F \to G$ such that for a generator $P$ of $\mathscr{A}$, $t: P^F \cong P^G$ is an isomorphism, then $t$ is a natural isomorphism. (Hint: apply $t$ to a 'presentation' of an object and use the 5-lemma.)

(17) (Eilenberg–Watts) Show that for any functor $S: \mathscr{M}_A \to \mathscr{M}_B$ the following are equivalent: (a) $S$ has a right adjoint $T: \mathscr{M}_B \to \mathscr{M}_A$, (b) $S$ is right exact and preserves coproducts, (c) $S = - \otimes_A P$, for some $(A, B)$-bimodule $P$, unique up to isomorphism. When (a)–(c) hold, the adjoint is $Y^T = \operatorname{Hom}_B(P, Y)$. (Hint: use Ex. (16); see also **6.4**.)

(18) Show that for a finitely generated projective $R$-module $P$, $\operatorname{Hom}_R(P, M) \cong P^* \otimes M$, where $P^* = \operatorname{Hom}_R(M, R)$. (Hint: use Ex. (16).)

(19) A complex $(C, d)$ is said to *split* if it is homotopic to $(H(C), 0)$. Show that $(C, d)$ splits iff (a) $C$ has a homotopy $s$ of degree 1 such that $dsd = d$ or equivalently, (b) $B(C)$, $Z(C)$ are direct summands of $C$. (Hint: if $s$ is a homotopy as in (a) then $sd$ is a projection on $B(C)$ and $1 - ds$ a projection on $Z(C)$.)

(20) Given a chain map $f: X \to X'$ between complexes, to obtain an exact sequence which

includes the maps $f_*: H_nX \to H_nX'$, define a new complex $M(f)$, the *mapping cone* of $f$, as $M_n = X_{n+1} \oplus X'_n$, with $(x, x')d = (-xd, x'd + xf)$. Verify that we have a short exact sequence of complexes

$$0 \to X' \to M \to \Sigma X \to 0,$$

where $(\Sigma X)_n = X_{n-1}$, and obtain the associated exact homology sequence. Deduce that $M(f)$ is acyclic iff $f_*$ is a chain equivalence.

(21) Given a chain map $f: X \to X'$ between complexes, define the *mapping cylinder* $N(f)$ on $f$ as the complex $N_n = C_n \oplus C_{n-1} \oplus C'_n$, with $(x, y, x')d = (xd + y, -yd, x'd - yf)$. Verify that we have a short exact sequence of complexes

$$0 \to X \to N(f) \to M(f) \to 0,$$

where $M(f)$ is the mapping cone from Ex. (20). Using the associated homology sequence, show that $\alpha: X' \to N(f)$, $\beta: N(f) \to X'$ are mutually homotopy inverse, where $\alpha: x' \mapsto (0, 0, x')$ and $\beta: (x, y, x') \mapsto x' + xf$.

(22) ('Eilenberg-trick') Let $P$ be a projective module, say $F = P \oplus P'$ is free. By expressing the direct sum $S$ of countably many copies of $F$ in two ways, show that $P \oplus S \cong S$.

(23) Show that the triangular matrix ring $\begin{pmatrix} k & k \\ 0 & k \end{pmatrix}$ over a field $k$ is hereditary. Is it principal?

(24) Given a short exact sequence of modules

$$0 \to A' \to A \to A'' \to 0,$$

show that $\mathrm{hd}(A) \leqslant \max\{\mathrm{hd}(A'), \mathrm{hd}(A'')\}$ with equality except possibly when $\mathrm{hd}(A'') = \mathrm{hd}(A') + 1$. Find similar bounds for $\mathrm{hd}(A')$, $\mathrm{hd}(A'')$ in terms of the other two.

(25) Let $R$ be a ring and $M$ an $R$-module. Show that if $\mathrm{gl}\cdot\dim(R) = r$ and $\mathrm{hd}(M) = r - 1$, then every submodule of $M$ has hd at most $r - 1$.

(26) A ring is said to be right *semihereditary* if every finitely generated right ideal is projective. Show that in a right semihereditary ring the finitely generated right ideals form a sublattice of the lattice of all right ideals.

(27) A ring $R$ is said to be *weakly semihereditary* if, given two maps $\alpha: P_0 \to P$, $\beta: P \to P_1$ between finitely generated projective modules such that $\alpha\beta = 0$, we can write $P = P' \oplus P''$ such that $\mathrm{im}\,\alpha \subseteq P' \subseteq \ker\beta$. Show that $R$ is weakly semihereditary iff for any matrices $A \in {}^rR^n$, $B \in {}^nR^s$ such that $AB = 0$ there exists an idempotent $n \times n$ matrix $E$ over $R$ such that $AE = A$, $EB = 0$. Deduce that the condition is left–right symmetric. Show also that every right (or left) semihereditary ring is weakly semihereditary.

(28) Show that over a right semihereditary ring $R$ every finitely generated submodule

of a projective module is projective, and that every finitely generated projective module is a direct sum of a finite number of finitely generated right ideals.

(29) Show that an injective module $E$ is a cogenerator iff $\mathrm{Hom}_R(S, E) \neq 0$ for every simple $R$-module $S$. Verify that $\mathbf{Q}/\mathbf{Z}$ is an injective cogenerator for $\mathbf{Z}$.

(30) Show that for any commutative ring $K$, $\mathrm{Ext}^n_K(A, B)$ and $\mathrm{Tor}^K_n(A, B)$ have a natural $K$-module structure. Show that if, moreover, $K$ is Noetherian and $A, B$ are finitely generated, then $\mathrm{Ext}^n_K(A, B)$ and $\mathrm{Tor}^K_n(A, B)$ are finitely generated $K$-modules.

# 4
# Further group theory

Group theory has developed so much in recent years that a separate volume would be needed even for an introduction to all the main areas of research. The most a chapter can hope to do is to give the reader a taste by a selection of topics; our choice was made on the basis of general importance or interest, or relevance in later applications. Thus ideas from extension theory are used in the study of simple algebras, while notions like the transfer, Frattini and Fitting subgroups have their counterparts in rings in the form of determinants or radicals. Hall subgroups are basic in the deeper study of finite groups, the linear groups lead to an important class of simple groups, while free groups and commutators have their analogues in rings and Lie algebras respectively (though we shall not get as far as Lie algebras in this volume).

We recall some standard notation from A.1. If $G$ is a group generated by a set $X$, we write $G = \mathrm{gp}\{X\}$. For subsets $X, Y$ of $G$, $XY$ denotes the set of all products $xy$, where $x \in X, y \in Y$. We write $N \lhd G$ to indicate that $N$ is a normal subgroup in $G$, i.e. mapped into itself by all inner automorphisms of $G$. If $H, K$ are subgroups of $G$, then $HK$ is a subgroup precisely when $HK = KH$; in particularly this holds if $H$ or $K$ is normal in $G$. We also recall the *modular law*: given subgroups $K, L, M$ of $G$, if $K \subseteq M$, then $K(L \cap M) = KL \cap M$.

## 4.1 Group extensions

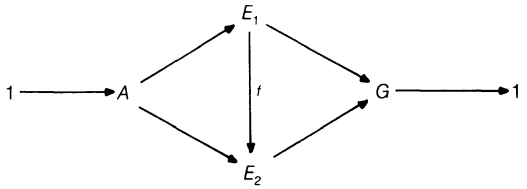
In A.1, **9.2** we saw that every finite group has a composition series; the factors of this series are simple groups and for the structure of $G$ one has (i) to study these simple groups, and (ii) to determine how $G$ is composed from them. For the moment we concentrate on (ii); in its simplest form this is the extension problem. It is of great general interest and some special cases will be of use to us later.

An extension of groups may be written as a short exact sequence

$$1 \to A \xrightarrow{\lambda} E \xrightarrow{\mu} G \to 1. \tag{1}$$

As in the case of abelian groups (**Z**-modules) this means that $A$ is isomorphic to a normal subgroup $A_1$ of $E$ and $E/A_1 \cong G$. We shall call $E$ an *extension of*

$A$ *by* $G$. Given any two groups $A, G$, their direct product $G \times A$ is such an extension, but in general there will be many others. Two extensions of $A$ by $G$, say $E_1$ and $E_2$ are said to be *isomorphic* if there is an isomorphism $f: E_1 \to E_2$ such that the diagram




commutes. By the 5-lemma any such homomorphism $f$ is an isomorphism. The aim of extension theory is to obtain a survey of all (isomorphism classes of) extensions of a prescribed pair of groups. In principle this is solved by Schreier's theorem (Th. 1.2 below), but the solution is not very explicit and the description in terms of cohomology groups is more illuminating, although not exhaustive.

We begin by discussing an important special case, that of split extensions. The extension (1) is said to *split* if there is a homomorphism $\alpha: G \to E$ such that $\alpha\mu = 1$; this means that we can choose the transversal for the subgroup $A\lambda$ in $E$ to be a subgroup (not necessarily normal in $E$). Consider any split extension (1) and denote the images of $G$, $A$ in $E$ by $G_1, A_1$ respectively. Any element $x$ of $E$ has the same image $x\mu$ as some element $g \in G_1$, thus $(g^{-1}x)\mu = 1$, and by exactness, $g^{-1}x \in A_1$. Hence each $x \in E$ has the form

$$x = ga, \qquad \text{where } g \in G_1, a \in A_1. \tag{2}$$

This just means that $E = G_1 A_1$. Next we observe that $G_1 \cap A_1 = 1$, for the restriction of $\mu$ to $G_1$ is injective and its kernel is $G_1 \cap A_1$. It follows that the expression (2) for $x$ is unique: if $x = ga = g'a'$, where $g, g' \in G_1, a, a' \in A_1$, then $a'a^{-1} = g'^{-1}g \in G_1 \cap A_1 = 1$, hence $a' = a, g' = g$. Let us identify $G$ with $G_1$ and $A$ with $A_1$, so that we have $E = G \times A$, as *sets*. By hypothesis $A$ is normal in $E$; if $G$ is also normal in $E$, then $E$ is just the direct product of $G$ and $A$ (A.1, Th. 1 of **9.6**). But in general $G$ need not be normal in $E$; each element of $G$ then defines an inner automorphism of $E$, which induces an automorphism of the normal subgroup $A$. If we write $\alpha_g$ for the automorphism of $A$ induced by $g \in G$, then we have the commutation rule

$$ag = g \cdot a\alpha_g, \qquad \text{for any } a \in A, g \in G. \tag{3}$$

It is easily verified that the mapping

$$g \mapsto \alpha_g \tag{4}$$

is a homomorphism from $G$ to Aut $A$, the automorphism group of $A$, and (3) is enough to determine the multiplication in $E$:

$$(ga)(g'a') = gg' \cdot (a\alpha_{g'})a'. \tag{5}$$

Thus the extension $E$ is determined by $G, A$ and the action of $G$ on $A$; this is just the *semidirect* product of $G$ and $A$ with the action $\alpha$, which we have met in A.1, **9.6**, denoted by $G \ltimes A$ or $G \ltimes_\alpha A$. Given any groups $G, A$ and a homomorphism $\alpha: G \to \mathrm{Aut}\, A$, we can define a group structure on the set $G \times A$ by the rule

$$(g, a)(g', a') = (gg', a\alpha_{g'} a').$$

We saw in A.1, **9.6** (and can easily verify directly) that $E$ is a group in which the elements $(g, 1)$ form a subgroup isomorphic to $G$ and the elements $(1, a)$ form a normal subgroup isomorphic to $A$, with quotient isomorphic to $G$; thus $E$ is a split extension of $A$ by $G$. We state the result as

THEOREM 1.1 *Let $G, A$ be groups and $\alpha: G \to \mathrm{Aut}\, A$ a homomorphism. Then the semidirect product $G \ltimes_\alpha A$ is a split extension of $A$ by $G$ with action $\alpha$, and all split extensions arise in this way.* ■

For example, the dihedral group $\mathbf{D}_m$ of order $2m$ can be written as a semidirect product:

$$\mathbf{D}_m \cong \mathbf{C}_2 \ltimes \mathbf{C}_m,$$

where $\mathbf{C}_2$ acts on $\mathbf{C}_m$ by the automorphism $x \mapsto x^{-1}$.

Let us now return to general extensions. Given any extension (1), where we take $A$ to be identified with its image in $E$ by means of $\lambda$, let us denote the elements of $A$ by latin letters and those of $G$ by greek letters. For each $\alpha \in G$ choose an element $g_\alpha \in E$ which maps to $\alpha$, taking $g_1 = 1$ in order to simplify later formulae. In general it will no longer be possible to choose the $g_\alpha$ to form a subgroup (they form a transversal of $G$ in $E$), but in any case we have $g_\alpha g_\beta \equiv g_{\alpha\beta} \pmod{A}$, hence

$$g_\alpha g_\beta = g_{\alpha\beta} m_{\alpha,\beta} \qquad \text{where } m_{\alpha,\beta} \in A, \tag{6}$$

and since $g_1 = 1$, we have

$$m_{\alpha,1} = m_{1,\alpha} = 1. \tag{7}$$

Further, each $g_\alpha$ induces an automorphism $\theta_\alpha$ of $A$:

$$g_\alpha^{-1} a g_\alpha = a\theta_\alpha \qquad \text{for all } a \in A, \alpha \in G, \tag{8}$$

where $\theta_1 = 1$. It is no longer true that $\theta$ has to be a homomorphism, i.e. $\theta_\alpha \theta_\beta$ will not in general be equal to $\theta_{\alpha\beta}$, but will differ from it by an inner automorphism: by (6) we have

$$\theta_\alpha \theta_\beta = \theta_{\alpha\beta} \iota(m_{\alpha,\beta}), \tag{9}$$

where for any $x \in A$, $\iota(x): u \mapsto x^{-1}ux$ is the inner automorphism defined by $x$. We recall that the set $\mathrm{Inn}\, A$ of all these inner automorphisms is a normal subgroup

of Aut $A$; the quotient $\mathrm{Aut}(A)/\mathrm{Inn}(A)$ is called the *automorphism class group* and (9) shows that we have a homomorphism

$$\bar{\theta}: G \to \mathrm{Aut}(A)/\mathrm{Inn}(A). \tag{10}$$

The set $\{m_{\alpha\beta}\}$ is called a *factor set* of the extension, *normalized* because it satisfies (7). In $E$ we have $g_\alpha(g_\beta g_\gamma) = g_\alpha g_{\beta\gamma} m_{\beta,\gamma} = g_{\alpha\beta\gamma} m_{\alpha,\beta\gamma} m_{\beta,\gamma}$ and $(g_\alpha g_\beta)g_\gamma = g_{\alpha\beta} m_{\alpha,\beta} g_\gamma = g_{\alpha\beta} g_\gamma (m_{\alpha,\beta}\theta_\gamma) = g_{\alpha\beta\gamma} m_{\alpha\beta,\gamma}(m_{\alpha,\beta}\theta_\gamma)$ hence by the associative law we obtain the *factor set condition*

$$m_{\alpha,\beta\gamma} m_{\beta,\gamma} = m_{\alpha\beta,\gamma}(m_{\alpha,\beta}\theta_\gamma) \qquad \text{for all } \alpha, \beta, \gamma \in G. \tag{11}$$

Conversely, given any groups $G$ and $A$ and mappings $\theta: G \to \mathrm{Aut}\, A$ and $m: G^2 \to A$ such that $\theta_1 = 1$ and (7), (9) and (11) hold, we can define a multiplication on the set $G \times A$ by putting

$$(\alpha, a)(\beta, b) = (\alpha\beta, m_{\alpha,\beta}(a\theta_\beta)b); \tag{12}$$

then it is straightforward to verify that we obtain a group in this way, which is an extension of $A$ by $G$ with factor set $m_{\alpha,\beta}$ and automorphisms $\theta_\alpha$. The associative law follows from (11), the neutral is $(1, 1)$ and the inverse of $(\alpha, a)$ is $(\alpha^{-1}, (m_{\alpha^{-1},\alpha} a^{-1})\theta_\alpha^{-1})$; the verification may be left to the reader. Here the normalization (7) is not essential, but without it the formulae become a little more complicated.

Let $E$ be an extension with factor set $\{m_{\alpha,\beta}\}$ arising from the transversal $\{g_\alpha\}$. If $\{g'_\alpha\}$ is a second transversal, then

$$g'_\alpha = g_\alpha c_\alpha, \qquad \text{where } c_\alpha \in A,\ c_1 = 1,$$

and the new factor set $\{m'_{\alpha,\beta}\}$ obtained from the $g'_\alpha$ is related to the old by the equations $g_{\alpha\beta} c_{\alpha\beta} m'_{\alpha,\beta} = g'_{\alpha\beta} m'_{\alpha,\beta} = g'_\alpha g'_\beta = g_\alpha c_\alpha g_\beta c_\beta = g_\alpha g_\beta (c_\alpha \theta_\beta) c_\beta = g_{\alpha\beta} m_{\alpha,\beta}(c_\alpha\theta_\beta) c_\beta$. Hence

$$m'_{\alpha,\beta} = c_{\alpha\beta}^{-1} m_{\alpha,\beta}(c_\alpha\theta_\beta) c_\beta; \tag{13}$$

we shall express (13) by saying that $\{m_{\alpha,\beta}\}$ and $\{m'_{\alpha,\beta}\}$ are *associated*. Similarly, from (8) we obtain

$$\theta'_\alpha = \theta_\alpha \iota(c_\alpha). \tag{14}$$

Conversely, if $m'_{\alpha,\beta}$ and $\theta'_\alpha$ are defined by (13) and (14), they lead to an extension isomorphic to the given one, as we see by retracing our steps. We sum up these results in

THEOREM 1.2 (Schreier's extension theorem) *Given two groups $G$ and $A$, a homomorphism $\bar{\theta}: G \to \mathrm{Aut}\,(A)/\mathrm{Inn}\,(A)$ and a map $m: G^2 \to A$ satisfying the factor set condition, we can define a multiplication on the set $G \times A$ by* (12) *and so obtain a group which is an extension of $A$ by $G$. All extensions of $A$ by $G$ are obtained in this way and two extensions are isomorphic if and only if their factor sets are associated.* ■

To prove this in full, some more verifications are needed, which are straightforward (though tedious) and will therefore be left to the reader. Instead we shall examine a special case, important for the applications, in more detail.

Let us assume that $A$ is abelian; in that case $\operatorname{Inn} A$ is trivial and so the map $\theta: G \to \operatorname{Aut} A$ is a homomorphism. Moreover, as (14) shows, the automorphism $\theta_\alpha$ depends only on $\alpha$ and not on the choice of $\theta$. So in this case we have an action of $G$ on $A$ (by automorphisms) and in place of $x\theta_\alpha$ we shall simply write $x^\alpha$. This action is trivial precisely when $A$ is contained in the centre of $E$; we call this a *central extension.*

In what follows we shall write $A$ as an additive group, but $G$ will still be multiplicative. The action of $G$ then turns $A$ into a $G$-module and the factor set condition (11) reads

$$m_{\alpha,\beta\gamma} + m_{\beta,\gamma} = m_{\alpha\beta,\gamma} + m_{\alpha,\beta}^{\gamma}, \tag{15}$$

while associated factor sets are related by the equations

$$m'_{\alpha,\beta} = m_{\alpha,\beta} + c_\alpha^\beta + c_\beta - c_{\alpha\beta}. \tag{16}$$

Our aim is to obtain a homological description of all extensions with abelian kernel. We recall that $G$-modules may equally well be regarded as $\mathbf{Z}G$-modules where $\mathbf{Z}G$ is the group algebra of $G$ over $\mathbf{Z}$. Moreover, any left $G$-module $A$ can be regarded as right $G$-module by using the canonical antiautomorphism of $\mathbf{Z}G$. Explicitly we put

$$a \cdot g = g^{-1}a \qquad \text{for all } a \in A,\ g \in G.$$

Let $A, B$ be any right $G$-modules and consider $\operatorname{Hom}(A, B)$, the group of all (abelian group) homomorphisms from $A$ to $B$. We can define a $G$-module structure on this group as follows: If $f \in \operatorname{Hom}(A, B)$, we put

$$f^s: a \mapsto (f(as^{-1}))s \qquad \text{for } s \in G. \tag{17}$$

This is indeed a $G$-module structure, for (17) can be written

$$f^s(as) = f(a)s,$$

hence $f^{st}(ast) = (f(a))st = (f^s(as))t = (f^s)^t(ast)$, therefore $f^{st} = (f^s)^t$, and the remaining laws are clear.

For any $G$-module $A$ we define

$$A^G = \{x \in A \mid xs = x \text{ for all } s \in G\};$$

thus $A^G$ is the largest submodule left fixed by $G$. For example, if $A$, $B$ are any $G$-modules, then $f \in \operatorname{Hom}(A, B)$ is left fixed by $G$ iff $f^s = f$, i.e. $f(as) = f(a)s$ for all $s \in G$. This just means that $f$ is a $G$-homomorphism from $A$ to $B$, hence we have

$$(\operatorname{Hom}(A, B))^G = \operatorname{Hom}_G(A, B). \tag{18}$$

In particular, regarding $\mathbf{Z}$ as $G$-module by the trivial action: $ns = n$ for all $n \in \mathbf{Z}$

and $s \in G$, we have

$$\mathrm{Hom}_G(\mathbf{Z}, A) = (\mathrm{Hom}(\mathbf{Z}, A))^G \cong A^G. \tag{19}$$

It is clear from (19) that the functor $A \mapsto A^G$ is left exact; we can therefore form the right derived functor, as described in **3.5–6**. The $n$th right derived functor of $A^G$ is written $H^n(G, A)$ and is called the $n$th *cohomology group* of the $G$-module $A$. By (19) we see that

$$H^n(G, A) = \mathrm{Ext}^n_G(\mathbf{Z}, A), \qquad H^0(G, A) = A^G,$$

where for the suffix on the right we have written $G$ rather than $\mathbf{Z}G$. We note that for any coinduced module $A$ we have $H^n(G, A) = 0$. For $\mathbf{Z}G$ is free as $\mathbf{Z}$-module, hence by the change-of-rings formula (13) of Th. 3.6.9, we have

$$\mathrm{Ext}^n_G(\mathbf{Z}, B^f) \cong \mathrm{Ext}^n_{\mathbf{Z}}(\mathbf{Z}, B) = 0 \qquad \text{for all } n \geqslant 1.$$

This holds for any coinduced module $B^f = \mathrm{Hom}_{\mathbf{Z}}(\mathbf{Z}G, B)$, where $f: \mathbf{Z} \to \mathbf{Z}G$.

Similarly the $n$th *homology group* of the $G$-module $A$ is defined as

$$H_n(G, A) = \mathrm{Tor}^G_n(\mathbf{Z}, A).$$

If $A$ is induced, say $A = B_f = \mathbf{Z}G \otimes_{\mathbf{Z}} B$, then $\mathrm{Tor}^G_n(\mathbf{Z}, B_f) \cong \mathrm{Tor}^{\mathbf{Z}}_n(\mathbf{Z}, B) = 0$ for all $n \geqslant 1$, hence we have $H_n(G, A) = 0$ for any induced $G$-module $A$.

Let $\varepsilon: \mathbf{Z}G \to \mathbf{Z}$ be the augmentation map, defined by $(\sum a_s s)\varepsilon = \sum a_s$. The kernel $IG$ is called the *augmentation ideal* of $\mathbf{Z}G$, and from the exact sequence

$$0 \to IG \to \mathbf{Z}G \to \mathbf{Z} \to 0 \tag{20}$$

we obtain the exact sequence

$$0 \to (IG)A \to A \to \mathbf{Z} \otimes A \to 0.$$

Hence $\mathbf{Z} \otimes A \cong A/(IG)A$ and we see that $H_n(G, A)$ is the left derived functor of $A/(IG)A$. We note that $A/(IG)A$, sometimes written $A_G$, is the largest quotient of $A$ with trivial $G$-action, as is easily seen. Hence we find that

$$H_0(G, A) \cong \mathbf{Z} \otimes A \cong A_G. \tag{21}$$

From the exact sequence (20) and the definitions of $H_n$, $H^n$ we obtain by shifting dimensions,

$$H_n(G, A) = \mathrm{Tor}^G_{n-1}(A, IG), \tag{22}$$

$$H^n(G, A) = \mathrm{Ext}^{n-1}_G(IG, A). \tag{23}$$

We recall that to construct $H_n(G, A)$ or $H^n(G, A)$ we take a projective resolution $X$ of $\mathbf{Z}$ (as trivial $G$-module) and form the complex $A \otimes X$ or $\mathrm{Hom}(X, A)$ respectively. Taking first $H_n(G, A)$ let us form $A \otimes X$, where the $X_n$ are now left modules; thus $X_n$ consists of all $n$-chains, i.e. $(n+1)$-tuples over $G$ with the action

$$g(s_0, s_1, \ldots, s_n) = (gs_0, gs_1, \ldots, gs_n).$$

The differential $d: X_n \to X_{n-1}$ is given by the formula

$$(s_0, \ldots, s_n)d = \sum_{i=0}^{n} (-1)^i (s_0, \ldots, \hat{s}_i, \ldots, s_n), \tag{24}$$

where the caret on $\hat{s}_i$ means that this term is to be omitted. The augmentation map $\varepsilon: X_0 \to \mathbf{Z}$ is defined by $s_0\varepsilon = 1$. With these definitions we have $d^2 = 0$, as we see by counting how often the term $(s_0, \ldots, \hat{s}_p, \ldots, \hat{s}_q, \ldots, s_n)$, where $p < q$, occurs in $(s_0, \ldots, s_n)d^2$.

To prove exactness, we define a homotopy $h: X_n \to X_{n+1}$ by the rule: $(s_0, \ldots, s_n)h = (1, s_0, \ldots, s_n)$ for $n \geqslant 0$, $1.h = 1$ for $n = -1$. Then it is easily verified that $hd + dh = 1$. This resolution is called the *standard* (or *bar*) *resolution* of $\mathbf{Z}$ over $\mathbf{Z}G$, in homogeneous form. The elements of $\ker d$, $\operatorname{im} d$ are called *cycles* and *boundaries* respectively. Sometimes the *inhomogeneous* form is more convenient to use; this is defined as

$$[t_1|t_2|\cdots|t_n] = (1, t_1, t_1t_2, \ldots, t_1 \ldots t_n) \qquad n \geqslant 1, [\ ] = (1). \tag{25}$$

Now the differential is given by

$$[t_1|t_2|\cdots|t_n]d = t_1[t_2|\cdots|t_n] + \sum(-1)^i[t_1|\cdots|t_{i-1}|t_it_{i+1}|\cdots|t_n] + (-1)^n[t_1|\cdots|t_{n-1}].$$

In low dimensions we have

$$(1)\varepsilon = (1), \quad [g]d = g[\ ] - [\ ] = g - 1, \quad [g|h]d = g[h] - [gh] + [g]. \tag{26}$$

As an illustration let us calculate $H_1(G, \mathbf{Z})$. We take the resolution $X$ of $\mathbf{Z}$ and tensor with $\mathbf{Z}$:

$$\cdots \mathbf{Z} \otimes X_2 \to \mathbf{Z} \otimes X_1 \to \mathbf{Z} \otimes X_0 \to \mathbf{Z} \to 0.$$

Since $\mathbf{Z}$ is a trivial $G$-module, we see that at $\mathbf{Z} \otimes X_1$ the kernel is all of $\mathbf{Z} \otimes X_1$ while the image is generated by the elements $1 \otimes u$, where $u = [h] - [gh] + [g]$, by (26). Thus $H_1(G, \mathbf{Z})$ consists of all sums of terms $1 \otimes [g]$ subject to the relation

$$1 \otimes [gh] = 1 \otimes [g] + 1 \otimes [h].$$

This is just $G$ made abelian: $G^{\mathrm{ab}} = G/G'$, hence we have

PROPOSITION 1.3 *For any group $G$ we have $H_1(G, \mathbf{Z}) = \operatorname{Tor}_1^G(\mathbf{Z}, \mathbf{Z}) \cong G^{\mathrm{ab}}$.* ■

We next turn to the construction of $H^n(G, A)$ using the standard resolution. The elements of $\operatorname{Hom}_G(X_n, A)$ are the *n-cochains*, those of $\ker d$, $\operatorname{im} d$ the *cocycles* and *coboundaries* respectively. Taking $X_n$ now as right $G$-module, we find that an $n$-cochain is a function $f: G^{n+1} \to A$ such that

$$f(s_0g, \ldots, s_ng) = f(s_0, \ldots, s_n)g.$$

Since $g$ is arbitrary in $G$, $f$ is completely determined by its values when its last argument is 1. Let us write

$$\varphi(t_1, \ldots, t_n) = f(t_1 \cdots t_n, t_2 \cdots t_n, \ldots, t_{n-1}t_n, t_n, 1).$$

$\varphi$ is the inhomogeneous cochain corresponding to the homogeneous cochain $f$. In terms of $\varphi$ the coboundary is given by the formula

$$(\varphi d)(t_1, \ldots, t_{n+1}) = \varphi(t_2, \ldots, t_{n+1}) + \sum(-1)^i \varphi(t_1, \ldots, t_i t_{i+1}, \ldots, t_{n+1}) + (-1)^{n+1} \varphi(t_1, \ldots, t_n) t_{n+1}. \tag{27}$$

Let us again write out the cases of low dimensions:

$n = 1$. A 1-cocycle is a map $\varphi: G \to A$ such that $\varphi d = 0$. Thus $C^1$ consists of all $\varphi$ satisfying

$$\varphi(gg') = \varphi(g') + \varphi(g)g'. \tag{28}$$

It is a coboundary iff

$$\varphi(g) = c - cg \qquad \text{for some fixed } c \in A. \tag{29}$$

A function $\varphi$ satisfying (28) is sometimes called a *derivation* (it is a $(1, \varepsilon)$-derivation in the sense of **2.2**) or a *crossed homomorphism*. If (29) holds it is said to be *inner* (or a *principal* crossed homomorphism). Writing $\mathrm{Der}(G, A)$ for the group of derivations and $\mathrm{IDer}(G, A)$ for the subgroup of inner derivations, we have

$$H^1(G, A) = C^1/B^1 \cong \mathrm{Der}(G, A)/\mathrm{IDer}(G, A).$$

If $G$ acts trivially on $A$, the derivations are just the homomorphisms and the inner derivations are 0, so in this case $H^1(G, A) = \mathrm{Hom}(G, A)$. Since $A$ is abelian, the homomorphisms $G \to A$ correspond to the homomorphisms $G^{\mathrm{ab}} \to A$, and so we obtain

PROPOSITION 1.4 *For any group $G$ and any module $A$ with trivial $G$-action $ag = a$ for $a \in A$, $g \in G$, we have*

$$H^1(G, A) \cong \mathrm{Hom}(G, A) \cong \mathrm{Hom}(G^{\mathrm{ab}}, A). \quad \blacksquare$$

$n = 2$. A 2-cocycle is a map $\varphi: G^2 \to A$ such that

$$\varphi(a, bc) + \varphi(b, c) = \varphi(ab, c) + \varphi(a, b)c.$$

This is precisely the condition (15) for a factor set; moreover, a 2-coboundary is a factor set of the form $\varphi(b) - \varphi(ab) + \varphi(a)b$; now (16) shows that two factor sets are associated iff they differ by a coboundary. This establishes

THEOREM 1.5 *Let $G$ be any group and $A$ a $G$-module. Then there is a natural bijection between the elements of $H^2(G, A)$ and the isomorphism classes of extensions of $A$ by $G$ with the given $G$-action on $A$.* $\blacksquare$

We observe that since $H^2(G, A)$ has a group structure, there is a multiplication of extensions of $A$ by $G$. This is obtained by taking the product (resp. sum) of the corresponding factor sets (resp. cocycles) and is known as the *Baer product* (resp. *sum*). Some sample calculations are suggested in the exercises; we add a general result which is often useful.

PROPOSITION 1.6 *Let $G$ be a group of finite order $r$ and $A$ any $G$-module. Then for any $n>0$, each element of $H^n(G, A)$ has order dividing $r$.*

*Proof.* We define a 'homotopy mod $r$' by the equation

$$(\varphi h)(s_2, \ldots, s_n) = \sum_g \varphi(g, s_2, \ldots, s_n).$$

If in (27) we sum over $t_1$ we obtain

$$(\varphi dh)(t_2, \ldots, t_{n+1}) = r\varphi(t_2, \ldots, t_{n+1}) - (\varphi hd)(t_2, \ldots, t_{n+1}).$$

Hence we have

$$dh + hd = r.1.$$

If $c$ is any cocycle, then $cd = 0$, hence $rc = cdh + chd = (ch)d$, therefore $rc$ is a coboundary and so $rH^n(G, A) = 0$, as claimed. ■

COROLLARY 1.7 *If $G, A$ are both finite, of coprime orders, then $H^n(G, A) = 0$.* ■

It is clear from Prop. 1.6 that if $A$ is uniquely divisible, as abelian group, then $H^n(G, A) = 0$ for any finite group $G$. This remark can be used to compute $H^2$ of a finite group over $\mathbf{Z}$:

PROPOSITION 1.8 *Let $G$ be any finite group and $\mathbf{K} = \mathbf{Q}/\mathbf{Z}$ the group of rational numbers* mod 1, *as trivial $G$-module. Then*

$$H^2(G, \mathbf{Z}) \cong \mathrm{Hom}(G, \mathbf{K}). \tag{30}$$

*Proof.* The exact sequence $0 \to \mathbf{Z} \to \mathbf{Q} \to \mathbf{K} \to 0$ leads to the derived sequence

$$H^1(G, \mathbf{Q}) \to H^1(G, \mathbf{K}) \to H^2(G, \mathbf{Z}) \to H^2(G, \mathbf{Q}).$$

Since $\mathbf{Q}$ is uniquely divisible, the extreme terms are 0 and so the other two are isomorphic: $H^2(G, \mathbf{Z}) \cong H^1(G, \mathbf{K})$. Since the $G$-action on $\mathbf{K}$ is trivial, $H^1(G, \mathbf{K}) \cong \mathrm{Hom}(G, \mathbf{K})$ and the result follows. ■

A similar proof, replacing $\mathbf{Q}$ by $\mathbf{R}$, shows that $H^2(G, \mathbf{Z}) \cong \mathrm{Hom}(G, \mathbf{T})$, where $\mathbf{T} \cong \mathbf{R}/\mathbf{Z}$. More directly we see that $\mathrm{Hom}(G, \mathbf{T}) = \mathrm{Hom}(G, \mathbf{K})$ for any finite group $G$, because every homomorphism of $G$ into $\mathbf{T}$ has its image in the torsion

subgroup of $\mathbf{T}$, and this is just $\mathbf{K}$. We also note that $\mathrm{Hom}(G, \mathbf{T}) \cong \mathrm{Hom}(G, \mathbf{C}^\times)$, because $\mathbf{T} \cong \mathbf{C}^\times$ by the isomorphism $x \mapsto \exp(2\pi i x)$.

The group $H^2(G, \mathbf{K})$ is called the *multiplicator* of $G$. It arises when we consider representations of $G$ by linear transformations of projective space, briefly *projective* transformations. If $\alpha \in G$ is represented by a matrix $\rho(\alpha)$ over $\mathbf{C}$ say, we have

$$\rho(\alpha)\rho(\beta) = c_{\alpha,\beta}\rho(\alpha\beta),$$

where $c_{\alpha,\beta} \in \mathbf{C}^\times$ is a factor set, corresponding to an element of $H^2(G, \mathbf{K})$.

As we have seen, from a short exact sequence of $G$-modules we can derive a long exact sequence by applying hom or the tensor product. This argument cannot be applied directly to an exact sequence describing a group extension

$$1 \to N \xrightarrow{\lambda} G \xrightarrow{\mu} L \to 1. \tag{31}$$

Nevertheless a similar result can be obtained on going via a short exact sequence of $L$-modules. It is given by

THEOREM 1.9 *Given an exact sequence* (31) *describing a group extension* $G$, *there is a* 5*-term exact sequence for any* $L$*-module* $A$:

$$0 \to H^1(L, A) \xrightarrow{\mu^*} H^1(G, A) \xrightarrow{\lambda^*} \mathrm{Hom}_L(N^{\mathrm{ab}}, A) \xrightarrow{t} H^2_L(L, A) \xrightarrow{\mu^*} H^2_G(G, A). \tag{32}$$

The map $\mu^*$, arising from $\mu$, is called the *inflation map* and $\lambda^*$, arising from the inclusion $\lambda$ is called the *restriction map.* The connecting map $t$ is called the *transgression.*

*Proof.* By tensoring (20) with $\mathbf{Z}L$ over $G$ we obtain

$$0 \to \mathrm{Tor}_1^G(\mathbf{Z}L, \mathbf{Z}) \to \mathbf{Z}L \otimes_G IG \to \mathbf{Z}L \otimes_G \mathbf{Z}G \to \mathbf{Z}L \otimes_G \mathbf{Z} \to 0.$$

Here the last two terms just represent the augmentation of $\mathbf{Z}L$, so the kernel in the third term is $IL$. For the first term we have, by Th. 3.6.9 and Prop. 1.3, $\mathrm{Tor}_1^G(\mathbf{Z}L, \mathbf{Z}) \cong \mathrm{Tor}_1^N(\mathbf{Z}, \mathbf{Z}) \cong N^{\mathrm{ab}}$. Hence we obtain the exact sequence of $L$-modules

$$0 \to N^{\mathrm{ab}} \to \mathbf{Z}L \otimes_G IG \to IL \to 0. \tag{33}$$

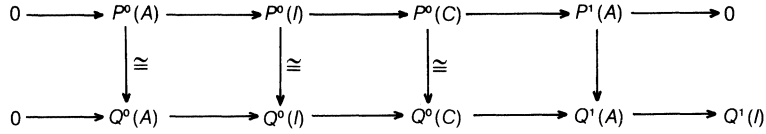
Now from the associativity of the tensor product we know that for any $L$-module $A$,

$$\mathrm{Hom}_L(ZL \otimes_G IG, A) \cong \mathrm{Hom}_G(IG, A).$$

Let us write the two sides of this formula as $P(A) \cong Q(A)$ and write $P^n, Q^n$ for the left derived functors. We take a short resolution of $A$ with $I$ injective:

$$0 \to A \to I \to C \to 0,$$

and apply $P, Q$, recalling that $P^1$ vanishes on injectives:




It follows that there is an injection from $P^1(A)$ to $Q^1(A)$, i.e. from $\mathrm{Ext}_L^1(\mathbf{Z}L \otimes_G IG, A)$ to $\mathrm{Ext}_G^1(IG, A)$. If we now apply $\mathrm{Hom}_L(-, A)$ to (33), we find

$$0 \to \mathrm{Hom}_L(IL, A) \to \mathrm{Hom}_L(\mathbf{Z}L \otimes_G IG, A) \to \mathrm{Hom}_L(N^{\mathrm{ab}}, A) \to \mathrm{Ext}_L^1(IL, A) \to \mathrm{Ext}_L^1(\mathbf{Z}L \otimes_G IG, A).$$

This reduces to (32) if we use (23) on the first two terms, replace the last term by $\mathrm{Ext}_G^1(IG, A)$ and again use (23). ■

Occasionally the cohomology groups are needed for a more general coefficient ring. If $K$ is any commutative ring and $KG$ is the group algebra of $G$ over $K$, then the standard resolution $X \to \mathbf{Z} \to 0$ on tensoring with $K$ over $\mathbf{Z}$ becomes

$$X \otimes_{\mathbf{Z}} K \to K \to 0,$$

and this is still a projective resolution, because the original resolution was $\mathbf{Z}$-split, by the homotopy constructed earlier. Hence we obtain, for any $KG$-module $A$,

$$\mathrm{Ext}_{\mathbf{Z}G}^n(\mathbf{Z}, A) \cong \mathrm{Ext}_{KG}^n(KG \otimes_G \mathbf{Z}, A) \cong \mathrm{Ext}_{KG}^n(K, A). \tag{34}$$

Similarly we have

$$\mathrm{Tor}_n^{\mathbf{Z}G}(\mathbf{Z}, A) \cong \mathrm{Tor}_n^{KG}(K, A). \tag{35}$$

We recall that (34) vanishes for $n \geqslant 1$ if $A$ is coinduced, and (35) vanishes if $A$ is induced. But for $G$-modules over a finite group $G$, induced and coinduced mean the same thing, for if $f: K \to KG$ is the inclusion map, then there is an isomorphism of $KG$-modules:

$$\mathrm{Hom}_K(KG, U) \cong U \otimes KG,$$

for any $K$-module $U$, given by $\alpha \mapsto \sum s\alpha \otimes s^{-1}$.

## Exercises

(1) Supply the details for the proof of Th. 1.2.

(2) Examine the case of an unnormalized factor set.

(3) (O. Hölder) Let $E$ be an extension of $A$ by $B$, where $A, B$ are cyclic groups of orders

$m, n$ respectively (such a group $E$ is called *metacyclic*). Show that $E$ has a presentation

$$\mathrm{gp}\{a, b \mid a^m = 1, b^n = a^r, b^{-1}ab = a^s\},$$

where $s^n \equiv 1 \pmod{m}$ and $r(s-1) \equiv 0 \pmod{m}$. Conversely, given $m, n, r, s$ satisfying these relations, there is a metacyclic group of order $mn$ defined by this presentation.

(4) Show that an extension of $A$ by $G$ with factor set $\{m_{\alpha,\beta}\}$ splits iff $c_\alpha \in A$ exists such that $m_{\alpha,\beta} = c_{\alpha\beta}^{-1}(c_\alpha\theta_\beta)c_\beta$. Let $E$ be an extension of an abelian group $A$ by $G$. By adjoining free abelian generators $c_\alpha$ to $E$ and using the above relation to define $c_\alpha\theta_\beta$ show that $E$ can be embedded in a group $\bar{E}$ which is a semidirect product of $\bar{A}$ and $G$, where $\bar{A} \supseteq A$.

(5) Show that the group of isometries of Euclidean $n$-space is a split extension of the normal subgroup of translations by the orthogonal group.

(6) Show that $H^1(G, A)$ is in natural bijection with the isomorphism classes of *A-torsors*, i.e. left $A$-sets with regular $A$-action such that $(a \cdot x)^\alpha = a^\alpha \cdot x^\alpha (\alpha \in G, a \in A)$.

(7) Show that the augmentation ideal $IG$, defined as in (20), is $\mathbf{Z}$-free on the $s-1$ $(1 \neq s \in G)$. Verify that $\mathrm{Hom}_G(IG, A)$ is isomorphic to the group of 1-cocycles.

(8) Using the mapping $s \mapsto s - 1 \pmod{IG^2}$ of $G$ into $IG/(IG)^2$ show that $G^{\mathrm{ab}} \cong IG/(IG)^2$. Deduce another proof of Prop. 1.3.

(9) Let $G$ be a finite group of order $r$. Show that every cocycle in $H^n(G, \mathbf{C}^\times)$ is cohomologous to a cocycle whose values are $r$th roots of 1. Deduce that the multiplicator of $G$ has order at most $r^{r^2}$.

(10) Let $G = \mathbf{C}_m$ with generator $s$ and write $D = s-1$, $N = 1 + s + \cdots + s^{m-1}$ in $\mathbf{Z}G$. Verify that there is a free resolution $W \to \mathbf{Z} \to 0$, where $W_n = \mathbf{Z}G$ and $d_{2n}: W_{2n} \to W_{2n-1}$ is $d_{2n} = N$, while $d_{2n-1} = D$. With the notation ${}_NA = \{a \in A \mid aN = 0\}$ show that $H^{2n}(G, A) = A^G/AN$, $H^{2n-1}(G, A) = {}_NA/AD$ and $H_n(G, A) = H^{n-1}(G, A)$ for all $n > 1$.

## 4.2 The Frattini subgroup and the Fitting subgroup

Let $G$ be any group and consider $\Phi(G)$, the intersection of all its maximal (proper) subgroups; of course if there are no such subgroups, then $\Phi(G) = G$. It is clear that $\Phi(G)$ is a subgroup of $G$, which is *characteristic*, i.e. invariant under all automorphisms of $G$. It is called the *Frattini subgroup* (after G. Frattini, who introduced it in 1885). We recall from A.2, **2.2** that if $G$ is finitely generated and non-trivial, then it has maximal subgroups, hence in that case $\Phi(G) \subset G$; in particular this holds for every finite group.

There is another characterization of $\Phi(G)$ which is often useful. In any group $G$, an element $c$ will be called a *non-generator* if any subset of $G$ which together with $c$ generates $G$ is itself a generating set of $G$.

PROPOSITION 2.1 *Let $G$ be any group. Then $\Phi(G)$ consists precisely of all the non-generators of $G$.*

*Proof.* Let $c \in G$ be a non-generator. If $M$ is a maximal subgroup, then $c \in M$, for otherwise the subgroup generated by $M \cup \{c\}$ would be larger than $M$, hence equal to $G$, and so by definition of $c$ as a non-generator, $M = G$, which is false. Thus $c$ lies in every maximal subgroup, and hence in $\Phi(G)$.

Conversely, if $c \in \Phi(G)$ and $X \cup \{c\}$ generates $G$, we have to show that $X$ generates $G$. If this is not the case, then the subgroup $H$ generated by $X$ is proper in $G$, and hence (by Zorn's lemma) there is a subgroup $M$ containing $H$ but not $c$ and maximal with these properties. Since $c \notin M$, $M \neq G$, but any subgroup properly containing $M$ must also contain $c$ and hence be equal to $G$. Thus $M$ is a maximal subgroup; but $c \in \Phi(G) \subseteq M$, which is a contradiction. Hence $X$ generates $G$ and $c$ is indeed a non-generator. ■

The property of elements described in Prop. 2.1 extends in the case of finitely generated groups to subgroups, as the next result shows.

PROPOSITION 2.2 *Let $G$ be a group and $U$ a subgroup. Then $U \subseteq \Phi(G)$ whenever $U$ satisfies the following condition:*

$$\text{for any proper subgroup } H \text{ of } G,\ \mathrm{gp}\{U, H\} \subset G. \tag{1}$$

*When $G$ is finitely generated, this condition is necessary as well as sufficient. In particular, we have $H\Phi(G) \subset G$ for every proper subgroup $H$ of $G$.*

*Proof.* If (1) holds, then in particular, for any maximal subgroup $M$ of $G$, $M \subseteq \mathrm{gp}\{U, M\} \subset G$, hence $M = \mathrm{gp}\{U, M\}$ and so $U \subseteq M$; therefore $U \subseteq \Phi(G)$. Now assume that $G$ is finitely generated; if $U \subseteq \Phi(G)$, and $H$ is a proper subgroup of $G$, then there is a maximal subgroup $M$ containing $H$; since also $U \subseteq M$, we have $\mathrm{gp}\{U, H\} \subseteq M \subset G$, and (1) is satisfied. ■

Let $G$ be a group and $H$ a subgroup. For any $x \in G$ we write $H^x = x^{-1}Hx$ and recall that the *normalizer* of $H$ in $G$ is defined as the largest subgroup of $G$ in which $H$ is normal. In symbols,

$$N_G(H) = \{x \in G \mid H^x = H\}.$$

We also recall some results on nilpotent groups from A.1. There the *upper central series* $Z_i(G)$ of a group $G$ was defined as follows. $Z_0(G) = 1$; if $Z_{i-1}(G)$ is defined as a normal subgroup of $G$, then $Z_i(G)$ is defined by the condition that $Z_i(G)/Z_{i-1}(G)$ is the centre of $G/Z_{i-1}(G)$. In particular, $Z_1(G)$, also written $Z(G)$, is the centre of $G$. So we have a series of characteristic subgroups

$$1 = Z_0(G) \subseteq Z_1(G) \subseteq \cdots \subseteq Z_n(G) \subseteq \cdots,$$

which is *central* in the sense that (mod $Z_{i-1}(G)$), $G$ centralizes $Z_i(G)$; moreover, in any other central series the $i$th term is contained in $Z_i(G)$, which justifies the term 'upper'. If $Z_r(G) = G$ for some $r$, $G$ is called *nilpotent* and the least such $r$ is the *nilpotency class* of $G$. We recall from A.1, **9.8** that a finite group $G$ is nilpotent iff $G/Z(G)$ is nilpotent, iff every Sylow subgroup of $G$ is normal in $G$.

The following result is often referred to as the *Frattini argument*:

LEMMA 2.3 *Let $G$ be a finite group, $N \lhd G$ and $P$ a Sylow subgroup of $N$. Then*

$$G = N \cdot N_G(P). \tag{2}$$

*Proof.* For any $x \in G$, $P^x \subseteq N^x = N$, hence $P^x$ is again a Sylow subgroup of $N$ of the same order as $P$, and hence conjugate to $P$ in $N$, by Sylow's theorem. Thus $P^x = P^y$ for $y \in N$, hence $z = xy^{-1} \in N_G(P)$; now $x = zy$ and (2) follows. ■

We shall also need another property of $\Phi$.

LEMMA 2.4 *Let $G$ be any group and $N$ a finitely generated normal subgroup. Then $\Phi(N) \subseteq \Phi(G)$.*

*Proof.* $\Phi(N)$ is a characteristic subgroup of $N$, hence normal in $G$. Now let $M$ be a maximal subgroup of $G$ and suppose that $\Phi(N) \nsubseteq M$. Then

$$\Phi(N)M = G, \tag{3}$$

and by the modular law,

$$N = \Phi(N)M \cap N = \Phi(N)(M \cap N).$$

By Prop. 2.2, $M \cap N = N$, i.e. $N \subseteq M$ and so $\Phi(N) \subseteq M$. But this contradicts (3). Hence $\Phi(N) \subseteq M$ for any maximal $M$, and therefore $\Phi(N) \subseteq \Phi(G)$. ■

We remark that without the normality of $N$ this result need not hold, as the case of a simple group and a cyclic subgroup shows.

LEMMA 2.5 *Let $G$ be a finite group and $N, K$ normal subgroups such that $N \subseteq K$, $N \subseteq \Phi(G)$. If $K/N$ is nilpotent, then so is $K$.*

*Proof.* Let $P$ be a Sylow $p$-subgroup of $K$. Then $PN/N$ is a Sylow $p$-subgroup of $K/N$, because $PN/N \cong P/(P \cap N)$ and this again has $p$-power order. Hence $PN/N$ is normal in $K/N$, therefore characteristic in $K/N$ and so normal in $G/N$. Hence $PN \lhd G$. Now $P$ is a Sylow $p$-subgroup of $PN$; hence by Lemma 2.3,

$$G = PN \cdot N_G(P) = N \cdot N_G(P).$$

Now $N \subseteq \Phi(G)$ by hypothesis, hence by Prop. 2.2, $N_G(P) = G$, and so $P \lhd G$. Thus $P \lhd K$ and this holds for all Sylow subgroups of $K$, hence $K$ is nilpotent, as claimed. ■

We can now establish a useful criterion for nilpotency in terms of $\Phi$.

PROPOSITION 2.6 (Wielandt) *Let $G$ be a finite group and $N$ a normal subgroup of $G$. Then $N$ is nilpotent if and only if $N' \subseteq \Phi(G)$.*

*Proof.* If $N' \subseteq \Phi(G)$, then $N' \lhd G$ and $N/N'$ is abelian, hence nilpotent, and so by Lemma 2.5, $N$ is nilpotent. Conversely, if $N$ is nilpotent, then every proper subgroup of $N$ is properly contained in its normalizer, hence every maximal subgroup of $N$ is normal in $N$, therefore of prime index and so it contains $N'$. This means that $N' \subseteq \Phi(N)$; by Lemma 2.4, $\Phi(N) \subseteq \Phi(G)$, hence $N' \subseteq \Phi(G)$. ■

This result has several useful consequences. In the first place, since $\Phi(G)$ is normal in $G$ and $\Phi' \subseteq \Phi$, we have

THEOREM 2.7 (Frattini) *The Frattini subgroup of a finite group is nilpotent.* ■

Secondly, taking $N = G$, we obtain Wielandt's characterization of finite nilpotent groups:

THEOREM 2.8 *A finite group $G$ is nilpotent if and only if $G' \subseteq \Phi(G)$.* ■

The Frattini subgroup may be thought of as a kind of radical, in analogy with the Jacobson radical. But whereas the latter is the largest of all the commonly used radicals in a ring, this is not so for the Frattini subgroup. Radicals are not as useful for groups as for rings; nevertheless there is one, first introduced by H. Fitting in 1938, which plays a role in group theory. To prove its existence, we shall need another property of nilpotent groups. In the proof we shall use the commutator notation: $(x, y) = x^{-1}y^{-1}xy$, while for subgroups $H, K, (H, K)$ denotes the subgroup generated by all commutators $(h, k)$, where $h \in H, k \in K$.

LEMMA 2.9 *If $G$ is nilpotent and $1 \neq K \lhd G$, then $K \cap Z(G) \neq 1$.*

*Proof.* We shall write $Z_i = Z_i(G), Z = Z_1$ and use induction on the class of $G$. If $K \cap Z(G) = 1$, consider the quotient $\bar{G} = G/Z(G)$; by the nilpotence of $G$ this has smaller class and it has the normal subgroup $\bar{K} = KZ(G)/Z(G) \cong K \neq 1$. By induction we have $\bar{K} \cap Z(\bar{G}) \neq 1$; thus $KZ(G) \cap Z_2(G) \supset Z(G)$, so there exists $a \in Z_2(G) \backslash Z$, such that $a = bc$, where $b \in K, c \in Z$. Either $b \in Z$ and $K \cap Z \neq 1$,

because $b \neq 1$ (for if $b = 1$, then $a = c \in Z$). Or $b \notin Z$; then $b = ac^{-1} \in Z_2(G)$ and for suitable $g \in G$, $1 \neq (b, g) \in Z$. Now $(b, g) \in K \cap Z$. ■

THEOREM 2.10 *Let $G$ be a finite group and denote by $F(G)$ the subgroup generated by all nilpotent normal subgroups of $G$. Then $F(G)$ is a nilpotent normal subgroup of $G$ which contains every nilpotent normal subgroup.*

$F(G)$ is called the *Fitting subgroup* of $G$.

*Proof.* Let $H, K$ be nilpotent normal subgroups of $G$. Clearly $HK \lhd G$; we claim that $HK$ is nilpotent, and here we may assume that $H, K \neq 1$ and use induction on $|G|$. Write $Z = Z(H)$; then $Z \neq 1$. If $N = (Z, K) = 1$, then $Z$ is in the centre of $HK$ and $HK/Z \cong (H/Z)(ZK/Z)$ is a product of nilpotent normal subgroups of $HK/Z$, hence nilpotent by the induction hypothesis, and so $HK$ is nilpotent. If $N \neq 1$, then $N \lhd K \cap Z$, hence $T = N \cap Z(K) \neq 1$, by Lemma 2.9, and we can repeat the argument with $N$ replaced by $T$. Thus in either case $HK$ is nilpotent.

Now let $F$ be a nilpotent normal subgroup of $G$ of maximal order. Then for any nilpotent normal subgroup $H, FH$ is again nilpotent and normal and $FH \supseteq F$, hence $FH = F$ by the maximality of $F$, so $H \subseteq F$, as we had to show. ■

Finally we establish a connexion between the Frattini and the Fitting subgroups.

THEOREM 2.11 *Let $G$ be a finite group, $\Phi = \Phi(G)$ its Frattini subgroup and $F = F(G)$ its Fitting subgroup. Then*

$$F' \subseteq \Phi \subseteq F, \tag{4}$$

*and*

$$F(G/\Phi) = F/\Phi. \tag{5}$$

*Proof.* (4) is clear from Prop. 2.6, Th. 2.7 and the definition of $F$. To prove (5), let us write $\bar{G} = G/\Phi$, $\bar{F} = F(\bar{G})$, and denote by $K$ the inverse image of $\bar{F}$ in $G$. Now take a Sylow $p$-subgroup $P \neq 1$ of $K$; its image $\bar{P}$ in $\bar{G}$ is a Sylow $p$-subgroup of $\bar{F}$ and since $\bar{F}$ is nilpotent, $\bar{P}$ is normal in $\bar{F}$, hence characteristic and so normal in $\bar{G}$. Therefore $P\Phi \lhd G$; we claim that $\Phi N_G(P) = G$. For $P$ is a Sylow $p$-subgroup of $L = P\Phi$; hence $G = LN_G(P)$ by Lemma 2.3, i.e. $G = \Phi N_G(P)$, hence $G = N_G(P)$ by Prop. 2.2, and so $P \lhd G$. Thus each Sylow subgroup of $K$ is normal and it follows that $K$ is nilpotent, so $K \subseteq F$. But the image of $F$ in $\bar{G}$ is certainly nilpotent and so is contained in $\bar{F}$. Therefore $F = K$; in particular, $\bar{F} = F/\Phi$. ■

We can use $\Phi$ to estimate the number of automorphisms of a finite group:

THEOREM 2.12 (P. Hall) *Let $G$ be a finite group. If $G$ can be generated by $d$ elements, then*

$$|\mathrm{Aut}(G)| \text{ divides } |\mathrm{Aut}(G/\Phi(G))| \cdot |\Phi(G)|^d. \tag{6}$$

*Proof.* Since $\Phi = \Phi(G)$ is a characteristic subgroup, each automorphism $\alpha$ of $G$ induces an automorphism $\bar{\alpha}$ of $G/\Phi$, and the correspondence $\alpha \mapsto \bar{\alpha}$ clearly is a homomorphism. Let $K$ be its kernel, i.e. $\alpha \in K$ iff $x^\alpha x^{-1} \in \Phi$ for all $x \in G$.

Let $\{u_1, \ldots, u_d\}$ be a generating system for $G$. The number of families $\{u_1 a_1, \ldots, u_d a_d\}$, where $a_i \in \Phi$, is $|\Phi|^d$ and each again generates $G$. The set $U$ of all these generating systems is permuted by $K$, and since any automorphism leaving $\{u_1, \ldots, u_d\}$ fixed must be the identity (because this is a generating set), it follows that the stabilizer of any member of $U$ is 1. Hence each orbit has $|K|$ members and if there are $r$ orbits, we have

$$r|K| = |\Phi|^d. \tag{7}$$

Now $|\mathrm{Aut}(G)| = |\mathrm{Aut}(G/\Phi(G))| \cdot |K|$, hence (6) follows. ■

## Exercises

(1) Show that $\Phi(\mathbf{Q}) = \mathbf{Q}$.

(2) Show that if $N \lhd G$, then $\Phi(G/N) \geqslant \Phi(G)N/N$, but that equality need not hold. (Hint: try $N = \Phi(G)$.)

(3) For a finite group $G$ show that if $G/\Phi$ is nilpotent, then so is $G$.

(4) Show that for any nilpotent group $G$ (not necessarily finite), $G' \subseteq \Phi(G)$. (Hint: use Prop. 2.2.)

(5) Show that a finite group with a single maximal subgroup is cyclic of prime power order.

(6) Show that for any $p$-group $G$, $\Phi(G) = G'G^p$, where $G^p = \mathrm{gp}\{x^p | x \in G\}$, while for $p = 2$, $\Phi(G) = G^2$.

(7) If $P$ is a finite $p$-group, then any subset $X$ generates $P$ iff its residues mod $\Phi(P)$ span the $\mathbf{F}_p$-space $P/\Phi(P)$. Deduce that all minimal generating sets have the same number of elements and that every element not in $\Phi(P)$ is contained in a minimal generating set.

(8) Let $G$ be a finite group and $A$ an abelian normal subgroup. Show that if $A \cap \Phi(G) = 1$, then $A$ has a complement in $G$.

(9) Show that in a faithful primitive permutation group any nilpotent normal subgroup is abelian.

(10) Let $G$ be a finite group and $\alpha$ an automorphism of $G$ of order prime to $|\Phi(G)|$ and inducing the identity on $G/\Phi(G)$. Show that $\alpha = 1$.

(11) Let $G$ be a group of order $p^n$ and $(G:\Phi(G)) = p^d$. Show that $\mathrm{Aut}(G)$ has order dividing $p^{d(n-d)}(p^d - 1)(p^d - p)\cdots(p^d - p^{d-1})$.

(12) Show that any non-abelian group of order 8 has a cyclic subgroup of index 2 and hence is either the dihedral or the quaternion group (cf. Ex. (4) of **9.9**, A.1, or **7.4** below).

## 4.3 Hall subgroups

One of the first results in group theory is Lagrange's theorem, which tells us that the order of a subgroup of $G$ is a divisor of the order of $G$. One soon discovers that the converse is false, for example $\mathrm{Alt}_4$ has order 12 but contains no subgroup of order 6. The first positive result in this direction was Sylow's theorem, which showed that for prime powers the converse of Lagrange's theorem is true. A significant generalization was found by P. Hall, who showed in 1928 that in a soluble group $G$, for every factorization of the order of $G$ into two coprime integers, $|G| = mn$, $(m, n) = 1$, there is a subgroup of order $m$, and in 1937 he showed that solubility is necessary for this to happen.

It is convenient to begin with some notation. Let $\pi$ be a set of prime numbers. By the *$\pi$-part* of an integer $n = \prod p^{\alpha_p}$ we understand the number $n_\pi = \prod_{p\in\pi} p^{\alpha_p}$. The complementary set of primes is written $\pi'$, so for any positive integer $n$, we have

$$n = n_\pi \cdot n_{\pi'}.$$

For any finite group $G$, $\pi(G)$ denotes the set of primes dividing $|G|$. If $\pi(G) \subseteq \pi$, the group $G$ is called a $\pi$-group. A *Hall subgroup* is a subgroup of $G$ whose order is prime to its index. A $\pi$-subgroup of $G$ whose index is prime to $\pi$ (i.e. not divisible by any prime in $\pi$) is called a *Hall $\pi$-subgroup*; e.g. when $\pi = \{p\}$, a Hall $p$-subgroup of $G$ is just a Sylow $p$-subgroup. A Hall $p'$-subgroup is also called a *$p$-complement*. To establish the existence of Hall subgroups in soluble groups we shall need some preliminary results.

LEMMA 3.1 *Let $G$ be a finite group and $H, K$ any subgroups. Then*

$$(HK:H) = (K:H \cap K). \tag{1}$$

*In particular, if $(G:H) = (K:H \cap K)$, then $HK = G$.*

*Proof.* Clearly $HK$ is a union of cosets of $H$, say $HK = Hk_1 \cup \cdots \cup Hk_r$, where $k_i \in K$ and $r = (HK:H)$. We claim that, writing $D = H \cap K$, we have the coset decomposition

$$K = Dk_1 \cup \cdots \cup Dk_r; \tag{2}$$

this will establish (1). Any $x \in K$ can by hypothesis be written as $x = hk_i$ where $h \in H$. Here $h = xk_i^{-1} \in H \cap K = D$; if $hk_i = h'k_j$, then $k_i k_j^{-1} \in H$ and it follows that $i = j$. This proves the coset decomposition (2), and (1) follows. Now if $(G:H) = (K:H \cap K)$, then $(G:H) = (HK:H)$, hence $G$ and $HK$ contain the same number of cosets of $H$ and so $HK = G$. ■

We note that in the special case where $H$ or $K$ is normal in $G$, the lemma follows from the second isomorphism theorem (Th. 1.2.6).

LEMMA 3.2 *If $H, K$ are subgroups of a finite group $G$ whose indices in $G$ are coprime, then $HK = G$ and $(G:H \cap K) = (G:H)(G:K)$.*

*Proof.* Put $(G:H) = m$, $(G:K) = n$; we have, by Lemma 3.1,

$$(G:H \cap K) = (G:K)(K:H \cap K) = (G:K)(HK:H).$$

Clearly $(HK:H)$ is a factor of $m = (G:H)$, hence

$$(G:H \cap K) = nm_1, \qquad \text{where } m_1 | m.$$

Similarly,

$$(G:H \cap K) = mn_1, \qquad \text{where } n_1 | n.$$

Hence $m_1 n = n_1 m$, so $n/n_1 = m/m_1 = 1$, because $m, n$ are coprime, and it follows that

$$(G:H \cap K) = mn.$$

Moreover, $(G:H) = m = m_1 = (HK:H)$, hence $HK = G$, by Lemma 3.1. ■

We shall need the result that a normal Hall subgroup always has a complement. When the normal subgroup is abelian, this follows from Cor. 1.7; the general case is deduced from this by an induction on the order.

THEOREM 3.3 (Schur–Zassenhaus) *Let $G, H$ be two finite groups. If the orders of $G, H$ are coprime, then every extension of $H$ by $G$ splits. Moreover, if either $G$ or $H$ is soluble, then any two complements of $H$ in an extension are conjugate.*

*Proof.* Write $|G| = n$, $|H| = r$, so that $(n, r) = 1$, and consider an extension

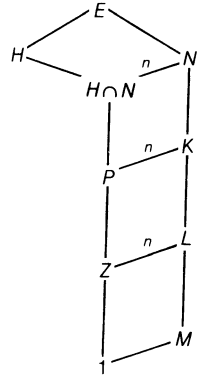
$$1 \to H \to E \xrightarrow{\lambda} G \to 1.$$

If $H$ is abelian, the extension is represented by an element in $H^2(G, H)$ and this group is 0 by Cor. 1.7, so the extension splits. In the general case we use induction on $r$. Clearly we need only find a subgroup $K$ of order $n$ in $E$; this must be a

complement to $H$, because $H \cap K = 1$ by comparing orders, and so $\lambda | K$ maps onto $G$.

Take a prime divisor $p$ of $r$ and let $P$ be a Sylow $p$-subgroup of $E$; if its normalizer is $N = N_E(P)$, then $N_H(P) = H \cap N$ and the number of conjugates of $P$ in $N$ is

$$(H : H \cap N) = (E : N).$$




It follows that $n = (E:H) = (N : N \cap H)$. Now $P$ and $H \cap N$ are normal in $N$ and $N/P$ is an extension of $(H \cap N)/P$ by $N/(H \cap N)$, whose orders divide $r$ and $n$ respectively, and so are coprime. Since $P \neq 1$, we can apply the induction hypothesis and find a subgroup $K$ of $N$ such that $P \subset K \subset N$ and $(K:P) = n$.

Let $Z$ be the centre of $P$; this is a non-trivial characteristic subgroup of $P$, hence $Z \lhd N$ and so $Z \lhd K$. By induction we obtain a subgroup $L$ of $K$ such that $Z \subset L \subset K$ and $(L:Z) = n$. Now $L$ is an extension of the abelian group $Z$ and the orders are again coprime, so we have a subgroup $M$ of $L$ of order $n$, and this is the required complement.

If an extension $E$ has two complements $G_1, G_2$ of $H$, suppose first that $H$ is abelian. An element $x$ of $G$ corresponds to $xf_1, xf_2 = xf_1 \cdot c_x$ of $G_1, G_2$, where $c_x$ is a cocycle of $G$ in $H$. Since $H^1(G, H) = 0$, we have $c_x = u^{-x}u$ for some $u \in H$, hence $xf_2 = xf_1 \cdot u^{-x}u = u^{-1}xf_1 u$ and so $G_2$ is conjugate to $G_1$, as claimed.

Next assume that $H$ is soluble and let $G_1, G_2$ be two complements. Applying the previous case to $E/H'$ we find that $G_2 H' = (G_1 H')^u$ for some $u \in H$, i.e. $G_1^u H' = G_2 H'$. By induction on the derived length of $H$ there exists $v \in H'$ such that $G_1^{uv} = G_2$ and the result follows.

Finally assume that $G$ is soluble and let $G_1, G_2$ again be complements of $H$ in $E$. If $G$ is a $p$-group, then $G_1, G_2$ are Sylow $p$-subgroups of $E$ and hence are conjugate by Sylow's theorem. In the general case let $M_1$ be a minimal normal subgroup of $G_1$ and $M_2$ the corresponding subgroup of $G_2$. Then $M_1 H = M_2 H$ and by induction there exists $x \in H$ such that $M_1^x = M_2$. Replacing $G_1$ by $G_1^x$ we may thus assume that $M_1 = M_2$. Now $G_1 H / M_1 = G_2 H / M_2$ and by the induction hypothesis there exists $y \in H$ such that $G_1^y / M_1 = G_2 / M_2$; hence $G_1^y = G_2$ as we had to show. ■

We remark that by the Feit–Thompson theorem every group of odd order is soluble. This shows that the hypothesis of Th. 3.3 is always satisfied, since of two coprime integers at least one must be odd.

Let $\pi$ be any set of primes and $G$ a finite group. It is clear that if $H$ is a Hall $\pi$-subgroup of $G$, then $H$ is also a Hall $\pi$-subgroup of any subgroup of $G$ containing $H$, and under any homomorphism $f$ from $G$, the image $Hf$ is a Hall $\pi$-subgroup of $Gf$. Further, if $p, q$ are distinct primes, $H$ is a Hall $p'$-subgroup and $K$ a Hall $q'$-subgroup of $G$, then $|HK|$ is divisible by $|G|_{p'}$ and $|G|_{q'}$, hence by $|G|$ and so $HK = G$. Moreover, $H \cap K$ is a Hall $p'$-subgroup of $K$, a $q'$-subgroup of $H$ and a $\{p, q\}'$-subgroup of $G$. With these preparations we can establish the existence of Hall subgroups:

THEOREM 3.4 (P. Hall) *Let $G$ be a finite soluble group and $\pi$ a finite set of primes. Then any $\pi$-subgroup of $G$ is contained in a Hall $\pi$-subgroup, hence Hall $\pi$-subgroups exist for any $\pi$, and any two Hall $\pi$-subgroups are conjugate in $G$.*

*Proof.* We shall use induction on $|G|$. Let $G$ be a finite soluble group and $M$ a minimal normal subgroup of $G$. By A.1, **9.6**, Th. 5, $M$ is an elementary abelian $p$-group, for some prime $p$. Write $\bar{G} = G/M$; then $\bar{G}$ contains a Hall $\pi$-subgroup $\bar{H}$, by the induction hypothesis, where $H \supseteq M$. Moreover, if $A$ is a $\pi$-subgroup of $G$, its image $\bar{A}$ in $\bar{G}$ is a $\pi$-subgroup and so is contained in some conjugate of $\bar{H}$, say $\bar{A} \subseteq \bar{H}^x$; it follows that $A \subseteq H^x$. If $p \in \pi$, then $(G:H)_\pi = 1$ and $H$ is a $\pi$-subgroup, hence a Hall $\pi$-subgroup of $G$, so is $H^x$ and $A$ has been embedded in a Hall $\pi$-subgroup. Since 1 is always a $\pi$-subgroup, this shows that Hall $\pi$-subgroups exist. Moreover, all Hall $\pi$-subgroups of $G$ are conjugate, for if $A$ is a Hall $\pi$-subgroup and $A \subseteq H^x$, then $A = H^x$, because $|A| = |H|$.

There remains the case $p \notin \pi$. By Th. 3.3 there is a complement $K$ of $M$ in $H$ and all complements are conjugate. Since $|K| = |\bar{H}| = |G|_\pi$, $K$ is a Hall $\pi$-subgroup of $G$, and all the Hall $\pi$-subgroups are conjugate. If $A$ is any $\pi$-subgroup of $G$, we have as before $A \subseteq H^x$ for some $x \in G$. Either $H \subset G$; then by induction, $A$ is contained in a Hall $\pi$-subgroup of $H^x$, which is also a Hall $\pi$-subgroup of $G$. Or $H = G$; then $A \subset G = KM$, hence

$$AM = AM \cap KM = (AM \cap K)M.$$

Now $A$ and $AM \cap K$ are Hall subgroups of $AM$ and hence are conjugate, so for some $y \in M$, $A^y = AM \cap K \subseteq K$. It follows that $A$ is contained in a Hall $\pi$-subgroup of $G$. ■

For the converse we shall use an interesting solubility criterion due to H. Wielandt:

THEOREM 3.5 *Let $G$ be a finite group. If $G$ has three soluble subgroups whose indices are pairwise coprime, then $G$ is soluble.*

*Proof.* Let the subgroups be $H_1, H_2, H_3$; if $H_1 = 1$, then $|G| = (G:H_1)$ is prime to $(G:H_2)$, so the latter is 1 and $H_2 = G$ is soluble. Hence we may assume that $H_i \neq 1$. Let $M$ be a minimal normal subgroup of $H_1$; since $H_1$ is soluble, $M$ is an elementary abelian $p$-group, for some prime $p$. Now $p$ cannot divide the indices of both $H_2$ and $H_3$, say $p \nmid (G:H_2)$. Then $p \mid\mid H_2|$, hence $H_2$ contains a non-trivial Sylow $p$-subgroup $P$, which is also a Sylow $p$-subgroup of $G$. Let $P_1$ be a Sylow $p$-subgroup of $H_1$; then $P_1 \subseteq P^x$ for some $x \in G$. We may replace $H_2$ by $H_2^x$ without affecting the hypothesis; then $P_1 \subseteq P$ and $M \subseteq P$, because $M \lhd H_1$. Thus we have $M \subseteq H_1 \cap H_2$. Now by Lemma 3.2, $G = H_1 H_2$; hence any $x \in G$ has the form $x = x_1 x_2$, $x_i \in H_i$; therefore $M^x = M^{x_2} \subseteq H_2$. Hence $H_2$ contains the normal closure of $M$ in $G$: $K = M^G = \text{gp}\{M^y \mid y \in G\} \subseteq H_2$. Since $H_2$ is soluble, so is $K$. Further, the subgroups $KH_i/K$ of $G/K$ satisfy the hypothesis, so $G/K$ is soluble, hence so is $G$. ■

We can now complete the proof of Hall's criterion; we recall that a $p$-complement is a subgroup of $p$-power index and order prime to $p$; in particular, if $p \nmid |G|$, then $G$ itself is the only $p$-complement.

THEOREM 3.6 (P. Hall) *Any finite group $G$ is soluble if and only if it contains a $p$-complement for each prime $p$.*

*Proof.* For soluble $G$ the result follows by Th. 3.4. Now assume that $|G| = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, where the $p_i$ are distinct primes, and that $G$ has a $p_i$-complement $H_i$ for $i = 1, \ldots, r$. If $r = 1$ or 2, $G$ is soluble by Burnside's $p^\alpha q^\beta$-theorem (A.2, Th. 7.8.3), hence we may assume $r \geqslant 3$. We claim that each $H_i$ is soluble. Clearly $(G:H_i) = p_i^{\alpha_i}$, hence by Lemma 3.2, $H_1 \cap H_i$ is a Hall $\pi_i$-subgroup of $G$, where $\pi_i = \pi(G) \setminus \{p_1, p_i\}$, and it follows that $H_1 \cap H_i$ is a $p_i$-complement of $H_1$. By induction $H_1$ is soluble, similarly for $H_2, \ldots, H_r$; now we apply Th. 3.5 to complete the proof. ■

## Exercises

(1) By a *Hall system* in a finite group $G$ is meant a family $\Sigma$ of Hall subgroups of $G$, one of order $d$ for each divisor $d$ of $|G|$ prime to $|G|/d$, such that each $H \in \Sigma$ is the intersection of all subgroups in $\Sigma$ whose orders are divisible by $|H|$. Verify that every family of $p$-complements, where $p$ runs over the prime divisors of $|G|$, gives rise to a Hall system, and that for any $H, K \in \Sigma$, $HK = KH \in \Sigma$.

(2) Find all Hall systems in $S_4$.

(3) Show that Lemma 3.1 holds for any group $G$ such that the subgroup $H$ is of finite index. Extend Lemma 3.2 similarly.

(4) Let $G$ be a finite group and $P_i$ a Sylow $p_i$-subgroup, for each prime divisor $p_i$ of $|G|$

$(i = 1, \ldots, r)$. Show that if $P_iP_j = P_jP_i$ for $i, j = 1, \ldots, r$, then $G$ is soluble and the different products of the $P_i$ constitute a Hall system. Show that $G$ is nilpotent iff it has a single Hall system.

(5) Let $G$ be a finite soluble group and $p$ a prime divisor of $|G|$. Show that the number of $p$-complements in $G$ is a power of $p$. Use Lemma 3.2 to deduce that all Hall systems in $G$ are conjugate.

(6) Let $G$ be a finite soluble group. Show that every subgroup containing the normalizer of a Hall subgroup is its own normalizer.

(7) By examining $\mathrm{Sym}_5$ show that Th. 3.5 does not remain true when the number of subgroups is reduced to 2.

## 4.4 The transfer

Let $G$ be a group and $H$ a subgroup of finite index $n$ say, and consider a fixed coset decomposition of $G$:

$$G = Hs_1 \cup Hs_2 \cup \cdots \cup Hs_n. \tag{1}$$

We can represent $G$ by matrices as follows: Each $a \in G$ permutes the cosets in (1) by right multiplication, thus there is a permutation $\sigma_a$ of $1, 2, \ldots, n$ such that $Hs_ia = Hs_{i\sigma_a}$. Hence $s_ias_{i\sigma_a}^{-1} \in H$ and we define an $n \times n$ matrix $\mu(a) = (\mu_{ij}(a))$ with entries in $H$ by the equation

$$\mu_{ij}(a) = \begin{cases} s_ias_{i\sigma_a}^{-1} & \text{if } j = i\sigma_a, \\ 0 & \text{otherwise.} \end{cases}$$

Thus $\mu(a)$ has a single non-zero entry in each row and one in each column; it is called a *monomial matrix* over $H$. To show that we have indeed a representation, take $a, b \in G$; we have $\sigma_{ab} = \sigma_a\sigma_b$ because $\sigma$ is the permutation representation of $G$ on the cosets (1). Hence we have $s_ias_{i\sigma_a}^{-1}s_{i\sigma_a}bs_{i\sigma_a\sigma_b}^{-1} = s_i(ab)s_{i\sigma_{ab}}^{-1} \in H$ and it follows that

$$\mu_{ik}(ab) = \begin{cases} \mu_{ii\sigma_a}(a)\mu_{i\sigma_ai\sigma_{ab}}(b) & \text{if } k = i\sigma_{ab}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus $\mu_{ik}(ab) = \sum_j \mu_{ij}(a)\mu_{jk}(b)$, and we have indeed a representation. Of course the representation still depends on the choice of transversal $(s_i)$; to free ourselves of this dependence, consider a homomorphism $f: H \to A$ to an abelian group (written multiplicatively) and put $\alpha_{ij}(a) = \mu_{ij}(a)f$. We claim that the mapping $V: G \to A$ defined by

$$aV = |\det \alpha(a)| = \prod_{i=1}^{n} (s_ias_{i\sigma_a}^{-1})f \tag{2}$$

is a homomorphism independent of the choice of transversal. For if $\bar{s}_i = h_i s_i$ is another transversal and we form the corresponding expression, we obtain

$$\prod_{i=1}^{n} (h_i s_i a s_{i\sigma_a}^{-1} h_{i\sigma_a}^{-1}) f = \prod_{i=1}^{n} (h_i f) \prod_{i=1}^{n} (s_i a s_{i\sigma_a}^{-1}) f \left( \prod_{i=1}^{n} h_i f \right)^{-1} = aV,$$

because $\sigma_a$ is a permutation of $1, \ldots, n$. Now the homomorphism property follows from the multiplication law for determinants. This mapping $V$ is called the *transfer* of $G$ into $A$ (German: Verlagerung). We remark that $\ker V \supseteq G'$, so that we can factor $V$ by the natural map from $G$ to $G/G'$, to obtain a homomorphism from $G/G'$ to $A$.

Let us examine the action of $a \in G$ on $G/H$ more closely. Suppose that $G/H$ consists of $r$ orbits of $n_i$ points ($i = 1, \ldots, r$) under the action of $\mathrm{gp}\{a\}$, so that $\sum_1^r n_i = n = (G:H)$. If we take our transversal in the form $t_i a^j$ ($i = 1, \ldots, r$, $j = 0, 1, \ldots, n_i - 1$), then the action on the $i$th orbit is represented by the cycle $(x_i, x_i a, \ldots, x_i a^{n_i - 1})$, and its contribution to the transfer is $x_i a^{n_i} x_i^{-1}$. Therefore we have

$$aV = \prod_{i=1}^{r} (x_i a^{n_i} x_i^{-1}) f, \qquad \text{where } x_i a^{n_i} x_i^{-1} \in H. \tag{3}$$

We shall apply the transfer to prove the existence of normal $p$-complements under suitable conditions. Here we need

LEMMA 4.1 *Let $G$ be a finite group, $P$ a Sylow $p$-subgroup of $G$ and $X$, $Y$ two subsets of $G$ normalized by $P$ and conjugate in $G$. Then $X$ and $Y$ are conjugate in $N_G(P)$.*

*Proof.* By hypothesis, $Y = X^a$ for some $a \in G$, and $X$, $Y$ are normalized by $P$, hence $Y = X^a$ is normalized by $P^a$. Thus $N = N_G(Y)$ contains $P$ and $P^a$; clearly they are Sylow subgroups of $N$ and by Sylow's theorem they are conjugate, say $P^{ab} = P$ for $b \in N$. Writing $c = ab$, we have $c \in N_G(P)$, hence $X^c = X^{ab} = Y^b = Y$, because $b \in N$. ■

THEOREM 4.2 (Burnside) *Let $G$ be a finite group and suppose that the Sylow $p$-subgroup $P$ of $G$ is contained in the centre of its normalizer. Then $P$ has a normal complement in $G$.*

*Proof.* By hypothesis $P$ is abelian, so we can take the transfer $V: G \to P$. For any $u \in P$ and $n_i, x_i$ as in (3) (for $H = P$), $u^{n_i}$ and $x_i u^{n_i} x_i^{-1}$ lie in $P$, are normalized by $P$ and conjugate in $G$. Hence by the lemma they are conjugate in $N_G(P)$, hence equal (because $P$ is in the centre of $N_G(P)$), so we obtain

$$x_i u^{n_i} x_i^{-1} = u^{n_i}.$$

It follows that $uV = u^n$, where $n = (G:P)$. Since $p \nmid (G:P)$ and $P$ is an abelian

$p$-group, it follows that $V: G \to P$ is surjective and the kernel is of index $|P|$ and so is the desired complement. ■

COROLLARY 4.3 *Let $G$ be a finite non-abelian group with a non-trivial cyclic Sylow 2-subgroup. Then $G$ has a normal 2-complement and so cannot be simple.*

*Proof.* Let $P$, of order $2^\alpha$, be a Sylow 2-subgroup of $G$. Since it is cyclic, its automorphism group has order $\varphi(2^\alpha) = 2^{\alpha-1}$, therefore $(N_G(P):P)$ is a factor of $2^\alpha$, but it also divides $(G:P)$, which is odd, hence $N_G(P) = P$, so the hypothesis of Th. 4.2 is satisfied and we obtain a normal complement of $P$. ■

## Exercises

(1) In the definition (2) the sign of the determinant was ignored. Show that taking the sign into account only amounts to taking the sign of the permutation representation of $G$ on $G/H$.

(2) Show that the corestriction mapping $H_1(G, \mathbf{Z}) \to H_1(H, \mathbf{Z})$ (induced by the restriction from $G$ to $H$ ) is just the transfer.

(3) Show that if $G$ has an abelian Sylow $p$-subgroup $P$ with normalizer $N$, then $P \cap G' = P \cap N'$ and $P = (P \cap N') \times (P \cap Z(N))$. Show also that the maximal $p$-factor group of $G$ (i.e. the maximal quotient group of $G$ which is a $p$-group) is $\cong P \cap Z(N)$.

(4) Show that if a Sylow $p$-subgroup $P$ of $G$ has trivial intersection with any of its distinct conjugates, then any two elements of $P$ which are conjugate in $G$ are conjugate in $N_G(P)$.

(5) Let $P, Q$ be two distinct Sylow $p$-subgroups of $G$, chosen so that (for fixed $p$) $P \cap Q$ has maximal order. Show that the only conjugates of $P \cap Q$ in $G$ that are contained in $P$ are conjugate to $P \cap Q$ in $N_G(P)$.

## 4.5 Free groups

Since groups form a class of algebras defined by laws (**1.3**), it is clear what is to be understood by a free group on a given set. But it is a peculiarity of the laws defining groups that allows the elements of a free group to be written in an easily recognized form, which we shall now describe.

Let $X$ be any set. By a *group word* in $X$ we understand an expression

$$u_1 u_2 \cdots u_n, \qquad n \geqslant 0, \tag{1}$$

where each $u_i$ is either an element of $X$ or $u_i = x^{-1}$ for some $x \in X$. Formally we can regard the expressions (1) as the elements of the free monoid on $X \cup X^{-1}$, where we have put $X^{-1} = \{x^{-1} | x \in X\}$. For $n = 0$, (1) reduces to the empty word, written 1. We define an *elementary reduction* of a word $u_1 \cdots u_n$ as the process

of omitting a pair of neighbouring factors of the form $xx^{-1}$ or $x^{-1}x$, for some $x \in X$. The inverse process, inserting a factor $xx^{-1}$ or $x^{-1}x$ at some point, is an *elementary expansion.* A word is said to be *reduced* if it contains no pairs of neighbouring factors $xx^{-1}$ or $x^{-1}x$; thus a word admits an elementary reduction precisely if it is not reduced. Two group words in $X$ are said to be *equivalent* if we can transform one into the other by a series of elementary reductions or expansions.

It is clear that in any group $G$ generated by a set $X$, two equivalent group words represent the same element. When there are relations in $G$ between the group words in $X$, there will be inequivalent group words representing the same group element, but as we shall now see, in a free group inequivalent group words represent distinct group elements.

THEOREM 5.1 *Let $X$ be a set and $F$ the free group on $X$. Then every element of $F$ is represented by exactly one reduced group word in $X$ and two group words represent the same element of $F$ if and only if they are equivalent.*

*Proof.* It is clear from the definitions that every element of $F$ is represented by a group word in $X$, and that equivalent group words represent the same element. The multiplication of group words by juxtaposition is associative, and it is easily checked that the equivalence class of a product depends only on those of the factors, not on the factors themselves. Further, the empty word 1 acts as neutral under multiplication. Hence the set of equivalence classes forms a monoid under multiplication, and this is in fact a group, since $u_1u_2 \cdots u_n$ has the inverse $u_n^{-1} \cdots u_1^{-1}$, where $u_i^{-1} = x^{-1}$ if $u_i = x$, $u_i^{-1} = x$ if $u_i = x^{-1}$:

$$u_1 \cdots u_n u_n^{-1} \cdots u_1^{-1} \sim u_1 \cdots u_{n-1} u_{n-1}^{-1} \cdots u_1^{-1} \sim \cdots \sim 1,$$

by elementary reductions. It only remains to show that each group word is equivalent to exactly one reduced group word. Given a group word (1), we apply elementary reductions as often as possible; each such reduction reduces the length, so we arrive at a reduced form after a finite number of steps. In order to show that this form is independent of the order in which the reductions are made, we shall use the diamond lemma (Lemma 1.5.1). We have to show that if a word $f$ is reduced to $g_1, g_2$ by different elementary reductions, then there is a word $h$ which can be obtained by reduction from $g_1$ as well as $g_2$. There are two cases. (i) The terms being reduced do not overlap, say $u_iu_{i+1} = xx^{-1}$, $u_ju_{j+1} = yy^{-1}$, where $j > i+1$ and $x, y \in X \cup X^{-1}$, and of course $(x^{-1})^{-1} = x$. Clearly these reductions can be performed in either order and the outcome will be the same. (ii) The terms overlap; then a subword $uu^{-1}u$ or $u^{-1}uu^{-1}$ occurs. Taking $uu^{-1}u$, we can either reduce $uu^{-1}$ to 1 or $u^{-1}u$ to 1, and are left with $u$ in each case, so again the outcome is the same, and similar reasoning applies to $u^{-1}uu^{-1}$. Now it follows by Lemma 1.5.1 that we have a unique reduced form. ■

Here is another proof, not using the diamond lemma. The essential step is to show that distinct reduced words represent distinct group elements. We define $F$ as a permutation group on the set $W$ of all reduced group words as follows. Given $w = u_1 \cdots u_n \in W$ and $x \in X$, we put

$$w\alpha_x = \begin{cases} u_1 \cdots u_n x & \text{if } u_n \neq x^{-1} \text{ or } n = 0, \\ u_1 \cdots u_{n-1} & \text{if } u_n = x^{-1}. \end{cases}$$

$$w\beta_x = \begin{cases} u_1 \cdots u_n x^{-1} & \text{if } u_n \neq x \text{ or } n = 0, \\ u_1 \cdots u_{n-1} & \text{if } u_n = x. \end{cases}$$

It is easily checked that $\alpha_x$ is a permutation of $W$ with inverse $\beta_x$. Thus $F$ has been defined as a permutation group on $W$. Now, given any reduced word $u_1 \cdots u_n$, if we apply $\alpha_{u_1} \cdots \alpha_{u_n}$ (where $\alpha_{x^{-1}} = \beta_x$), to the empty word, we find $u_1 \cdots u_n$, hence distinct reduced words define distinct permutations of $W$ and so represent different elements of $F$. ■

Another proof of this theorem, using permutations of finite sets, was given in A.1, Th. 2 of **9.9**.

A free generating set in a free group will also be called a *basis*. Let $F$ be the free group with basis $X = \{x_1, \ldots, x_d\}$ and denote by $F^2$ the subgroup generated by all squares. Then $F/F^2$ is the elementary abelian 2-group generated by the images of $x_1, \ldots, x_d$, and hence is of order $2^d$. In particular this shows that the number $d$ of elements in a basis is independent of the choice of basis. It is called the *rank* of $F$, written $\mathrm{rk}\, F$, and the above argument shows that free groups of different ranks cannot be isomorphic (cf. also Ex. (1), Ch. 1).

We remark that by **1.3** every group $G$ can be written as a homomorphic image of a free group; more precisely, if $G$ can be generated by $d$ elements, then it can be expressed as a quotient of a free group of rank $d$.

We note that Th. 5.1 solves the word problem for free groups, for it provides a means of deciding when two group words (in a given free generating set) represent the same group element (cf. A.1, **9.9**). Let us now look at the conjugacy problem and show that in free groups this can be solved in a similar manner.

Two group words are said to be *cyclic conjugates* if $f = uv$, $g = vu$ for suitable words $u, v$. Thus $xy^{-1}z^{-1}xy$ and $xyxy^{-1}z^{-1}$ are cyclic conjugates. We remark that even when a word is in reduced form, it may have a cyclic conjugate which is not reduced, e.g. $x^{-1}yx$. A group word is said to be *cyclically reduced* if all its cyclic conjugates are reduced. Now we have

PROPOSITION 5.2 *Let $F$ be a free group on a set $X$. Then two elements of $F$ are conjugate if and only if their cyclically reduced forms are cyclic conjugates.*

It is clear that one can check in a finite number of steps whether two reduced

words are cyclic conjugates; for words of length $n$ we need only compare $n$ different words.

*Proof.* Let $f, g \in F$; if their reduced forms are cyclic conjugates, then $f = uv$, $g = vu$ and hence $g = u^{-1}fu$. Conversely, suppose that $f, g$ are conjugate. By passing to appropriate conjugates we may suppose that both $f$ and $g$ are in cyclically reduced form. Let

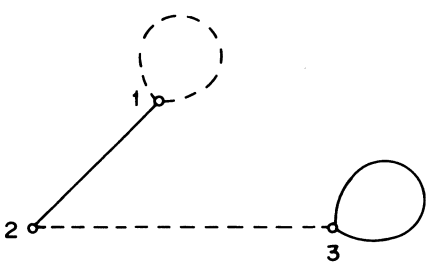
$$g = u^{-1}fu, \tag{2}$$

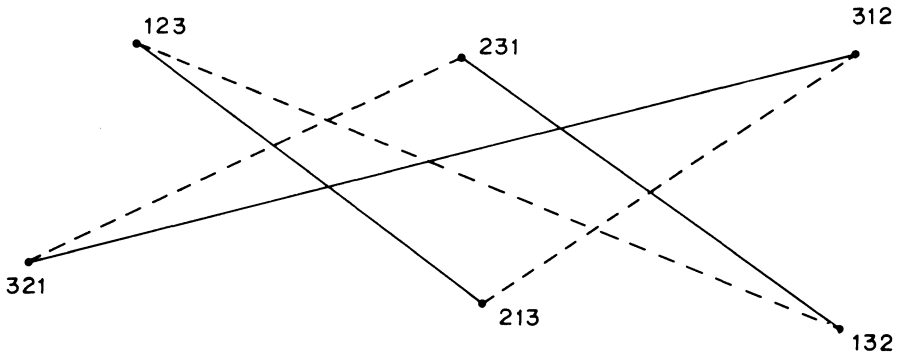
where $u$ is in reduced form. Since $g$ is cyclically reduced, the right-hand side cannot be in reduced form, say $u$ and $f$ begin with the same letter $x$. Then $f$ cannot end in $x^{-1}$ (because $f$ is cyclically reduced), so no cancellation can take place between $f$ and $u$, and the only way to reduce the right-hand side of (2) is by cancelling an initial portion of $u$ against that of $f$. So we have either $u = u_1u_2$, $f = u_1f_1$, hence $u^{-1}fu = u_2^{-1}f_1u_1u_2$; this is reduced and so must equal $g$; but the latter is cyclically reduced, so $u_2 = 1$, $u_1 = u$ and $f = uf_1, g = f_1u$, showing $f, g$ to be cyclic conjugates. Or we have $u = fu_2$; then $u^{-1}fu = u_2^{-1}fu_2$ and the argument shows as before that $u_2 = 1$, and it follows again that $f, g$ are cyclically conjugate. ■

We conclude this section by proving the Schreier subgroup theorem; this is an important result in its own right, while the proof illustrates some of the methods used in the study of free groups.

Given any group $G$ with generating set $X$ and a $G$-set $M$, we define its *diagram* $(G, X, M)$ as the graph whose vertices are the points of $M$, with an edge from $p$ to $q$ whenever $q = px$ for some $x \in X$. More precisely, the diagram may be regarded as a directed graph (digraph), in which each $x \in X$ is represented by a directed edge and $x^{-1}$ by its opposite.

For example, the symmetric group $S_3$ with generating set $\{(12), (23)\}$ acting on the set $\{1, 2, 3\}$ has the diagram shown below; the second diagram represents $S_3$ acting on the set of all arrangements of 1, 2, 3. In both cases continuous lines represent (12) and broken lines (23).






We note that the second diagram may be regarded as the diagram of the regular representation $(G, X, G)$; such a diagram contains all the information contained in the multiplication table, but in more accessible form. For example, we can from the diagram read off the relation $[(12)(23)]^3 = 1$, corresponding to a circuit based at 123.

It is easily seen that the diagram $(G, X, M)$ has a connected graph iff $G$ acts transitively on $M$. When this is so, we can take $M$ to consist of the cosets in $G$ of the stabilizer of a point, and from this point of view a diagram may be called a *coset diagram.* Any path in such a coset diagram $(G, X, G/H)$ represents an element of $G$ applied to a certain coset; in particular any element of $H$ may be represented by a circuit beginning and ending at the coset $H$. Since the graph is connected, it contains a subgraph on all the vertices which is a tree (A.2, Th. 1.4.3), i.e. a connected graph without circuits. Such a tree including all the vertices of our graph is also called a *spanning tree* for the graph. Let us express this in terms of our group. Given any coset $Hu$, there is a path in the coset diagram from $H$ to $Hu$ (where the orientation in forming the path is disregarded). Each such path corresponds to a group word $w$ on $X$ such that $Hw = Hu$. Now the choice of spanning tree in our graph amounts to the choice of a particular representative $w = w_1 \cdots w_r (w_i \in X \cup X^{-1})$ for our coset. It has the property that any left factor $w_1 \cdots w_i$ of $w$ is the chosen representative in its coset. A transversal with this property is called a *Schreier transversal.* So the choice of a spanning tree amounts to choosing a Schreier transversal, and it is not hard to verify that any Schreier transversal in turn leads to a spanning tree. Therefore the existence of a spanning tree assures us that every subgroup of $G$ has a Schreier transversal, relative to the given generating set of $G$. We note that neither $X$ nor $(G:H)$ need be finite for the spanning tree to exist.

Consider a coset diagram $(G, X, G/H)$ and choose a spanning tree $\Gamma$. Any edge $\alpha$ in our diagram, from $p$ to $q$ say, gives rise to a circuit as follows. Denote by $p_0$ the vertex corresponding to the coset $H$. There is a unique path $w_1$ within $\Gamma$ from $p_0$ to $p$, and a unique path $w_2$ from $q$ to $p_0$; hence $w_1 \alpha w_2$ is a circuit

on $p_0$. Here we have $w_2 = (w_1\alpha)^{-1}$ precisely if $\alpha \in \Gamma$, so our circuit is trivial (reduced to $p_0$) precisely if $\alpha \in \Gamma$. If $\alpha \notin \Gamma$, so that the circuit is non-trivial, it will be called a *basic* circuit, corresponding to $\alpha$. As we saw earlier, any circuit on $p_0$ corresponds to an element of $H$. We now observe that any circuit on $p_0$ can be written as a product of basic circuits. For, given edges $\alpha_1, \ldots, \alpha_r$ forming a circuit on $p_0$, suppose that $\alpha_i$ goes from $p_{i-1}$ to $p_i$, where $p_r = p_0$, and let $w_i$ be the path within $\Gamma$ from $p_0$ to $p_i$; in particular, $w_0$ is the empty path.

Then we have

$$\alpha_1 \cdots \alpha_r = \alpha_1 w_1^{-1} w_1 \alpha_2 w_2^{-1} \cdots w_{r-1}^{-1} w_{r-1} \alpha_r. \tag{3}$$

Here $w_{i-1}\alpha_i w_i^{-1}$ is either trivial and so can be omitted, or it is basic. This expresses our circuit as a product of basic circuits, and it shows that $H$ is generated by the elements corresponding to basic circuits. Thus we have

PROPOSITION 5.3 *Let $G$ be a finitely generated group. Then any subgroup of finite index is finitely generated; more precisely, if $G$ has a generating set of $d$ elements and $(G:H) = m$, then $H$ can be generated by $m(d-1)+1$ elements.*

It only remains to prove the last part. The coset diagram $(G, X, G/H)$ has $m$ vertices and $dm$ edges; by Th. 1.4.3 of A.2, a spanning tree has $m-1$ edges. We saw that $H$ has a generating set whose elements correspond to the non-tree edges, and their number is $dm - (m-1) = m(d-1)+1$. ■

By a slight refinement of the argument we obtain Schreier's subgroup theorem:

THEOREM 5.4 *Any subgroup of a free group is free. If $F$ is free of finite rank $d$ and $H$ is a subgroup of finite index $m$ in $F$, then $H$ has rank given by*

$$\operatorname{rk} H = (F:H)(\operatorname{rk} F - 1) + 1. \tag{4}$$

Equation (4) is known as *Schreier's formula* (cf. A.2, **11.5**).

*Proof.* We take the coset diagram of $(F, X, F/H)$ and choose a spanning tree $\Gamma$. As we saw, $H$ is generated by the elements corresponding to basic circuits; our aim is to show that these elements generate $H$ freely. For each non-tree edge $\alpha$ we have a basic circuit $\bar{\alpha}$ and it will be enough to show that if $\alpha_1 \cdots \alpha_r$ is a reduced word $\neq 1$ in the non-tree edges, then

$$\bar{\alpha}_1 \cdots \bar{\alpha}_r \neq 1, \tag{5}$$

where $\bar{\alpha}_i = w_{i-1}\alpha_i w_i^{-1}$ as before. If we write out the left-hand side of (5) as a word in $X$, we find that although there may be cancellation, this cannot involve the $\alpha_i$; they cannot cancel against any part of any $w_j$ because the $\alpha_i$ are not in $\Gamma$, and they cannot cancel against each other because the original word $\alpha_1 \cdots \alpha_r$ was reduced. Hence (5) follows and this shows the basic circuits to be a free

generating set of $H$. Now (4) follows as before by counting the non-tree edges. ■

## Exercises

(1) Give a direct proof that every abelian subgroup of a free group is cyclic.

(2) In a free group, if $w = u^n$, where $n \geqslant 1$, $u$ is called a *root* of $w$, *primitive* if $n$ is maximal for a given $w$. Show that every element of a free group has a unique primitive root. Show that two elements $u, v \neq 1$ have a common root iff they commute.

(3) Let $F$ be a free group. Show that every subgroup of finite index meets every subgroup $\neq 1$ non-trivially.

(4) Define a group $G$ to be *projective* (for this exercise only) if any homomorphism $G \to H_1$, where $H_1$ is a quotient of a group $H$, can be lifted to a homomorphism $G \to H$. Show that a group is projective iff it is free. (Note that this allows one to define free groups without reference to a basis.)

(5) Let $G$ be any group and $X$ a subset such that $X \cap X^{-1} = \varnothing$ and any product of elements of $X \cup X^{-1}$ with no factors $xx^{-1}$ or $x^{-1}x$ is $\neq 1$. Show that $X$ is a free generating set of the subgroup generated by it.

(6) Show that a free group of rank $d$ cannot be generated by fewer than $d$ elements.

(7) A group is called *Hopfian* if every homomorphism of $G$ onto $G$ is an isomorphism. Show that every free group of finite rank is Hopfian.

(8) Show that in a free group of rank 3 any subgroup of finite index has odd rank.

(9) In the free group on $x, y, z$ find a basis for the subgroup generated by all the squares.

(10) Show that in the free group on $x, y$, the elements $y^{-n}xy^n$ form a free generating set. Deduce that any non-cyclic free group contains a free subgroup of countable rank.

(11) Show that in any free group $F$ of rank $> 1$, the derived group $F'$ has countable rank.

(12) Let $G$ be a group with generating set $X$ and $H$ a subgroup. Show how to construct a Schreier transversal and verify that it corresponds to a spanning tree of the graph of $(G, X, G/H)$.

## 4.6 Commutators

The commutator of two elements or subgroups is a useful tool for studying group properties; below we give some of the basic formulae and simple applications.

Let $G$ be any group and $x, y \in G$. The *conjugate* of $x$ by $y$ is $x^y = y^{-1}xy$ and the *commutator* of $x$ and $y$ is $(x, y) = x^{-1}y^{-1}xy$. We note the following general formulae:

$$x^y = x(x, y) \qquad xy = yx(x, y), \tag{1}$$

$$(x, y)^{-1} = (y, x) = (x^{-1}, y)^x, \tag{2}$$

$$(xy, z) = (x, z)^y(y, z) = (x, z)((x, z), y)(y, z), \tag{3}$$

$$(x, yz) = (x, z)(x, y)^z = (x, z)(x, y)((x, y), z). \tag{4}$$

The proofs are straightforward; we have e.g. for (2): $(x^{-1}, y)^x = x^{-1}(xy^{-1}x^{-1}y)x = y^{-1}x^{-1}yx = (y, x)$, and for (4): $(x, yz) = x^{-1}z^{-1}y^{-1}xyz = x^{-1}z^{-1}xz \cdot z^{-1}x^{-1}y^{-1}xyz = (x, z)(x, y)^z$.

We further note

LEMMA 6.1 *Let $G$ be any group and $x, y \in G$.*

(i) *If $(x, y)$ commutes with $x$, then*

$$(x^n, y) = (x, y)^n \qquad \text{for all } n \in \mathbf{Z}. \tag{5}$$

(ii) *If $(x, y)$ commutes with $x$ and $y$, then*

$$(xy)^n = x^n y^n (y, x)^{\binom{n}{2}} \qquad \text{for all } n \geqslant 0. \tag{6}$$

*Proof.* (i) By (3) we have $(x^{n+1}, y) = (x, y)^{x^n}(x^n, y)$; since $x$ commutes with $(x, y)$, we can omit the exponent. Now (5) follows for $n \geqslant 0$ by induction; for negative $n$, say $n = -m$, we have $(x^{-m}, y) = (x, y)^{-m}$ by (2).

(ii) Clearly (6) holds for $n = 0$. If it holds for some $n \geqslant 0$, then on writing $m = \binom{n}{2}$, we have

$$\begin{aligned}(xy)^{n+1} &= (xy)^n(xy) = x^n y^n (y, x)^m (xy) \\ &= x^n y^n (xy)(y, x)^m \\ &= x^{n+1} y^n (y^n, x) y (y, x)^m \\ &= x^{n+1} y^n (y, x)^n y (y, x)^m \\ &= x^{n+1} y^{n+1} (y, x)^{m+n},\end{aligned}$$

which is the required formula, since $m + n = \binom{n}{2} + n = \binom{n+1}{2}$. ∎

We define higher commutators recursively by putting

$$(x_1, x_2, \ldots, x_n) = ((x_1, \ldots, x_{n-1}), x_n). \tag{7}$$

Thus a higher commutator within a single pair of brackets is understood to be

*left-normed*, i.e. all brackets begin on the left of the first argument. We remark that the identities between commutators correspond to identities in a Lie algebra, but this correspondence will not be pursued further here. For the moment we note the analogue of the Jacobi identity:

PROPOSITION 6.2 (Witt identity) *For any $x, y, z$ in a group $G$,*

$$(x, y^{-1}, z)^y (y, z^{-1}, x)^z (z, x^{-1}, y)^x = 1. \tag{8}$$

*Proof.* We have

$$(x, y^{-1}, z)^y = y^{-1}(yx^{-1}y^{-1}x)z^{-1}(x^{-1}yxy^{-1})zy = x^{-1}y^{-1}xz^{-1}x^{-1}\cdot yxy^{-1}zy.$$

This has the form $u^{-1}v$, where $u = xzx^{-1}yx$, $v = yxy^{-1}zy$. If we put $w = zyz^{-1}xz$, the left-hand side of (8) can be written $u^{-1}v\cdot v^{-1}w\cdot w^{-1}u$, which reduces to 1. ■

For any subgroups $H, K$ of a group $G$ we define the commutator subgroup $(H, K)$ as the subgroup generated by all $(x, y), x\in H, y\in K$. Thus the general element of $(H, K)$ is a product of commutators, but need not itself be a commutator. Higher commutator subgroups are again defined recursively as

$$(H_1, \ldots, H_n) = ((H_1, \ldots, H_{n-1}), H_n). \tag{9}$$

If each $H_i$ is normal in $G$, then this is the subgroup generated by all $(x_1, \ldots, x_n)$, where $x_i\in H_i$, but in general equality need not hold. We note the following obvious properties of commutator subgroups:

PROPOSITION 6.3 *If $H, K$ are subgroups of a group $G$, then*

(i) $(H, K) = (K, H)$,

(ii) $(H, K) \lhd \mathrm{gp}\{H, K\}$,

(iii) $(H, K) \lhd H \Leftrightarrow K \subseteq N_G(H)$,

(iv) *for any homomorphism $\varphi$, $(H, K)^\varphi = (H^\varphi, K^\varphi)$,*

(v) $(H, K) = 1$ *if and only if $H$ and $K$ commute elementwise.*

The proof may be left to the reader. ■

LEMMA 6.4 (P. Hall's three subgroup lemma) *Given three subgroups $A, B, C$ of a group, if $(A, B, C) = (B, C, A) = 1$, then $(C, A, B) = 1$.*

*Proof.* Let $x\in A, y\in B, z\in C$. By hypothesis, $(x, y^{-1}, z)^y = (y, z^{-1}, x)^z = 1$, hence Witt's identity shows that $(z, x^{-1}, y)^x = 1$, so $(z, x^{-1}, y) = 1$. Hence $(z, x^{-1})$ centralizes $B$, i.e. $((C, A), B) = 1$. ■

Commutators may be used to study nilpotent subgroups, which we

encountered in **4.2** and A.1, **9.8**. In a group $G$ a series of subgroups

$$G = H_1 \supset H_2 \supset \cdots \supset H_{r+1} = 1 \tag{10}$$

is called a *central series* if $(G, H_i) \subseteq H_{i+1}$. Since we then have $(G, H_i) \subseteq H_i$, it follows that each $H_i$ is normal in $G$. A group $G$ is nilpotent precisely if it has a finite central series (10). In particular, such a group has a *lower* and an *upper* central series from 1 to $G$ and $G$ to 1 respectively, defined as follows. Put $\gamma_1(G) = G$ and $\gamma_{i+1}(G) = (G, \gamma_i(G))$. Each $\gamma_i(G)$ is a fully invariant subgroup and the series

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

is called the *lower central series* of $G$; the definition shows that it is indeed a central series. Moreover, it has the property that for any central series (10) of $G$, we have

$$\gamma_i(G) \subseteq H_i. \tag{11}$$

For this holds when $i = 1$, and if it holds for a given $i \geqslant 1$, then $\gamma_{i+1}(G) = (G, \gamma_i(G)) \subseteq (G, H_i) \subseteq H_{i+1}$. In particular this shows that $G$ is nilpotent iff $\gamma_{r+1}(G) = 1$ for some $r \geqslant 0$. The least such $r$ is called the (precise) *class of nilpotence* of $G$.

We recall from **4.2** that the upper central series of $G$ is defined in a similar way. Namely, $Z_1(G)$ is the centre of $G$ and $Z_{i+1}(G)$ is defined as the subgroup of $G$ such that $Z_{i+1}(G)/Z_i(G)$ is the centre of $G/Z_i(G)$. Writing $Z_i = Z_i(G)$, we have

$$1 = Z_0 \subseteq Z_1 \subseteq \cdots$$

and if $G$ is nilpotent of precise class $r$, then

$$Z_i(G) \supseteq H_{r+1-i}. \tag{12}$$

For we clearly have $Z_1 \supset H_r$ and if (12) holds for some $i \geqslant 1$, then $Z_{i+1} = \{x \in G \mid (G, x) \subseteq Z_i\}$ and $(G, H_{r-i}) \subseteq H_{r+1-i} \subseteq Z_i$, therefore $H_{r-1} \subseteq Z_{i+1}$. This proves (12) generally.

We recall from A.1, Th. 5 of **9.8**, that a finite group is nilpotent iff it is the direct product of its Sylow subgroups. We also note that in any finite group $G$ with Sylow $p$-subgroup $P$, any subgroup $H$ containing $N_G(P)$ must be its own normalizer. For let $a \in G$ be such that $H^a = H$. Then $P$ and $P^a$ are Sylow $p$-subgroups of $H = H^a$, hence there exists $b \in H$ such that $P^a = P^b$, i.e. $c = ab^{-1} \in N_G(P)$, and so $a = cb \in H$; this shows that $N_G(H) = H$.

We now have the following characterizations of a nilpotent group:

THEOREM 6.5 *For any finite group $G$ the following conditions are equivalent:*

(a) *$G$ is nilpotent,*

(b) *$\gamma_{r+1}(G) = 1$ for some $r \geqslant 0$,*

(c) $Z_r(G) = G$ *for some* $r \geqslant 0$,
(d) *every proper subgroup is distinct from its normalizer,*
(e) *every maximal subgroup is normal in* $G$.
*Moreover, the least* $r$ *in* (*b*) *and* (*c*) *are the same and are equal to the precise class of nilpotence.*

*Proof.* We have already seen the equivalence of (a), (b), (c) and the assertion about the class of nilpotence. Now assume that $G$ is nilpotent with central series (10) and let $K$ be a proper subgroup, say $K \subseteq H_i, K \nsubseteq H_{i-1}$. Then $(G, H_{i-1}) \subseteq H_i \subseteq K$, hence $H_{i-1} \subseteq N_G(K)$ and so $K \subset N_G(K)$ and (d) follows. Clearly (d)$\Rightarrow$(e), so assume (e). If $P$ is any Sylow subgroup of $G$ and $P$ is not normal, then $N_G(P) \neq G$, hence there is a maximal subgroup $M \supseteq N_G(P)$ and $M$ is its own normalizer, but this contradicts (e). Thus all Sylow subgroups are normal, so $G$ is the direct product of its Sylow subgroups and hence is nilpotent. ■

Finally we shall prove that free groups are residually finite $p$-groups, for any prime $p$. We recall from **1.2** that this means: given any element $c \neq 1$ in a free group $F$, there exists a normal subgroup $N$ of $F$ not containing $c$, such that $F/N$ is a finite $p$-group. This will also show that for any prime $p, F$ can be expressed as a subdirect product of finite $p$-groups.

Let $F$ be the free group on $x_1, \ldots, x_d$ and consider a non-trivial word

$$x_{i_1}^{\alpha_1} \cdots x_{i_r}^{\alpha_r}, \qquad i_\rho \neq i_{\rho+1}, \qquad \alpha_\rho \neq 0, \qquad r \geqslant 1. \tag{13}$$

Let $m$ be a positive integer so large that $p^m$ does not divide $\alpha_1 \cdots \alpha_r$ and take $G$ to be the group of all upper unitriangular matrices of order $r+1$ over $\mathbf{Z}/p^m$, i.e. matrices of the form $I + N$, where $N = (n_{ij})$, $n_{ij} = 0$ for $i \geqslant j$. It follows that $N^{r+1} = 0$, hence $(I+N)^{p^t} = I + N^{p^t} = I$ for $p^t \geqslant r+1$; this shows $G$ to be a finite $p$-group. Our object will be to find a homomorphism $F \to G$ such that (13) is not mapped to 1. We map $x_i$ to $A_i$, where

$$A_i = \prod_k (I + e_{kk+1}),$$

where the product is taken over all $k$ such that $i_k = 1$. Then we have $e_k A_{i_k} = e_k + e_{k+1}$, where the $e_k$ are unit row vectors; hence

$$e_1 A_{i_1}^{\alpha_1} \cdots A_{i_r}^{\alpha_r} = e_1 + \alpha_1 \cdots \alpha_r e_{r+1} + \cdots$$

where dots indicate terms in $e_2, \ldots, e_r$. We have thus found a homomorphism of the required form; we remark that $G$ does not depend on the rank $d$ of $F$, and this may be taken to be infinite. We thus have

THEOREM 6.6 *The free group* (*of any rank*) *is residually a finite p-group, for any prime p.* ■

COROLLARY 6.7 (W. Magnus) *In a free group* $F$, $\bigcap_n \gamma_n(F) = 1$.

For if $\cap\gamma_n(F) \neq 1$, take a word $c \neq 1$ in the intersection. By Th. 6.6 there is a normal subgroup $N$ not containing $c$ such that $F/N$ is a finite $p$-group. Since $F/N$ is nilpotent, of class $h$ say, we have $\gamma_{h+1}(F) \subseteq N$, and so $c \notin \gamma_{h+1}(F)$, which is a contradiction. ■

For another proof of this Cor. see Ex. (15) of the further exercises.

## Exercises

(1) Let $A$ be a group acting by automorphisms on $G$, such that in the semidirect product, $(G, A) = G$. Show that if $N \lhd G$ is such that $(A, N) = 1$, then $N \subseteq Z(G)$.

(2) Let $G$ be a group and $K, L, N$ subgroups such that $N \lhd K$, $K \lhd G$, $(L, K) = K$, $KL = G$, $(L, N) = 1$. Show that $N \subseteq Z(K)$.

(3) Prove Prop. 6.3.

(4) Show that if $H_1, \ldots, H_n$ are normal subgroups of $G$, then $(H_1, \ldots, H_n)$ is the subgroup generated by all $(x_1, \ldots, x_n)$, where $x_i \in H_i$ and give an example to show that normality cannot be omitted.

(5) (H. Wielandt) Give a direct proof that a finite group is nilpotent iff every maximal subgroup is normal.

(6) In any group $G$, for fixed $a \in G$, write $x\alpha = x^{-1}ax$, $x\lambda = (a, x)$. Show that $x\alpha^n = a \cdot x\lambda^n$.

(7) Show that for a finite $p$-group $G$, $G/\Phi(G)$ is a vector space over $\mathbf{F}_p$; if its dimension is $d$, then any minimal generating set of $G$ has $d$ elements. Show that every element of $G$ which is not a non-generator occurs in a minimal generating set.

(8) Show that in a symmetric group every cycle of odd length is a commutator (and not merely a product of commutators).

(9) (D. Ornstein) Show (by induction on $r$) that $(a, b)^r$ can be written as a product of $(ba)^{-r}(ab)^r$ and $r - 1$ commutators. Deduce that if in a group $G$, $Z(G)$ has finite index in $G$, then $G'$ is finite.

(10) Show that a group satisfying the law $x^2 = 1$ is abelian. Deduce that any commutator can be written as a product of squares and find an explicit expression of $(x, y)$ in this form.

## 4.7 Linear groups

In **4.1** we saw how in principle at least every finite group can be built up from simple groups. This leaves the problem of determining the simple finite groups,

and it was only around 1980 that the classification of simple finite groups was completed. The full list contains 18 families (including the cyclic groups of prime order) and another 26 'sporadic' groups. This is not the place to give a detailed account (cf. Gorenstein [1982]), but we shall present some of the families that have been known since the work of Dickson [1901]. In A.1 it was shown that the alternating group $A_n$ is simple for $n \geqslant 5$ (A.1, Th. 4 of **9.5**; also A.2, **3.11**). In this section we shall describe the linear groups and later, in Ch. 8, the orthogonal and symplectic groups, collectively usually known under the name of *classical groups* since the time of H. Weyl [1946]. These families of groups can be shown to correspond to certain infinite families of simple Lie algebras, though we shall not do so here. It was Chevalley's paper in 1955, describing how these families of Lie algebras could also be defined over finite fields, that provided the impetus for research in the 1960s and 1970s that led to the classification of the simple finite groups.

Let $k$ be any field and $V$ an $n$-dimensional vector space over $k$. The automorphisms of $V$ form a group whose elements may be described, after the choice of a basis in $V$, by invertible $n \times n$ matrices over $k$. This is the *general linear group* of degree $n$ over $k$, written $\mathbf{GL}(V)$ or $\mathbf{GL}_n(k)$. The determinant function provides a homomorphism from $\mathbf{GL}_n(k)$ to $k^\times$ whose kernel $\mathbf{SL}_n(k)$ is the *special linear group*, consisting of all $n \times n$ matrices with determinant 1. We remark that the general linear group may be defined more generally over any ring $R$ as $\mathbf{GL}_n(R)$, the set of all invertible $n \times n$ matrices over $R$. When $R$ is commutative we also find the subgroup $\mathbf{SL}_n(R)$ as before, but this definition breaks down for more general rings. Later in Ch. 11 we shall find a way of defining $\mathbf{SL}_n(K)$ for a skew field $K$, but for the present we shall mainly be concerned with a commutative field as ground ring.

We begin by finding generating sets for $\mathbf{GL}_n$ and $\mathbf{SL}_n$. We denote by $E_{ij}$ the matrix with $(i,j)$-entry 1 and all others zero. By an *elementary matrix* we understand a matrix of the form

$$B_{ij}(a) = I + aE_{ij}, \qquad \text{where } a \in k, i \neq j.$$

It is clear that this matrix lies in $\mathbf{SL}_n$, with inverse $B_{ij}(-a)$. A diagonal matrix $\sum a_i E_{ii}$ lies in $\mathbf{GL}_n$ iff $\prod a_i \neq 0$, while $\prod a_i = 1$ is the condition for it to belong to $\mathbf{SL}_n$.

Taking $V$ again as our vector space over $k$, let $u \neq 0$ be a vector in $V$. By a *transvection* along $u$ we understand a linear mapping $T$ of $V$ keeping $u$ fixed and such that $xT - x$ is in the direction of $u$. Thus we can express $T$ as

$$xT = x + \lambda(x)u,$$

where $\lambda$ is a linear functional on $V$ such that $\lambda(u) = 0$. For example, with the standard basis $e_1, \ldots, e_n$ in $k^n$, $B_{1n}(a)$ maps $x = \sum x_i e_i$ to $x + ax_1 e_n$ and so is a transvection along $e_n$ (for $n > 1$). Given any transvection $T$ along $u$, let us choose a basis $u_1, \ldots, u_n$ of $V$ such that $u_2, \ldots, u_n$ forms a basis for the kernel of $T - I$,

while $u_n = u$. Then $T$ takes the form $xT = x + \lambda(x)u_n$, so $T$ is represented by an elementary matrix relative to this basis.

For our first result we can take our ring to be any Euclidean domain; we recall (from A.1, **6.5**) that a *Euclidean domain* is an integral domain in which the Euclidean algorithm holds (relative to a norm function).

THEOREM 7.1 *For any $n \geqslant 1$ and any (commutative) Euclidean domain $R$, $\mathbf{SL}_n(R)$ is generated by all elementary matrices and $\mathbf{GL}_n(R)$ by all elementary and diagonal matrices with units along the main diagonal. In particular, this holds for any field.*

*Proof.* The result is clear for $n = 1$, so assume $n > 1$. Let $A \in \mathbf{SL}_n(R)$; it is clear that right multiplication by $B_{ij}(c)$ corresponds to the operation of adding the $i$th column, multiplied by $c$, to the $j$th column. Now the Euclidean algorithm shows that multiplication on the right alternatively by $B_{12}(p)$ and $B_{21}(q)$ for suitable $p, q$ reduces $a_{11}, a_{12}$ to $d, 0$ respectively, where $d$ is an HCF of $a_{11}$ and $a_{12}$. By operating similarly on other columns we reduce all elements of the first row after the first one to zero. We next repeat this process on the rows, to reduce all elements in the first column after the first one to zero, by multiplying by suitable elementary matrices on the left. This either leaves the top row unchanged or it replaces the (1, 1)-entry by a proper factor. By an induction on the number of factors we reduce $A$ to the form $a \oplus A_1$, where $A_1$ is $(n-1) \times (n-1)$, and another induction, this time on $n$, reduces $A$ to diagonal form. Using the factorization

$$\begin{pmatrix} c & 0 \\ 0 & c^{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c^{-1} & 1 \end{pmatrix}\begin{pmatrix} 1 & -c \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ c^{-1} & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix},$$

we can express $\mathrm{diag}(a_1, a_2, \ldots, a_n)$ by $\mathrm{diag}(1, a_1 a_2, a_3, \ldots, a_n)$ and a product of elementary matrices, and hence by another induction express $A$ as a product of elementary matrices.

The same process applied to a matrix of $\mathbf{GL}_n(R)$ reduces it to a diagonal matrix and this proves the second assertion. ■

Let $R$ be any commutative ring; the determinant may be regarded as a homomorphism from $\mathbf{GL}_n(R)$ to $\mathbf{U}(R)$ and it yields the exact sequence

$$1 \to \mathbf{SL}_n(R) \to \mathbf{GL}_n(R) \xrightarrow{\det} \mathbf{U}(R) \to 1. \tag{1}$$

Since $R$ is commutative, it follows that $\mathbf{GL}_n(R)' \subseteq \mathbf{SL}_n(R)$; for fields we have the following more precise relation:

PROPOSITION 7.2 *Let $k$ be any field and $n \geqslant 2$. Then*

$$\mathbf{SL}_n(k)' = \mathbf{GL}_n(k)' = \mathbf{SL}_n(k), \tag{2}$$

*except when $n=2$ and $k$ has 2 or 3 elements. Thus $\mathbf{SL}_n(k)$ is perfect (with the exceptions listed).*

*Proof.* As we have remarked, we have $\mathbf{SL}_n(k)' \subseteq \mathbf{GL}_n(k)' \subseteq \mathbf{SL}_n(k)$. To establish equality it is enough, by Th. 7.1, to show that every elementary matrix is a product of commutators. Now it is easily checked that for any distinct indices $i, j, k$,

$$(B_{ik}(a), B_{kj}(1)) = B_{ik}(-a)B_{kj}(-1)B_{ik}(a)B_{kj}(1) = B_{ij}(a).$$

This expresses every elementary matrix as a commutator when $n \geqslant 3$. For $n=2$ we have

$$\left(\begin{pmatrix} b^{-1} & 0 \\ 0 & b \end{pmatrix}, \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & (b^2-1)c \\ 0 & 1 \end{pmatrix}; \tag{3}$$

hence if $k$ contains an element $b$ such that $b \neq 0$, $b^2 \neq 1$, then we can express any elementary matrix $B_{12}(a)$ as a commutator by taking $c = (b^2-1)^{-1}a$ in (3), and similarly for $B_{21}(a)$. This shows that $\mathbf{SL}_n(k)' = \mathbf{SL}_n(k)$ except when $n=2$ and $b^3 = b$ for all $b \in k$, and this happens only when $|k| = 2$ or 3. ■

We shall see that $\mathbf{SL}_2(\mathbf{F}_2)$ and $\mathbf{SL}_2(\mathbf{F}_3)$ are soluble (cf. Ex. (1)), so the exceptions made in Prop. 7.2 do in fact occur.

Our next aim is to show that $\mathbf{SL}_n(k)$ modulo its centre is simple (except when $n=2$ and $|k| \leqslant 3$). By taking $k$ to be finite we thus obtain a family of simple groups. The centre of $\mathbf{GL}_n(k)$ or $\mathbf{SL}_n(k)$ clearly consists of all scalar matrices and the quotients are known as the *projective* linear groups, written

$$\mathbf{PGL}_n(k) = \mathbf{GL}_n(k)/Z, \qquad \mathbf{PSL}_n(k) = \mathbf{SL}_n(k)/Z \cap \mathbf{SL}_n(k),$$

where $Z$ is the centre of $\mathbf{GL}_n(k)$. Let us define projective $n$-space $\mathbf{P}^n(k)$ as the set of all $(n+1)$-tuples $x = (x_0, x_1, \ldots, x_n) \in k^{n+1}$, where the $x_i$ are not all 0 and $x = y$ iff $x_i = \lambda y_i$ for some $\lambda \in k^\times$. Thus the points of $\mathbf{P}^n(k)$ are given by the ratios of $n+1$ coordinates. It is clear that $\mathbf{PGL}_{n+1}(k)$ and $\mathbf{PSL}_{n+1}(k)$ act in a natural way on the points of $\mathbf{P}^n(k)$.

To prove the simplicity of $\mathbf{PSL}_n$ (following Iwasawa) we shall need to recall the notion of primitivity and derive some of its properties. We recall that a permutation group $G$ acting on a set $S$ is *transitive* if for any $p, q \in S$ there exists $g \in G$ such that $pg = q$, thus $S$ consists of a single $G$-orbit. The *stabilizer* of $p \in S$ is $\mathrm{St}_p = \{x \in G \mid px = p\}$, a subgroup of $G$. When $G$ acts on $S$, it acts in a natural way on $S^n$ for $n \geqslant 1$, and if this action is transitive, $G$ is said to be *$n$-fold* transitive. $G$ is said to be *primitive* if it acts transitively and there is no partition of $S$ (other than the trivial ones consisting of $S$ alone and of the 1-element subsets) which is compatible with the $G$-action. As we saw in A.1, **9.5**, $G$ is imprimitive precisely when $S$ contains a proper subset $S_0$ with more than one element such that for each $g \in G$ either $S_0 g = S_0$ or $S_0 g \cap S_0 = \varnothing$. For example, any 2-fold transitive

group is primitive, for if $p, q \in S_0$ then there exists $g \in G$ with $pg \in S_0$, $qg \notin S_0$, hence $S_0 g$ meets $S_0$ but is not equal to it. We also recall from A.1, Th. 5 of **9.5** that a transitive permutation group $G$ is primitive iff the stabilizer of any point is a maximal subgroup of $G$. Further, we shall need the following necessary condition for primitivity.

LEMMA 7.3 *Let $G$ be a primitive permutation group acting on a set $S$. Then any non-trivial normal subgroup $N$ is transitive, and $G = \mathrm{St}_p \cdot N$ for the stabilizer of any point $p$ of $S$.*

*Proof.* Let $N \lhd G$ and consider the orbits under the action of $N$. For any $p \in S, g \in G$, we have $pNg = pgN$, and this equals $pN$ if $pg = p$ and otherwise is disjoint. By primitivity it follows that $pN = S$. Now fix $p \in S$ and take any $g \in G$; then $pg = pu$ for some $u \in N$, hence $gu^{-1} \in \mathrm{St}_p$ and so $g \in \mathrm{St}_p \cdot N$, as claimed. ■

This leads to a criterion for simplicity:

LEMMA 7.4 *Let $G$ be a primitive permutation group, acting on $S$. If for some $p \in S$, $\mathrm{St}_p$ contains an abelian normal subgroup $A$ such that $G$ is generated by all the conjugates $g^{-1}Ag (g \in G)$, then any non-trivial normal subgroup of $G$ contains $G'$. In particular, if $G$ is perfect, then it must be simple.*

*Proof.* Let $N \neq 1$ be a normal subgroup of $G$. By Lemma 7.3, $N$ is transitive and $G = \mathrm{St}_p \cdot N$, where $p \in S$. Consider $AN$; this is easily seen to be a normal subgroup of $G$ and so contains all conjugates $g^{-1}Ag$, hence $AN = G$ by the hypothesis. Now $G/N \cong A/(A \cap N)$ is abelian, therefore $N \supseteq G'$ and the conclusion follows. ■

To prove that $\mathbf{PSL}_n(k)$ is simple we shall use this lemma, taking the action of $\mathbf{PSL}_n(k)$ on $\mathbf{P}^{n-1}(k)$. We note that the action is doubly transitive when $n \geqslant 2$. For let $p_1, p_2$ and $q_1, q_2$ be two pairs of distinct points in $\mathbf{P}^{n-1}(k)$. If $x_i, y_i$ are coordinate vectors for $p_i, q_i$ so that $x_1, x_2$ are linearly independent, as well as $y_1, y_2$, then there exists $A \in \mathbf{SL}_n(k)$ such that $x_1 A = y_1, x_2 A = cy_2$ for some $c \in k^\times$, hence $A$ maps $p_i$ to $q_i$ for $i = 1, 2$. By what was said earlier it follows that $\mathbf{PSL}_n(k)$ is primitive.

THEOREM 7.5 *Let $k$ be a field and $n \geqslant 2$. Then $\mathbf{PSL}_n(k)$ is simple except when $n = 2$ and $|k| \leqslant 3$.*

*Proof.* As we have just seen, the action of $G = \mathbf{PSL}_n(k)$ on $\mathbf{P}^{n-1}(k)$ is primitive; moreover, $G$ is perfect, for if $H = \mathbf{SL}_n(k)$ and $Z$ denotes the centre of $H$, then $gG = H/Z$, hence by Prop. 7.2, $G' = H'Z/Z = H/Z = G$. We shall complete the proof by verifying the conditions of Lemma 7.4.

Let $p$ be the point with coordinates $(1, 0, \ldots, 0)$. This is left fixed by any matrix

$$\begin{pmatrix} a & 0 \\ b & A \end{pmatrix} \text{ with inverse } \begin{pmatrix} a^{-1} & 0 \\ -A^{-1}ba^{-1} & A^{-1} \end{pmatrix}, \qquad \text{where } a \in k, b \in {}^{n-1}k, A \in k_{n-1}. \tag{4}$$

Consider the subgroup $L$ consisting of all matrices

$$\begin{pmatrix} 1 & 0 \\ c & I \end{pmatrix}, \qquad \text{where } c \in {}^{n-1}k. \tag{5}$$

Clearly this is an abelian subgroup, and transforming by (4) we find

$$\begin{pmatrix} a^{-1} & 0 \\ -A^{-1}ba^{-1} & A^{-1} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ c & I \end{pmatrix}\begin{pmatrix} a & 0 \\ b & A \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ A^{-1}ca & I \end{pmatrix}.$$

Thus $L$ is a normal abelian subgroup of $\mathrm{St}_p$. Clearly every elementary matrix is conjugate to a matrix in $L$, so by Th. 7.1, $L$ and its conjugates generate $G$. Thus $G$ satisfies the conditions of Lemma 7.4, and it is perfect, therefore it is simple. ■

Let us determine the order of $\mathbf{PSL}_n(k)$ when $k$ is finite, with $q$ elements say. We begin with $\mathbf{GL}_n(k)$; this group acts on $V = k^n$, a vector space with $q^n$ elements. Any element of $\mathbf{GL}_n(k)$ is completely determined by the image of the standard basis $e_1, e_2, \ldots, e_n$ of $V$, and this image can be any basis of $V$. Thus $e_1$ can map to any non-zero vector, giving $q^n - 1$ choices, $e_2$ can map to any vector independent of the image of $e_1$, giving $q^n - q$ choices, $e_3$ can map to any vector not a linear combination of the images of $e_1, e_2$, giving $q^n - q^2$ choices, and so on. In this way we find that

$$|\mathbf{GL}_n(k)| = (q^n - 1)(q^n - q)\cdots(q^n - q^{n-1}), \qquad \text{where } |k| = q.$$

To determine $|\mathbf{SL}_n(k)|$ we use the exact sequence (1) and find

$$|\mathbf{SL}_n(k)| = (q^n - 1)(q^n - q)\cdots(q^n - q^{n-2})q^{n-1}.$$

To find the order of $\mathbf{PSL}_n(k)$ we need to calculate the order of its centre $Z$. We recall that $Z$ consists of all matrices $cI$ such that $c^n = 1$, so we need to find the number of solutions of $x^n = 1$ in $k$. But $k^\times$ is a cyclic group of order $q - 1$, so the number we want is $d = (n, q - 1)$. Thus

$$|\mathbf{PSL}_n(k)| = (q^n - 1)(q^n - q)\cdots(q^n - q^{n-2})q^{n-1}/d, \qquad \text{where } d = (n, q - 1).$$

## Exercises

(1) By examining the action of $\mathbf{PSL}_2(k)$ on $\mathbf{P}^1(k)$, show that $\mathbf{PSL}_2(\mathbf{F}_2) \cong S_3, \mathbf{PSL}_2(\mathbf{F}_3) \cong A_4$.

(2) Show that $\mathbf{PSL}_2(\mathbf{F}_4) \cong \mathbf{PSL}_2(\mathbf{F}_5) \cong A_5$.

(3) Show that $\mathbf{PSL}_2(\mathbf{F}_9) \cong A_6, \mathbf{PSL}_4(\mathbf{F}_2) \cong A_8$.

(4) Apply Lemma 7.4 to show that $A_5$ is simple.

(5) How much remains true of Th. 7.1 when commutativity is dropped?

(6) Show that $\mathbf{PSL}_3(\mathbf{F}_4)$ and $\mathbf{PSL}_4(\mathbf{F}_2)$ have the same order but are not isomorphic. (Hint: compare the Sylow 2-subgroups.)

## Further exercises on Chapter 4

(1) Let $A$ be a group and $\alpha$ an automorphism of $A$. Show that there is a split extension $E$ of $A$ by an infinite cyclic group such that $\alpha$ is induced by an inner automorphism of $E$. Show also that any extension of $A$ by an infinite cyclic group splits.

(2) Defining an *automorphism* of an extension $E$ as an isomorphism of $E$ with itself, show that the group of all automorphisms of an extension $E$ of an abelian group $A$ by a group $G$ is isomorphic to the group of 1-cocycles of $G$ in $A$. Show that if the extension is split and $H^1(G, A) = 0$, then any two complements of $A$ in $E$ are conjugate.

(3) Let $F$ be the free group on $X$ and $kF$ the group algebra over a field $k$. Obtain the following analogue of (14) of **2.2**:

$$0 \to {}^{X}(kF \otimes_k kF) \to kF \otimes_k kF \to kF \to 0.$$

By applying $k \otimes_{kF}$ deduce that the augmentation ideal $IF$ is free on $x - 1$ $(x \in X)$ as $F$-module. Hence show that $H^n(F, A) = H_n(F, A) = 0$ for $n > 1$.

(4) Let $G$ be a finite primitive permutation group acting faithfully on a set $X$ (i.e. such that no element $\neq 1$ leaves every point of $X$ fixed). Show that either $F(G) = 1$ or (i) $F = F(G)$ is a minimal normal subgroup of $G$, and is an elementary abelian $p$-group, (ii) $F$ is its own centralizer, (iii) every element $\neq 1$ in $F$ moves all points of $X$, (iv) $|X| = p^r$. (Hint: by primitivity all the stabilizers are conjugate maximal subgroups of $G$. Fix a prime $p$ dividing the centre $Z(F)$ of $F$ and take a maximal elementary abelian $p$-subgroup of $Z(F)$.)

(5) Let $G$ be a finite group and $H$ a subgroup. A *partial complement* for $H$ is a subgroup $K \neq G$ such that $HK = G$. Show that a normal subgroup of $G$ has a partial complement iff it is not contained in $\Phi(G)$.

(6) Let $G$ be a finite group and $A$ an abelian normal subgroup. Show that if $A \cap \Phi(G) = 1$, then $A$ has a complement in $G$.

(7) Show that in a finite soluble group $G$, with a Hall subgroup of order $r$, the number

$h_r$ of subgroups of order $r$ in $G$ has the form $h_r = c_1 \cdots c_t$, where $c_i$ is a prime power dividing the order of a chief factor of $G$ and $c_i \equiv 1$ modulo a prime factor of $r$. (Hint: let $|G| = rs$, $(r,s) = 1$. Treat first the case where $G$ has a normal subgroup of index $r's'$, $r'|r$, $s'|s$ and $s' > 1$.)

(8) Let $E/k$ be a finite Galois extension of degree $m = q_1 \cdots q_r$, where $q_i = p_i^{\alpha_i}$ and the $p_i$ are distinct primes. Show that if $E$ contains a subfield $E_i$ of degree $q_i$ over $k$, for $i = 1, \ldots, r$, then

$$E = E_1 \otimes_k \cdots \otimes_k E_r, \qquad [E_i : k] = q_i. \tag{1}$$

Use Hall's theorem to show that a decomposition (1) of $E$ exists iff $G = \mathrm{Gal}(E/k)$ is soluble and that any two such decompositions of $E$ are conjugate by an element of $G$.

(9) Show that for any finite Galois extension $E/k$, a decomposition (1) into *Galois* extensions $E_i/k$ of prime power degrees exists iff $\mathrm{Gal}(E/k)$ is nilpotent.

(10) Show that the order of a finite simple group is either divisible by 12 or by the cube of the least prime dividing its order. (Hint: if a Sylow $p$-group $P$ has order $p$ or $p^2$, it is abelian. Now use Th. 4.2 to describe the action of $N_G(P)$ on $P$.)

(11) Let $X$ be a subset of a free group $F$ and define an *elementary transformation* of $X$ as one of the following: (i) replace $x$ by $xy$ $(x, y \in X, x \neq y)$, (ii) replace $x$ by $x^{-1}$ $(x \in X)$, (iii) omit 1. A *Nielsen transformation* is a series of elementary transformations. Show that if $X_1$ is obtained by a Nielsen transformation from $X$, then $\mathrm{gp}\{X_1\} = \mathrm{gp}\{X\}$. Show also that any finite subset $X$ can be *reduced* by a Nielsen transformation, i.e. brought to the form where $1 \notin X$, $x, y \in X \cup X^{-1}$, $xy \neq 1$ implies $|xy| \geqslant |x|, |y|$ and $x, y, z \in X \cup X^{-1}$, $xy \neq 1, yz \neq 1$ implies $|xyz| > |x| - |y| + |z|$. ($|w|$ is the length of $w$ in terms of a basis.)

(12) (H. Zieschang) Let $U$ be a reduced set in a free group $F$ (cf. Ex. (11)). For each $u \in U \cup U^{-1}$ let $a(u)$ be the longest initial segment of $u$ which is cancelled in any product $vu \neq 1, v \in U \cup U^{-1}$. Show that $u = a(u)m(u)a(u^{-1})^{-1}$ is reduced, where $m(u) \neq 1$, and if $w = u_1 \cdots u_n$ then in the reduced form of $w, m(u_1), \ldots, m(u_n)$ are uncancelled.

(13) Show that a reduced subset $X$ of a free group is a basis of $\mathrm{gp}\{X\}$. Deduce that every finitely generated subgroup of a free group is free.

(14) Let $F$ be a free group of finite rank on a basis $X$ and let $\alpha$ be an automorphism of $F$. By applying Ex. (11) and (12) to the set $X^\alpha$ show that every automorphism of $F$ can be obtained by a permutation and a Nielsen transformation of the generating set.

(15) (M. Takahasi) Let $F$ be a free group and $F = F_1 \supset F_2 \supset \cdots$ a series of subgroups such that $F_{i+1}$ contains no element of a basis of $F_i$. Show that relative to any basis of $F_1$ any element $w \neq 1$ of $F_{i+1}$ satisfies $|w| > i$. Deduce that any infinite series of characteristic subgroup of $F$ intersects in 1, and so obtain another proof of Magnus's theorem (Cor. 6.7).

(16) Show that $\mathbf{PSL}_3(\mathbf{F}_2) \cong \mathbf{PSL}_2(\mathbf{F}_7)$ and find its order. (Hint: this is the subgroup of

$S_7$ in the action on the columns of the array

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 4 & 5 & 6 & 7 & 1 \\ 4 & 5 & 6 & 7 & 1 & 2 & 3 \end{matrix}$$

which maps each column to another; cf. Ex. (8) of **3.6**, A.1. The columns may be interpreted as lines in $\mathbf{P}^2(\mathbf{F}_2)$, the projective plane over $\mathbf{F}_2$.)

# 5
# Further field theory

The chapter on field theory in A.2 was almost entirely devoted to extensions of finite degree and concentrated on Galois theory. However, even an introductory account should make some mention of infinite extensions, and we shall discuss them in the present chapter, including transcendental extensions and infinite Galois theory. In addition there are concise accounts of topics that are useful in commutative ring theory and algebraic geometry, besides being of independent interest: separability, linearly disjoint extensions, Galois cohomology and Kummer extensions.

## 5.1 Algebraic dependence

There is a notion of algebraic dependence that is entirely analogous to the notion of linear dependence introduced in A.1, **4.1**. We shall find that it satisfies the axioms for an abstract dependence relation given in **1.4**, and we shall therefore be able to use the consequences deduced there.

Given any field extension $E/k$, let $u_1,\dots,u_m,v\in E$; we shall say that $v$ is *algebraically dependent* on $u_1,\dots,u_m$ over $k$ if $v$ is algebraic over the field $k(u_1,\dots,u_m)$. This just means that $v$ satisfies an equation with coefficients in $k(u_1,\dots,u_m)$; on multiplying by a common denominator, we obtain an equation for $v$:

$$a_0v^d + a_1v^{d-1} + \cdots + a_d = 0, \tag{1}$$

where $a_i\in k[u_1,\dots,u_m]$ and $a_0 \neq 0$. It is easily seen that this dependence relation satisfies the rules **D**.0–2 in **1.4**:

**D**.0 *Each $u_i$ is algebraically dependent on $u_1,\dots,u_m$.*

**D**.1 *If $w$ is algebraically dependent on $v_1,\dots,v_n$ and each $v_i$ is algebraically dependent on $u_1,\dots,u_m$, then $w$ is algebraically dependent on $u_1,\dots,u_m$.*

**D**.2 *If $v$ is algebraically dependent on $u_1,\dots,u_m$ but not on $u_2,\dots,u_m$, then $u_1$ is algebraically dependent on $v,u_2,\dots,u_m$.*

**D**.0 holds trivially; to prove **D**.1, we are given that each $v_i$ is algebraic over

$k(u_1,\ldots,u_m)$, hence in the tower

$$k(u_1,\ldots,u_m)\subseteq k(u_1,\ldots,u_m,v_1)\subseteq\cdots\subseteq k(u_1,\ldots,u_m,v_1,\ldots,v_n)$$

each extension is of finite degree; moreover, $w$ is algebraic over $k(v_1,\ldots,v_n)$ and *a fortiori* also over $k(u_1,\ldots,u_m,v_1,\ldots,v_n)$. By the product formula for the degrees, $k(u_1,\ldots,u_m,v_1,\ldots,v_n,w)$ is of finite degree over $k(u_1,\ldots,u_m)$, and this shows $w$ to be algebraically dependent on $u_1,\ldots,u_m$.

Finally, to prove **D**.2, we have by hypothesis a relation (1), i.e. a linear dependence over $k$ between the power products in $u_1,\ldots,u_m,v$. Let us arrange the left-hand side as a polynomial in $u_1$:

$$b_0u_1^r+b_1u_1^{r-1}+\cdots+b_r=0, \tag{2}$$

where $b_i\in k[u_2,\ldots,u_m,v]$. By hypothesis, $a_0\neq 0$ in (1); if $b_0,b_1,\ldots,b_{r-1}$ all vanish, then (2) reduces to $b_r=0$, a polynomial relation between $u_2,\ldots,u_m,v$ and since this is obtained by rearranging (1), it would show that $v$ is algebraically dependent on $u_2,\ldots,u_m$, against the hypothesis. Hence $b_0,\ldots,b_{r-1}$ do not all vanish and now (2) shows that $u_1$ is algebraically dependent on $u_2,\ldots,u_m,v$.

Thus algebraic dependence satisfies the axioms **D**.0–2 of **1.4**, and we can use the concepts introduced there, and the facts proved about them. From the definitions given there we see that a subset $X$ of $E$ is *algebraically independent over* $k$ if no element of $X$ is algebraically dependent on the rest. If $X=\{x_1,\ldots,x_n\}$, this means that the power products $x_1^{a_1}\cdots x_n^{a_n}$ are linearly independent over $k$; but the definition applies even if $X$ is infinite. A subset $B$ of $E$ which is algebraically independent and such that every element of $E$ is algebraically dependent on $B$ over $k$ is called a *transcendence basis* of $E$ over $k$. From Lemma 1.4.2, Cor. 1.4.3 and Prop. 1.4.4 we now obtain

THEOREM 1.1 *Given a field extension $E/k$, any algebraically independent subset of $K$ can be completed to a transcendence basis of $E/k$ and any two such bases have the same number of elements.* ■

The number of elements in a transcendence basis of $E/k$ is called the *transcendence degree* or *dimension* of $E/k$ and is written tr. deg$(E/k)$. An extension of transcendence degree zero is called *algebraic*. Thus an extension $E/k$ is algebraic precisely if every element of $E$ is algebraic over $k$. As we saw in Cor. 3.1.5 of A.2, it is enough to assume that $E$ has a generating set of algebraic elements, for $E/k$ to be algebraic. We remark that a field extension $E/k$ which is finitely generated, by $x_1,\ldots,x_n$ say, is of finite degree over $k$ iff each $x_i$ is algebraic over $k$, but in general an extension may well be algebraic and yet of infinite degree over $k$. For example, the algebraic closure of a field is always algebraic but not generally of finite degree.

The transcendence degree of a repeated extension is given by the following formula:

PROPOSITION 1.2 *For any fields $k \subseteq E \subseteq F$ we have*

$$\text{tr. deg}(F/k) = \text{tr. deg}(F/E) + \text{tr. deg}(E/k). \tag{3}$$

*Proof.* Let $X$ be a transcendence basis of $E/k$ and $Y$ one of $F/E$. Then no element of $Y$ is in $E$, but every element of $X$ is, hence $X \cap Y = \varnothing$ and to establish (3) it will be enough to show that $X \cup Y$ is a transcendence basis for $F/k$. Any element $z$ of $F$ is algebraic over $E(Y)$; by writing down an equation for $z$ we see that $z$ is algebraically dependent on a finite set over $k$, say $y_1, \ldots, y_r, u_1, \ldots, u_s$, where $y_i \in Y$, $u_j \in E$. Each $u_j$ is algebraically dependent on $X$ over $k$, hence (by transitivity) $z$ itself is algebraically dependent on $X \cup Y$ over $k$. Secondly we must show that $X \cup Y$ is algebraically independent. Suppose not; then there is a non-trivial polynomial relation

$$f(x_1, \ldots, x_r, y_1, \ldots, y_s) = 0, \qquad x_i \in X,\ y_j \in Y, \tag{4}$$

with coefficients in $k$, i.e. the power products in the $x_i, y_j$ are linearly dependent over $k$. Here some $y$'s must occur because $X$ is algebraically independent over $k$. Since $x_i \in E$, (4) is an algebraic relation between the $y$'s over $E$; but $Y$ was algebraically independent over $E$, so if we rewrite (4) as a polynomial in the $y$'s, all the coefficients vanish. But these coefficients are polynomials in the $x$'s, so we obtain an algebraic relation between the $x$'s, again a contradiction. Thus $X \cup Y$ is algebraically independent, and hence is a transcendence basis for $F/k$ and (3) follows. ■

In particular, for algebraic extensions this shows

COROLLARY 1.3 *An algebraic extension of an algebraic extension is again algebraic.* ■

A field extension is called *transcendental* if it is not algebraic; thus a transcendental field extension is one of positive transcendence degree. An extension $E/k$ is said to be *purely transcendental* if there is an algebraically independent subset $X$ of $E$ such that $E = k(X)$. Such a purely transcendental extension of $k$ is determined up to isomorphism by its transcendence degree: given any cardinal $\alpha$, take a set $X$ of cardinal $\alpha$, construct the polynomial ring $k[X]$ in the indeterminates $X$ over $k$ and form its field of fractions $k(X)$. This field, consisting of all rational functions in the indeterminates $X$ over $k$, is a purely transcendental extension of tr. deg. $\alpha$ over $k$, and it is clearly the only one, up to isomorphism.

If $E = k(x_1, \ldots, x_n)$ is a purely transcendental extension of $k$, then any permutation of $x_1, \ldots, x_n$ defines a $k$-automorphism of $E$; thus any permutation group of the $x$'s acts in a natural way as a group of $k$-automorphisms of $E$. It is an old problem (known as E. Noether's problem) whether the fixed field

under a given permutation group is a purely transcendental extension of $k$. This was answered negatively by R. G. Swan (1969) for $n = 47$ and a cyclic transitive group. More general results have been obtained by V. E. Voskresenskii (1973) and H. W. Lenstra (1974). The latter gives a complete solution of Noether's problem for abelian transitive groups in a most readable account which also includes a survey of earlier results.

### Exercises

(1) Show that the fixed field in $k(x_1, \ldots, x_n)$ under the group of all permutations of the $x$'s is a purely transcendental extension of $k$, of transcendence degree $n$.

(2) Show that the extension of $\mathbf{C}$ defined by $x, y$ subject to $x^2 + y^2 = 1$ is purely transcendental, but for the equation $x^3 + y^3 = 1$ it is not. (Hint: write $x = f/h$, $y = g/h$ and examine the degrees of $f, g, h$. Try differentiating.)

(3) Let $E$ be the extension of $\mathbf{R}$ defined by $x, y$ subject to $x^2 + y^2 + 1 = 0$. Show that any element of $E \backslash \mathbf{R}$ is transcendental over $\mathbf{R}$ and deduce that $E/\mathbf{R}$ is not purely transcendental.

(4) Show that $\mathrm{tr.deg}(\mathbf{C}/\mathbf{Q}) = \mathfrak{c}$, where $\mathfrak{c} = 2^{\aleph_0}$. Deduce that the automorphism group of $\mathbf{C}$ has order $2^{\mathfrak{c}}$.

(5) Show that a field extension $E/k$ is algebraic iff every subalgebra of $E$ over $k$ is a subfield.

(6) Let $F/k$ be any finitely generated field extension and $k \subseteq E \subseteq F$. Show that $E$ is finitely generated over $k$.

(7) Let $k$ be any field and $E = k((x))$ the field of formal Laurent series $\sum_{-N}^{\infty} a_\nu x^\nu$ in an indeterminate $x$. Show that $E$ has $k(x)$ as a subfield and that $E$ is not algebraic over $k(x)$, even though it contains elements that are algebraic over $k(x)$.

## 5.2 Simple transcendental extensions

We have seen that a purely transcendental extension of a field $k$, of finite transcendence degree $d$, is just the field $k(x_1, \ldots, x_d)$ of rational functions in $d$ independent indeterminates. The general study of such fields raises some difficult questions, but in the case $d = 1$ the situation is rather better. We then have a field of rational functions of one variable, $k(x)$, also called a *simple transcendental extension*, and we shall now examine this case.

Let $E = k(x)$ be a simple transcendental extension; its elements are rational functions in $x$ over $k$, and so may be written

$$u = \frac{f(x)}{g(x)}, \qquad \text{where } f, g \in k[x],\ (f, g) = 1 \text{ and } g \text{ is monic.} \tag{1}$$

The *degree* of $u$ is defined as $\deg u = \max\{\deg f, \deg g\}$. It is positive unless $u$ is constant, i.e. $u \in k$, and although the representation (1) depends on the choice of the generator $x$, the degree of $u$ is independent of this choice, by the next result. Given any field extension $E/k$, we shall say that $k$ is *relatively algebraically closed* in $E$ if every element of $E \setminus k$ is transcendental over $k$; in other words, any element of $E$ algebraic over $k$ already lies in $k$.

PROPOSITION 2.1 *Let $k(x)$ be a simple transcendental extension of $k$. Then $k$ is relatively algebraically closed in $k(x)$, and for any $u$ in $k(x) \setminus k$, the degree of $u$ is given by*

$$\deg u = [k(x):k(u)]. \tag{2}$$

*Moreover, if $u$ is given by* (1), *then an irreducible polynomial for $x$ over $k(u)$ is $f(X) - ug(X)$.*

*Proof.* If $u$ is given by (1), we know that $x$ satisfies the equation in $X$:

$$f(X) - ug(X) = 0. \tag{3}$$

Let $b_i$ be a non-zero coefficient in $g(X)$ and $a_i$ the corresponding coefficient in $f(X)$; then the corresponding coefficient in (3) is $a_i - ub_i$. This is non-zero, for otherwise $u = a_i/b_i \in k$, against the hypothesis. Thus not all the coefficients of (3) vanish, and it follows that $x$ is algebraic over $k(u)$. If $u$ were algebraic over $k$, then $x$ would also be algebraic over $k$, which is clearly not the case; hence $u$ is transcendental over $k$.

The left-hand side of (3), as a polynomial in $X$, is irreducible over $k(u)$, for if it could be factorized over $k(u)[X]$, then by Gauss's lemma, 3.7.1 of A.2, it can be factorized over $k[u, X]$. Since it is linear in $u$, one factor must be independent of $u$ and the other linear in $u$, but this is impossible, because $f, g$ are coprime. So (3) is the irreducible polynomial for $x$ over $k(u)$. It has degree $\deg u$ and so (2) follows. ■

The case $n = 1$ is of particular importance; it applies iff $u$ generates $k(x)$, so we obtain

COROLLARY 2.2 *Let $u$ be an element of a simple transcendental extension $k(x)$. Then $k(u) = k(x)$ if and only if*

$$u = \frac{ax + b}{cx + d},$$

*where $a, b, c, d \in k$ and $ad - bc \neq 0$.*

Here the condition on the constants just ensures that $u \notin k$. ■

This result makes it easy to determine the automorphisms of $k(x)$ over $k$. Any such automorphism must map $x$ to a generator, and conversely, any mapping from $x$ to a generator determines a unique automorphism. Thus we have

PROPOSITION 2.3 *The group of automorphisms of a simple transcendental extension $k(x)/k$ is the group* $\mathbf{PGL}_2(k)$ *of projective transformations*

$$x \mapsto \frac{ax+b}{cx+d}, \qquad \text{where } a, b, c, d \in k \text{ and } ad-bc \neq 0. \quad \blacksquare \tag{4}$$

We observe that $\mathbf{PGL}_2(k)$ is the quotient of the general linear group $\mathbf{GL}_2(k)$ by its centre; each matrix $A$ in $\mathbf{GL}_2(k)$ defines a transformation (4), which is the identity precisely when $A$ lies in the centre, i.e. $A = \lambda I$ for some $\lambda \in k^\times$. Thus the constants in (4) are determined up to a constant multiple; this means that we can normalize our transformations, at least in the case $k = \mathbf{C}$, by requiring that $ad - bc = 1$.

Prop. 2.1 has an important generalization:

THEOREM 2.4 (Lüroth's theorem) *Let $k(x)$ be a simple transcendental extension. Then every field $E$ such that $k \subset E \subseteq k(x)$ is of the form $E = k(u)$, where $u$ is transcendental over $k$.*

*Proof.* Since $E \supset k$, there exists $u \in E \setminus k$; in Prop. 2.1 we saw that $x$ is algebraic over $k(u)$, hence $x$ is algebraic over $E$. Let the minimal polynomial for $x$ over $E$ be

$$\varphi(X) = X^n + u_1 X^{n-1} + \cdots + u_n, \qquad \text{where } u_i \in E.$$

Here the $u_i$ are rational functions of $x$ and on multiplying by a lowest common denominator we obtain a polynomial

$$\Phi(X, x) = v_0(x)X^n + v_1(x)X^{n-1} + \cdots + v_n(x), \tag{5}$$

in which the coefficients $v_i(x)$ are polynomials in $x$ without a common (non-constant) factor, i.e. $\Phi$ is primitive, as a polynomial in $X$. Since $x$ is transcendental over $k$, the $u_i = v_i/v_0$ are not all in $k$; we pick $u_j \notin k$, so that $m = \deg u_j > 0$. By Prop. 2.1, $x$ has degree $m$ over $k(u_j)$, while its degree over $E$ is $n$, thus

$$m = [k(x):k(u_j)] = [k(x):E][E:k(u_j)] = n[E:k(u_j)],$$

and it follows that

$$[E:k(u_j)] = \frac{m}{n}. \tag{6}$$

Now the proof will be complete if we can show that $m = n$, for then $E = k(u_j)$,

by (6). Let us write $u_j$ in lowest terms as

$$u_j = \frac{f(x)}{g(x)}, \tag{7}$$

where $f, g$ are coprime and $g$ is monic, while $m = \max\{\deg f, \deg g\}$, by the definition of $m$. The polynomial $f(X) - u_j g(X)$ does not vanish identically, but has $x$ as a zero, and so is divisible by $\varphi(X)$. If we replace $u_j$ by its value from (7) and multiply up by $g$, we obtain by (5),

$$f(X)g(x) - g(X)f(x) = Q(X, x)\Phi(X, x), \tag{8}$$

where $Q$ is a polynomial in $X$ and by the primitivity of $\Phi$, also one in $x$. The left-hand side of (8) has degree $m$ in $x$, and on the right $\Phi$ has degree at least $m$ in $x$. Hence $Q$ must be independent of $x$; if $Q$ has positive degree in $X$, it has a zero $\alpha$ in an algebraic extension of $k$. Thus $f(\alpha)g(x) - f(x)g(\alpha) = 0$, and $f(\alpha), g(\alpha) \neq 0$, because $f, g$ are coprime, so

$$\frac{f(\alpha)}{g(\alpha)} = \frac{f(x)}{g(x)} = u_j,$$

hence $u_j$ is algebraic over $k$, which is a contradiction. This shows $Q$ to be independent of $X$ as well. Now the left-hand side of (8) has degree $m$ in $X$, while the right-hand side has degree $n$ in $X$, by (5). Hence $m = n$ and so $E = k(u_j)$, as we wished to show. ■

For fields in two variables $k(x, y)$ the analogue need not hold, though it does hold when $k$ is algebraically closed. But for more than two variables the result fails even over algebraically closed fields (cf. Deligne (1973)).

Let us return to the subject of Prop. 2.3. By Galois theory, any finite subgroup of $\mathbf{PGL}_2(k)$ corresponds to a subfield $F$ of $k(x)$ such that $[k(x):F]$ is finite. In the special case $k = \mathbf{C}$ any such subfields may be determined as follows*.

By stereographic projection the unit-sphere in real 3-space $\mathbf{R}^3$ corresponds to the complex line, i.e. the plane, with the North pole corresponding to the point at infinity. Any rotation in $\mathbf{R}^3$ corresponds to a fractional linear transformation (of the type (4)) of the complex line with two fixed points, corresponding to the axis of rotation. Taking the fixed points to be $0, \infty$, we obtain the form

$$x' = \kappa x, \qquad \text{where } |\kappa| = 1, \tag{9}$$

for our transformation; this is known as an *elliptic* transformation of the line. In the form (4) (with $ad - bc = 1$) an elliptic transformation is characterized by the condition that $a + d$ is real and $|a + d| < 2$; for this is an invariant condition

*The remarks that follow assume some acquaintance with the rudiments of complex function theory; they will not be used later in this work.

which just ensures that $|\kappa| = 1$ in (9). Conversely, an elliptic transformation corresponds to a rotation, as we see by transforming the fixed points to 0, $\infty$ respectively. Now any transformation (4) of finite order is elliptic. To see this we observe that the fixed points of (4) are given by

$$\frac{x}{y} = \frac{ax+by}{cx+dy}, \qquad \text{i.e. } A\begin{pmatrix} x \\ y \end{pmatrix} = \lambda\begin{pmatrix} x \\ y \end{pmatrix}, \qquad \text{where } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

If $A^n = I$, then the minimal equation of $A$ has distinct roots, so $A$ is similar to a diagonal matrix, which gives the fixed points 0, $\infty$ on the line. Moreover, the roots are roots of 1, reciprocal and so conjugate over $\mathbf{R}$, hence $a+d$ is real and $|a+d| < 2$.

Thus it follows that any finite subgroup of $\mathbf{PGL}_2(\mathbf{C})$ corresponds to a finite group of rotations in $\mathbf{R}^3$. But we know what these groups are, from **3.6** of A.1: the only finite rotation groups in $\mathbf{R}^3$ are the cyclic and dihedral groups, and the tetrahedral, octahedral and icosahedral groups. The latter three groups are isomorphic to $A_4, S_4$ and $A_5$ respectively; thus the only non-abelian simple group that occurs is the alternating group of degree 5 (for more details, see Klein [1884], Weber [1894]).

By contrast, there are many different finite extensions of $k(x)$; they form the field of study of the theory of algebraic functions of one variable, of importance in function theory as well as algebraic geometry (cf. Chevalley [1951], Cohn [1991]).

## Exercises

(1) Show that if $\varphi, \psi$ are any rational functions of degrees $m, n$ (as defined after (1)), then $\varphi(\psi(x))$ is of degree $mn$.

(2) Show that any simple subextension of a purely transcendental extension is purely transcendental.

(3) Let $k(x)/k$ be a simple transcendental extension. Show that for any $n \geqslant 1$ there exists an extension $E$ of $k(x)$ of degree $n$, such that $E$ is again simple transcendental over $k$. Use this result to prove the existence of an infinite algebraic extension $F$ of $k(x)$ such that any finitely generated subextension of $F$ is simple transcendental over $k$.

(4) Show that $\alpha: x \mapsto x^{-1}$ and $\beta: x \mapsto 1-x$ generate an automorphism group $\Gamma \cong S_3$ of $k(x)$ and that the fixed field is $k((x^2-x+1)^3/x^2(x-1)^2)$.

(5) Let $k$ be a finite field of $q$ elements. Show that $\mathrm{Gal}(k(x)/k)$ has order $q^3 - q$ and fixed field $k((x^{q^2}-x)^{q+1}/(x^q-x)^{q^2+1})$.

(6) Let $k$ be a field and $x$ an indeterminate. Show that a $k$-subalgebra $R$ of $k[x]$ is of the form $k[y]$ or $k$ iff $R$ is integrally closed. (Hint: use Lüroth's theorem to show that

$R$ has a field of fractions of the form $k(y)$ and prove that $R$ is contained in all valuation subrings of $k(y)$ except at most one. Now use the characterization of integrally closed rings in Th. 8.4.4 of A.2.)

(7) (E. Noether) Show that a subfield $E$ of $k(x)$ has a generator in $k[x]$ iff $E \cap k[x] \supset k$. (Hint: use Ex. (6).)

## 5.3 Separable and $p$-radical extensions

Separable extensions were defined in A.2, **3.4**; we shall now take a closer look at the inseparable case, bearing in mind that it only arises in prime characteristic. Over a field of characteristic 0 every algebraic extension is separable.

We have seen in A.2, Th. 3.5.4 that finite separable extensions may be characterized as finite extensions $E/k$ possessing exactly $[E:k]$ $k$-homomorphisms into a normal closure of $E/k$. Here we may instead of a normal closure take an algebraic closure of $E$, for the latter contains such a normal closure.

Let us define, for any finite extension $E/k$, the *separable degree* $[E:k]_s$ as the number of $k$-homomorphisms of $E$ into an algebraic closure of $E$. Then we have, by A.2, Th. 3.5.4,

$$[E:k]_s \leqslant [E:k],$$

with equality holding precisely when $E/k$ is separable. Our first observation is that the separable degree is again multiplicative:

PROPOSITION 3.1 *Let $F \supseteq E \supseteq k$ be algebraic extensions. Then*

$$[F:k]_s = [F:E]_s[E:k]_s, \tag{1}$$

*whenever either side is finite.*

*Proof.* Let $\bar{F}$ be an algebraic closure of $F$; since $F/E$ is algebraic, $\bar{F}$ is also an algebraic closure of $E$. We have $s$ $E$-homomorphisms of $F$ into $\bar{F}$, where $s = [F:E]_s$; it follows that any isomorphism $\varphi: E \to E'$ can be extended in just $s$ ways to a homomorphism of $F$ into an algebraic closure of $E'$, for if we use $\varphi$ to identify $E$ with $E'$, the problem is reduced to extending the identity mapping on $E$, i.e. to finding $E$-homomorphisms of $F$, and we saw that there are just $s$ such extensions. We now apply this argument to a $k$-homomorphism $\varphi$ of $E$ into $\bar{F}$. There are $[E:k]_s = r$ such mappings, and each can be extended in $s$ ways, giving $rs$ $k$-homomorphisms of $F$ into $\bar{F}$. But every $k$-homomorphism of $F$ into $\bar{F}$ can be obtained in this way; thus there are just $rs$ of them, and so (1) follows whenever either side is finite. ■

In contrast to the separable extensions we now consider the *p-radical* or *purely inseparable* extensions, defined as algebraic extensions $E/k$ of prime

characteristic $p$, such that there is only *one* $k$-homomorphism of $E$ into an algebraic closure of $E$. In particular, this means that every element $\alpha$ of $E$ has only one conjugate over $k$, for if $\alpha$ had a conjugate $\alpha' \neq \alpha$, then the $k$-isomorphism $k(\alpha) \cong k(\alpha')$ in which $\alpha$ maps to $\alpha'$ could be extended to a $k$-homomorphism different from 1 on $E$, which contradicts the hypothesis.

Let $E/k$ be any $p$-radical extension, where $E \supset k$. This means that $p = \text{char } k$ is a prime and the minimal polynomial $f$ of any $\alpha \in E$ over $k$ has only one zero. If its degree is $n = mq$, where $q = p^r$ and $p \nmid m$, let us write $\alpha^q = a$. The minimal polynomial for $\alpha$ is $f = (x-\alpha)^n = (x^q - a)^m = x^{mq} - max^{(m-1)q} + \cdots$; it follows that $a \in k$ and so $m = 1$ and the minimal polynomial for $\alpha$ is of the form

$$f = (x-\alpha)^q = x^q - a, \qquad \text{where } \alpha^q = a \in k \text{ and } q = p^r. \tag{2}$$

An element $\alpha$ of $E$ is said to be *p-radical* or *purely inseparable* over $k$ if $\alpha^q \in k$, where $q = p^r$ ($p = \text{char } k$). Here the least value of $r$ is sometimes called the *height* of $\alpha$. By (2) every element of a $p$-radical extension is $p$-radical; the converse also holds, in a slightly sharper form:

PROPOSITION 3.2 *An extension $E/k$ is p-radical if and only if it is generated by p-radical elements.*

*Proof.* We have seen that a $p$-radical extension consists entirely of $p$-radical elements. Conversely, assume that $E$ is generated over $k$ by a set $X$ of elements that are $p$-radical over $k$. In any $k$-homomorphism of $E$, any $x \in X$ must map to a root of the minimal equation for $x$ and hence remains fixed. Therefore $E = k(X)$ has only one $k$-homomorphism into an algebraic closure of $E$, so $E/k$ is $p$-radical, as claimed. ■

Our next task is to see how a general algebraic extension is built up from separable and $p$-radical extensions. We shall need:

LEMMA 3.3 *Let $\alpha$ be algebraic over a field $k$, where $\text{char } k = p$. Then for some $q = p^r$, $\alpha^q$ is separable over $k$.*

*Proof.* We shall use induction on the degree of $\alpha$ over $k$. Let $f$ be the minimal polynomial of $\alpha$ over $k$. If $f$ has distinct zeros, it is separable and the result follows. Otherwise $f$ has a repeated zero and so is not prime to its derivative $f'$. Since $f$ is irreducible over $k$, this means that $f \mid f'$, and by comparing degrees we see that this can happen only when $f' = 0$. If $f = \sum a_i x^i$, this means that $ia_i = 0$; hence $a_i = 0$ for $p \nmid i$, and so $f(x) = g(x^p)$ for some polynomial $g$ of lower degree than $f$. Now $\alpha^p$ is a zero of $g$, hence by induction, $(\alpha^p)^{p^r} = \alpha^{p^{r+1}}$ is separable over $k$, as we wished to show. ■

Let $E/k$ be any algebraic extension. Then the elements of $E$ that are separable

over $k$ form a subfield, by A.2, Th.3.5.4, which will be denoted by $k_s$ and called the *separable closure* of $k$ in $E$. We note that $E$ is $p$-radical over $k_s$, for if $\alpha \in E$, then for some $q = p^r$, $\alpha^q$ is separable over $k$, by Lemma 3.3, and so $\alpha^q$ lies in $k_s$. Thus we have $[E:k_s]_s = 1$; now if $E/k$ is finite, then $[E:k]_s = [E:k_s]_s[k_s:k]_s$ by (1), hence

$$[k_s:k] = [E:k]_s. \tag{3}$$

We shall define the degree of inseparability or simply the *inseparable degree* of $E/k$ as

$$[E:k]_i = [E:k_s]. \tag{4}$$

Like the total and the separable degree, it is again multiplicative. This is clear from the definition, which may be written

$$[E:k] = [E:k]_s[E:k]_i. \tag{5}$$

We remark that $[E:k]_i$ must be a power of $p$, if finite, since it is the degree of the $p$-radical extension $E/k_s$. The situation may be summed up as follows:

THEOREM 3.4 *Every algebraic extension $E/k$ may be obtained by taking a separable extension followed by a $p$-radical extension: $k \subseteq k_s \subseteq E$ where $p = \operatorname{char} k$. Moreover, when $E/k$ is finite, then*

$$[E:k]_s = [k_s:k], \qquad [E:k]_i = [E:k_s]. \quad \blacksquare \tag{6}$$

The set of all $p$-radical elements of an algebraic extension $E/k$ also forms a subfield $k_i$ say, called the *perfect closure* of $k$ in $E$. For if $\alpha, \beta \in k_i$, say $\alpha^{p^r}, \beta^{p^s} \in k$ and $r \leqslant s$ say, then $(\alpha - \beta)^{p^s} = \alpha^{p^s} - \beta^{p^s} \in k$ and of course $(\alpha\beta)^{p^s} = \alpha^{p^s}\beta^{p^s} \in k$. Strictly speaking, $k_i$ should be described as the *relative* perfect closure of $k$ in $E$, to distinguish it from the following absolute notion.

Let $k$ be any field of prime characteristic $p$. A *perfect closure* of $k$ is a field $\hat{k}$ containing $k$ as subfield such that $\hat{k}$ is perfect and $\hat{k}/k$ is $p$-radical.

PROPOSITION 3.5 *Every field $k$ of prime characteristic $p$ has a perfect closure, unique up to isomorphism.*

*Proof.* We define $\hat{k}$ as the relative perfect closure of $k$ in its algebraic closure. Every element of $\hat{k}$ is $p$-radical over $k$, hence $\hat{k}/k$ is $p$-radical, and $\hat{k}$ is perfect because every element of $\hat{k}$ has a $p$th root in the algebraic closure of $k$, and so this root also lies in $\hat{k}$. The uniqueness is clear: If $E, F$ are two perfect closures of $k$ and $\alpha \in E$, then $\alpha^q \in k$ for some $q = p^r$ and there is a unique element $\alpha' \in F$ such that $\alpha'^q = \alpha^q$. The correspondence $\alpha \mapsto \alpha'$ is a homomorphism, for if $\beta \mapsto \beta'$, then for a high enough power $q = p^r$ we have $\alpha^q, \beta^q \in k$ and so $(\alpha + \beta)^q = \alpha^q + \beta^q = \alpha'^q + \beta'^q = (\alpha' + \beta')^q$, therefore $\alpha + \beta \mapsto \alpha' + \beta'$ and similarly $\alpha\beta \mapsto \alpha'\beta'$. Clearly it is bijective and preserves 1, so it is an isomorphism and $E \cong F$. $\blacksquare$

This result may also be proved directly, without using the notion of algebraic closure of $k$ (cf. Ex. (3)). The relation of the perfect closure to its relative version should be clear: if $\hat{k}$ denotes the perfect closure of a field $k$, and $E$ is an extension field of $k$, then the relative perfect closure of $k$ in $E$ is the largest subfield of $E$ which is $k$-isomorphic to a subfield of $\hat{k}$.

Let $E/k$ be an algebraic extension. We have in $E$ the separable closure $k_s$ and the perfect closure $k_i$ of $k$. As we saw in Th. 3.4, $E/k_s$ is then $p$-radical, so $E$ can be obtained by a separable extension followed by a $p$-radical extension. However, $E/k_i$ need not be separable, so the order of the extensions cannot in general be reversed (cf. Ex. (4)). In the special case where it can, we have a more precise description of the extension. First we prove an auxiliary result. We write $E^p = \{x^p | x \in E\}$.

LEMMA 3.6 *Let $E/k$ be a finite separable extension of characteristic $p$ and let $u_1, \ldots, u_m$ be a basis. Then $u_1^q, \ldots, u_m^q$ is again a basis, for any $q = p^r, r \geqslant 1$.*

*Proof.* Since $E \supseteq E^q k \supseteq k$, it follows that $E$ is $p$-radical as well as separable over $E^q k$, hence $E = E^q k$. Therefore $u_1^q, \ldots, u_m^q$ again spans $E$ over $k$, and so is a basis. ■

We can now achieve our aim, to prove that a $p$-radical extension followed by a separable extension always decomposes as a tensor product.

THEOREM 3.7 *Given extensions $k \subseteq E \subseteq F$, where $E/k$ is $p$-radical and $F/E$ is separable, we have*

$$F \cong E \otimes_k k_s. \tag{7}$$

*Proof.* In proving (7) we may assume $F/k$ to be finite; for when that case has been established, the general case follows because the algebraic extensions $F/k$, $E/k$ can be written as ascending unions of their finite subextensions, and tensor products preserve ascending unions.

Further, if we can find a $k$-basis for $k_s$ which is also an $E$-basis for $F$, then the natural homomorphism $E \otimes_k k_s \to F$ is injective, by the independence property of the tensor product. Now the image $Ek_s$ in $F$ is a $k$-algebra, finite-dimensional over $k$ and without zero-divisors, and therefore a field. Moreover, $F/E$ is separable and $F/k_s$ $p$-radical, so $F$ is both separable and $p$-radical over $Ek_s$ and therefore $F = Ek_s$. So it only remains to find a $k$-basis of $k_s$ which is an $E$-basis for $F$.

Let $u_1, \ldots, u_m$ be any $k$-basis of $k_s$; then $u_1, \ldots, u_m$ span $F$ over $E$, by what we have seen, and we must show that they are linearly independent over $E$. Suppose then that $\sum a_i u_i = 0$ $(a_i \in E)$. Since $E/k$ is $p$-radical, $a_i^q \in k$ for some $q = p^r$, so $\sum a_i^q u_i^q = 0$, but the $u_i^q$ are linearly independent over $k$, by Lemma 3.6, hence $a_i^q = 0$, so $a_i = 0$ and this proves that the $u_i$ are linearly independent over $E$. ■

It is clear that in Th. 3.7, $E$ is just the perfect closure of $k$ in $F$. The result can be applied to describe the structure of normal extensions. We recall that an extension $E/k$ is called *normal* (or sometimes *quasi-Galois*) if it is algebraic and every irreducible polynomial over $k$ which has a zero in $E$ splits completely over $E$. We shall need a generalization of a property of normal extensions (A.2, Cor. 3.2.5) to the infinite case.

PROPOSITION 3.8 *Let $\Omega/k$ be a normal extension. Then any isomorphic subextensions are conjugate, i.e. any isomorphism $\varphi: E \to F$ of subextensions can be extended to a $k$-automorphism of $\Omega$.*

*Proof.* Consider the family of all $k$-homomorphisms $\varphi_\lambda: E_\lambda \to \Omega$, where $E_\lambda \supseteq E$ and $\varphi_\lambda | E = \varphi$. These homomorphisms can be partially ordered by writing $(\varphi_\lambda, E_\lambda) \leqslant (\varphi_\mu, E_\mu)$ if $E_\lambda \subseteq E_\mu$ and $\varphi_\mu | E_\lambda = \varphi_\lambda$. This family clearly is inductive; by Zorn's lemma there is a maximal element: $\varphi_0: E_0 \to \Omega$, and we must show that $E_0 = \Omega$. If $E_0 \neq \Omega$, take $\alpha \in \Omega \setminus E_0$, let $f$ be the minimal polynomial of $\alpha$ over $E_0$ and $g$ its minimal polynomial over $k$. Then $g\varphi_0 = g$ and so both $f, f\varphi_0$ divide $g$. Since $f$ has the zero $\alpha$ in $\Omega$, it follows that $g$ has a zero in $\Omega$; by normality $g$ splits completely over $\Omega$, hence so does $f\varphi_0$. We can therefore extend $\varphi_0$ to $E_0(\alpha)$ by mapping $\alpha$ to a zero of $f\varphi_0$. This contradicts the maximality of $\varphi_0$, hence $E_0 = \Omega$ and $\varphi_0$ is the desired automorphism of $\Omega$ over $k$. ■

This result can be used to characterize normal extensions:

COROLLARY 3.9 *An algebraic extension $E/k$ is normal if and only if every $k$-automorphism of any field $\Omega$ containing $E$ maps $E$ into itself.*

*Proof.* The condition is necessary because when $E/k$ is normal, then every element of $E$ satisfies an irreducible equation over $k$ whose roots are permuted by any $k$-automorphism of $\Omega$. Conversely, when it holds, let $f$ be any irreducible polynomial over $k$ which has a zero $\alpha$ in $E$. Take a normal extension $\Omega/k$ containing $E/k$ such that $f$ splits completely in $\Omega$ (e.g. a normal closure of the splitting field of $f$ over $E$). If $\beta$ is any zero of $f$ in $\Omega$, then $\alpha \mapsto \beta$ under an isomorphism $k(\alpha) \cong k(\beta)$, hence $\alpha$ and $\beta$ are conjugate in $\Omega$, by Prop. 3.8, and so $\beta \in E$. Thus $f$ splits over $E$ and this shows $E/k$ to be normal. ■

We now come to the structure theorem for normal extensions. For simplicity we shall limit ourselves to the finite case; the general case is quite analogous but requires some ideas from **5.6**.

THEOREM 3.10 *Let $E/k$ be any finite normal extension and denote by $k_s$ the separable closure and by $k_i$ the perfect closure of $k$ in $E$. Then $k_s$ is a Galois*

*extension of k and we have*

$$E \cong k_s \otimes_k k_i.$$

*Proof.* Any $k$-automorphism of an extension of $E$ maps $E$ into itself, hence it maps $k_s$ into itself. Therefore $k_s/k$ is normal, and being also separable it is Galois (A.2, Prop. 3.6.1). Now the fixed field under all automorphisms is $k_i$, so the result follows from Th. 3.7. ■

### Exercises

(1) Show that a finite extension $E/k$ of characteristic $p$ is separable iff $E = E^p k$.

(2) Show that for a given extension $E/k$, the perfect closure of $k$ in $E$ need not be perfect. Under what conditions on $E$ is the perfect closure of every subfield perfect?

(3) Let $k$ be any field of prime characteristic $p$. Show that there is a tower of fields $k = k_0 \subseteq k_1 \subseteq \cdots$, such that the $p$th power mapping $x \mapsto x^p$ induces an isomorphism $k_r \to k_{r-1}$, and that the union $\cup k_r$ is the perfect closure of $k$.

(4) Let $p$ be an odd prime and $k = \mathbf{F}_p(s, t)$, where $s, t$ are indeterminates. Show that if $\alpha$ is a root of the equation

$$x^{2p} + 2sx^p + t = 0,$$

then the perfect closure of $k$ in $k(\alpha)$ is $k$ itself, but that $k(\alpha)/k$ is not separable (such extensions are sometimes called *exceptional*).

(5) Show that the rule which assigns to each field its perfect closure can be regarded as the left adjoint of a forgetful functor (in the category of fields and homomorphisms), but that the algebraic closure (or the separable algebraic closure) cannot be so regarded. (Hint: observe that the perfect closure is unique, up to a *unique* isomorphism.)

(6) Let $\Omega/k$ be a normal extension and $E, F$ any subfields of $\Omega$ that are conjugate. Show that any normal extension of $k$ containing $E$ also contains $F$.

(7) Show that if $F/k$ is generated by finitely many algebraic elements, all but at most one separable over $k$, then $F/k$ is simple. (Hint: use Steinitz' criterion, A.2, Th. 3.9.3: a finite extension is simple iff the number of intermediate fields is finite.)

(8) For a finite $p$-radical extension $F/k$ define the *height* as the least exponent $r$ such that $F^{p^r} \subseteq k$. Show that for any finite extension $F/k$ there exists a $p$-radical extension $E$ of $k$, of the same height as $F$ over $k_s$, such that $F/k$ is contained in $k_s \otimes_k E$.

## 5.4 Derivations

It is well known and easily checked that a polynomial $f$ has no repeated factor iff it is prime to $f'$, its formal derivative. From this fact it is not hard to derive

a criterion for the separability of simple extensions in terms of derivations, and it is reasonable to expect that derivations can also be used to study more general extensions. In a ring $R$ generated by a set $X$, any derivation is uniquely determined by its values on $X$, but usually not every map of $X$ into $R$ can be extended to a derivation (cf. **2.2**). One of our main tasks in this section is to find conditions for this to be possible, for certain field extensions.

The derivations used here are always (1, 1)-derivations, in the terminology of **2.2**, i.e. they are mappings $a \mapsto da$ satisfying

$$d(ab) = da \cdot b + a \cdot db,$$

besides being additive. We recall that the kernel of a derivation $d$ is a subring, the ring of *constants* for $d$, and a derivation of a $K$-algebra is said to be *over* $K$ if $K \subseteq \ker d$. When a derivation is defined on a field, the set of constants actually forms a subfield.

We begin by noting how a derivation on an integral domain extends to the field of fractions.

PROPOSITION 4.1 *Let $R$ be an integral domain with field of fractions $K$. Then any derivation $d$ on $R$ can be extended in just one way to a derivation of $K$, again written $d$, and given by the formula*

$$d\left(\frac{a}{b}\right) = \frac{da \cdot b - a \cdot db}{b^2}. \tag{1}$$

*Proof.* The mapping $a \mapsto \begin{pmatrix} a & da \\ 0 & a \end{pmatrix}$ of $R$ into $\begin{pmatrix} K & K \\ 0 & K \end{pmatrix}$ maps each non-zero element of $R$ to an invertible element and so extends to a unique homomorphism $u \mapsto \begin{pmatrix} u & du \\ 0 & u \end{pmatrix}$. Hence $u \mapsto du$ is a derivation of $K$, whose precise form (1) follows by evaluating $\begin{pmatrix} a & da \\ 0 & a \end{pmatrix} \cdot \begin{pmatrix} b & db \\ 0 & b \end{pmatrix}^{-1}$. ■

The chain rule, familiar from calculus, can be extended to arbitrary derivations, where it takes the following form. Let $R = K[x_1, \ldots, x_n]$ be a polynomial ring in $n$ variables over a ring $K$. We shall write $D_1$ or $\partial/\partial x_1$ for the derivation with respect to $x_1$; this is the derivation over $K$ which maps $x_1$ to 1 and the other variables to 0. Since $R$ may be regarded as a polynomial ring in $x_1$ over $K[x_2, \ldots, x_n]$, this derivation certainly exists. Now $D_i = \partial/\partial x_i$ $(i = 2, \ldots, n)$ is defined similarly. Next consider an arbitrary commutative ring $R$, generated by $n$ elements $a_1, \ldots, a_n$ over $K$. Any element $c$ of $R$ may be written (possibly in more than one way) as $c = f(a_1, \ldots, a_n)$, where $f$ is a polynomial with coefficients in $K$. If $d$ is a derivation of $R$ and $f^d$ denotes the polynomial obtained

from $f$ by applying $d$ to the coefficients, then the chain rule asserts

$$df(a_1,\dots,a_n) = f^d(a_1,\dots,a_n) + \sum_1^n D_i f(a_1,\dots,a_n)\cdot da_i. \tag{2}$$

Here $D_i f$ is understood to mean formal differentiation with respect to the $i$th argument in $f$. For the case when $f$ is a monomial, (2) follows by induction on the degree, and the general case then follows by linearity. In fact the result holds even for rational functions, as we see by writing $f = g/h$ and using the formula (1), but we shall not need this fact. We now establish conditions under which a derivation can be extended; the reader should compare this with the parallel discussion in **2.2**.

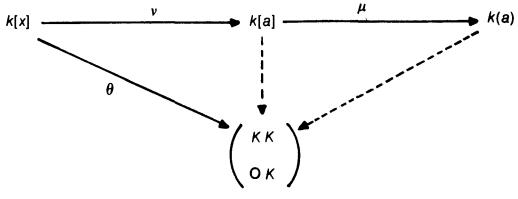
THEOREM 4.2 *Let $E = k(a_1,\dots,a_n)$ be a finitely generated field extension of $k$, denote by $\mathfrak{a}$ the ideal in $k[x_1,\dots,x_n]$ of all polynomials $f(x_1,\dots,x_n)$ vanishing for $x_i = a_i$ and let $\{f_\lambda(x)\}$ be a generating set for $\mathfrak{a}$. Given any derivation $d$ on $k$ and any $u_1,\dots,u_n \in E$, the derivation $d$ can be extended to a derivation $d^*$ of $E$ such that $d^* a_i = u_i$ $(i = 1,\dots,n)$ if and only if*

$$f_\lambda^d(a) + \sum_i D_i f_\lambda(a)\cdot u_i = 0. \tag{3}$$

*Moreover, $d^*$ is then uniquely determined and is given by*

$$d^* f(a) = f^d(a) + \sum D_i f(a)\cdot u_i. \tag{4}$$

*Proof.* Consider the diagram




where $v$ is the reduction mod $\mathfrak{a}$ and $\mu$ is the inclusion, while $\theta$ maps $x_i$ to $\begin{pmatrix} a_i & u_i \\ 0 & a_i \end{pmatrix}$. The subring of $\begin{pmatrix} K & K \\ 0 & K \end{pmatrix}$ generated by these elements over $k$ is commutative, and hence this mapping extends to a homomorphism $\theta: f(x) \mapsto \begin{pmatrix} f(a) & d'f(a) \\ 0 & f(a) \end{pmatrix}$, where $d'$ is the unique $(v, v)$-derivation extending $d$ such that $a_i \mapsto u_i$. Now the chain rule shows that

$$d'f(a) = f^d(a) + \sum_i D_i f(a)\cdot u_i.$$

Hence if an extension $d^*$ is to exist, we must have $d'f = 0$ for all $f \in \mathfrak{a}$ and (3) follows. Conversely, when (3) holds for a generating set $\{f_\lambda\}$ of $\mathfrak{a}$, then $\ker \theta \supseteq \mathfrak{a}$, hence by the factor theorem (Th. 1.2.4) we obtain the vertical dotted arrow, and now by Prop. 4.1 we obtain the slanting arrow on the right. This is a homomorphism of the form $c \mapsto \begin{pmatrix} c & d^*c \\ 0 & c \end{pmatrix}$, and $d^*$ is the required derivation. ■

Consider in particular a simple extension $k(x)$. If $x$ is transcendental over $k$, there is no condition, so any derivation of $k$ has a unique extension mapping $x$ to a prescribed element of $k(x)$. If $x$ is algebraic over $k$, let $f$ be its minimal polynomial; the condition that there should exist an extension of $d$ mapping $x$ to $u \in K$ is that

$$f^d(x) + f'(x)u = 0. \tag{5}$$

Now two cases can arise: if $x$ is separable over $k$, then $f'(x) \neq 0$ and we see that $d$ has an extension only if $u = -f'(x)^{-1} f^d(x)$, and with this value the extension is unique. If $x$ is inseparable, then $f'(x) = 0$, and (5) holds precisely when $f^d(x) = 0$, and $u$ can then be chosen arbitrarily. But $f^d$ is a polynomial of lower degree than $f$, for the highest coefficient of $f$ is 1 and so maps to 0. Hence $f^d(x) = 0$ only when $f^d$ vanishes identically, i.e. when all the coefficients of $f$ are constant. We can sum up the result as

THEOREM 4.3 *Let $k$ be a field with a derivation $d$ and let $x$ be an element of some extension $E$ of $k$. Then*

(i) *if $x$ is transcendental over $k$ and $u$ is any element of $E$, then $d$ extends to a unique derivation of $k(x)$ into $E$ such that $dx = u$;*

(ii) *if $x$ is separable over $k$, with minimal polynomial $f$, then $d$ can be extended in just one way to a derivation of $k(x)$ into $E$, and the value of $dx$ then satisfies*

$$f^d(x) + f'(x) \cdot dx = 0;$$

(iii) *if $x$ is inseparable over $k$, then $d$ can be extended to $k(x)$ if and only if the minimal polynomial for $x$ has constant coefficients, and when this is so, there is just one extension for each value of $dx$, where $dx$ can be chosen arbitrarily.* ■

By induction we obtain

COROLLARY 4.4 *Let $E/k$ be a finitely generated field extension. Then there is no non-zero derivation of $E$ over $k$ if and only if $E$ is separably algebraic over $k$.* ■

This leads to another useful criterion for separability:

PROPOSITION 4.5 *Given a finitely generated extension $E = k(x_1, \ldots, x_n)$ over*

*k, if there are n polynomials in n variables over k such that* $f_i(x_1, \ldots, x_n) = 0$ $(i = 1, \ldots, n)$ *and* $\det(\partial f_i/\partial x_j) \neq 0$, *then E/k is separably algebraic.*

For by Cor. 4.4 we need only show that every derivation of $E$ over $k$ is zero. Let $d$ be a derivation of $E/k$; then $df_i(x) = 0$, hence by the chain rule,

$$\sum D_j f_i(x) \cdot dx_j = 0.$$

Since $\det(D_j f_i) \neq 0$ by hypothesis, these equations have only the zero solution, hence $dx_j = 0$ and $d$ must be zero on $E$. ■

If a field $E$ is generated by $x_1, \ldots, x_n$ over $k$, as in Prop. 4.5, then the universal derivation module $\Omega_k(E)$ is spanned by the $dx_i$. This leads to the following criterion for algebraic independence:

THEOREM 4.6 *Let E/k be a field extension in characteristic* 0, *and let* $(x_\lambda)$ *be any family of elements of E. Then* (i) *the* $x_\lambda$ *are algebraically dependent if and only if the* $dx_\lambda$ *are linearly dependent over E,* (ii) $E/k(x_\lambda)$ *is algebraic if and only if the* $dx_\lambda$ *span* $\Omega_k(E)$ *as E-space,* (iii) $(x_\lambda)$ *is a transcendence basis for E over k if and only if the* $dx_\lambda$ *form a basis for* $\Omega_k(E)$ *over E.*

This follows from the fact that any polynomial relation $f(x_1, \ldots, x_n) = 0$ corresponds to a relation $\sum D_i f(x) \cdot dx_i = 0$. ■

## Exercises

(1) Let $p$ be an odd prime and $k$ a field of characteristic $p$. Show that any extension generated by $x, y$ such that $x^p + ay^2 = 1, bx^2 + y^p = 1$ $(a, b \in k)$ is separably algebraic iff $ab \neq 0$.

(2) Complete the details in the proof of Prop. 4.1.

(3) Use Prop. 4.5 to establish the transitivity of separably algebraic extensions.

(4) Prove Cor. 4.4.

(5) Let $E/k$ be a separably algebraic extension. Given an $E$-space $V$, show that any derivation of $k$ into $V$ has a unique extension to a derivation of $E$ into $V$. (Hint: adapt the proof of Th. 4.5 and use induction on the degree).

(6) Find an extension of Th. 4.6 to the case of prime characteristic.

(7) Apply Th. 4.6 to give another proof that the transcendence degree (in characteristic 0) is an invariant of the extension (cf. Th. 1.1). What can be said in prime characteristic?

(8) Let $E/k$ be a finitely generated extension of characteristic $p$. Show that an element of $E$ lies in $E^p k$ iff it is constant for every derivation of $E/k$.

## 5.5 Linearly disjoint extensions

Let $k$ be a field and $\Omega$ an algebraic extension; in this section we examine the way in which different extensions of $k$ in $\Omega$ interact. All our extensions will be algebraic and we may take $\Omega$ to be an algebraic closure of $k$ (cf. **3.3** of A.2 or **5.6** below), but this is not essential.

Given any subalgebras $A, B$ of $\Omega$, we have a homomorphism

$$A \otimes_k B \to \Omega, \tag{1}$$

obtained by mapping $\sum a_i \otimes b_i \mapsto \sum a_i b_i$. If (1) is injective, $A$ and $B$ are said to be *linearly disjoint* over $k$. We shall be concerned with conditions for linear disjointness of subalgebras, and in particular, of subextension fields of $\Omega/k$.

It is clear that if $A, B$ are linearly disjoint and $A' \subseteq A$, $B' \subseteq B$, then $A', B'$ are again linearly disjoint. In the opposite direction, $A, B$ are linearly disjoint whenever all finitely generated subalgebras of $A$ and $B$ are linearly disjoint; this follows because the condition involves only finite sets of elements. We also note that any subalgebra of $\Omega$ is an integral domain; if it is algebraic, it must be a field, for then any element $\alpha$ satisfies an equation

$$\alpha^n + c_1 \alpha^{n-1} + \cdots + c_n = 0,$$

where $c_n \neq 0$ if $\alpha \neq 0$, and then $\alpha^{-1} = -c_n^{-1}(\alpha^{n-1} + c_1 \alpha^{n-2} + \cdots + c_{n-1})$.

From the structure of tensor products we obtain the following equivalent conditions for linear disjointness:

PROPOSITION 5.1 *For any $k$-algebras $A$ and $B$ in $\Omega$ the following conditions are equivalent*:

(a) *$A$, $B$ are linearly disjoint over $k$,*

(b) *any $k$-basis of $A$ remains linearly independent over $B$,*

(c) *any $k$-basis of $B$ remains linearly independent over $A$,*

(d) *if $(u_i)$ is a $k$-basis for $A$ and $(v_\lambda)$ a $k$-basis for $B$, then the products $u_i v_\lambda$ are linearly independent over $k$,*

(e) *if $K$, $L$ are the subfields of $\Omega$ generated by $A, B$ respectively, then $K, L$ are linearly disjoint over $k$.*

*Proof.* The equivalence of (a)–(d) is an immediate consequence of the independence property of the tensor product (A.2, **4.7**). (e)$\Rightarrow$(a) is clear, and to prove (a)$\Rightarrow$(e), let $a_i \in K$, $b_i \in B$ be such that $\sum a_i b_i = 0$. Write $a_i = a_i'/s$, where $a_i', s \in A$; then $\sum a_i' b_i = (\sum a_i b_i)s = 0$, so if $A, B$ are linearly disjoint, it follows that $\sum a_i' \otimes b_i = 0$, hence $\sum a_i \otimes b_i = 0$. Thus if $A, B$ are linearly disjoint, then so are $K, B$, and by another application, so are $K$ and $L$. ■

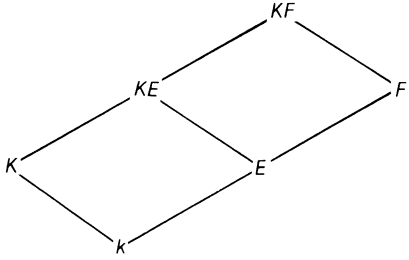
As an example, suppose that $K/k$ is $p$-radical and $L/k$ separable; then $KL = \{\sum a_i b_i | a_i \in K, b_i \in L\}$ is separable over $K$, and as we shall see in Prop. 5.3, $K$ and $L$ are linearly disjoint over $k$ (cf. also Ex. (6) of the further exercises).

We record the following transitivity property:

PROPOSITION 5.2 *Let $k \subseteq K$, $k \subseteq E \subseteq F$ be extension fields of $k$. Then $K$ and $F$ are linearly disjoint over $k$ if and only if $K$ and $E$ are linearly disjoint over $k$ and $KE$ and $F$ are linearly disjoint over $E$.*

*Proof.* We have the evident relation

$$K \otimes_k F \cong (K \otimes_k E) \otimes_E F.$$




Thus the mapping

$$K \otimes_k F \to \Omega \tag{2}$$

can be carried out in two stages, first mapping $K \otimes_k E$ to $\Omega$, and then $KE \otimes_E F$ to $\Omega$, and here it is immaterial, by Prop. 5.1, whether we take $KE$ to be the subfield or the subalgebra generated by $K$ and $E$. Thus the mapping (2) is injective iff the latter two mappings are, and this gives the assertion. ■

PROPOSITION 5.3 *Given subfields $K, L$ of $\Omega$, if $K/k$ is $p$-radical and $L/k$ separable, then $K, L$ are linearly disjoint over $k$ and $L$ is the separable closure of $k$ in $KL$.*

*Proof.* Put $E = KL$; since $L/k$ is separable, so is $E/K$, and the result will follow from Th. 3.7 if we can show that $L$ is the separable closure of $k$ in $E$. If this separable closure is denoted by $k_s$, then clearly $L \subseteq k_s$ and we have to prove equality. Thus, given $a \in k_s$, we must show that $a \in L$, and for this we may assume $K, L$ to be finitely generated, and hence finite, over $k$. Any $k$-basis of $L$ spans $E$ over $K$, and its elements are separable over $K$, hence $[E:K]_s \leqslant [L:k]_s$. Now $[E:k]_s = [E:K]_s[K:k]_s = [E:K]_s \leqslant [L:k]_s \leqslant [E:L]_s[L:k]_s = [E:k]_s$, hence equality holds throughout, and $[E:L]_s = 1$, so $L$ has no separable extension in $E$. It follows that any $a \in k_s$ lies in $L$, as claimed. ■

Let $K, L$ be two algebraic subextensions of $\Omega/k$; to find whether $K, L$ are linearly disjoint, we consider the natural mapping

$$K \otimes_k L \to \Omega. \tag{3}$$

If the tensor product on the left is a field, then the kernel must be trivial and so $K, L$ are then linearly disjoint. Conversely, when $K, L$ are linearly disjoint, then $K \otimes_k L$ can be embedded in $\Omega$ by (3). As a subalgebra of $\Omega$ generated by algebraic elements it is algebraic and hence a field. Thus we have

PROPOSITION 5.4 *Two algebraic subextensions $K, L$ of $\Omega/k$ are linearly disjoint if and only if $K \otimes_k L$ is a field.* ■

It is clear that for linearly disjoint extensions $K, L$ of $k$ we must have $K \cap L = k$. In an important case this necessary condition is also sufficient.

THEOREM 5.5 *Let $K, L$ be two subextensions of $\Omega/k$ of which one is normal and one (possibly the same) separable. Then $K, L$ are linearly disjoint if and only if $K \cap L = k$.*

In order to prove this result we restate it in absolute form:

THEOREM 5.5′ *Let $K/k$, $L/k$ be two extensions of which one is normal and one is separable. Then $K \otimes_k L$ is a field if and only if $K, L$ have no isomorphic subfield properly containing $k$.*

*Proof.* Suppose that $K, L$ each have a subfield isomorphic to a proper extension of $k$. We may as well take this to be simple, say $k(\alpha)$, where $\alpha$ is algebraic of degree $n > 1$. Let its minimal equation over $k$ be

$$c_0 x^n + c_1 x^{n-1} + \cdots + c_n = 0 \qquad (c_0 = 1).$$

Then $\sum c_i \alpha^{n-i} = 0$; consider the product

$$\left( \sum_{i=0}^{n-1} \sum_{j=1}^{n-i} c_i \alpha^{n-i-j} \otimes \alpha^{j-1} \right) (\alpha \otimes 1 - 1 \otimes \alpha). \tag{4}$$

The coefficient of $c_i$ is

$$\sum (\alpha^{n-i-j+1} \otimes \alpha^{j-1} - \alpha^{n-i-j} \otimes \alpha^j) = \alpha^{n-i} \otimes 1 - 1 \otimes \alpha^{n-i}.$$

Hence (4) becomes

$$\sum_i c_i \alpha^{n-i} \otimes 1 - 1 \otimes \sum_i c_i \alpha^{n-i} = 0.$$

This shows that $K \otimes L$ has zero-divisors, so it cannot be a field.

To prove the converse it will be enough to show that under the given

conditions $K \otimes L$ has no zero-divisors, and for this we may take $K, L$ to be finitely generated and hence finite over $k$. Suppose first that $K/k$ is both separable and normal, hence Galois, and so is simple, say $K = k(\alpha)$, where $\alpha$ has minimal polynomial $f$ over $k$. Over $L$ we have a factorization

$$f = p_1 \cdots p_r, \qquad p_i \text{ irreducible over } L, \tag{5}$$

and $K \otimes L = L(\alpha)$ is a field iff $r = 1$. Assume that $K \otimes L$ is not a field; then $r > 1$ and if $L_1$ is the subfield of $L$ generated by the coefficients of the $p_i$, then $L_1 \neq k$. Let us enlarge $L$ to a splitting field of $f$; this will contain a splitting field $L_0$ of $f$ over $k$ and clearly $L_0 \supseteq L_1$. But $K$ as Galois extension of $k$ generated by $\alpha$ is also a splitting field of $f$ over $k$. By uniqueness, $K \cong L_0$; now $L_1$ is a subfield of $L$ and is contained in $L_0$, hence it is isomorphic to a subfield of $K$. This contradicts the hypothesis and it shows the condition to be necessary.

There remains the case when one of $K, L$, say $K$ is normal and $L$ is separable. Taking $K$ again finite, we have by Th. 3.10,

$$K = k_s \otimes_k k_i, \tag{6}$$

where $k_s$ is the separable closure and $k_i$ the perfect closure of $k$ in $K$. Here $k_s$ is Galois and we have

$$K \otimes_k L = k_s \otimes k_i \otimes L.$$

Under the given hypothesis $k_s$ and $L$ have no isomorphic subfield $k$, hence $k_s \otimes L$ is a field, by what has been proved. It is generated by separable elements over $k$ and hence is a separable extension. Therefore $k_s \otimes L \otimes k_i$ is a field, by Prop. 5.3, and thus $K \otimes L$ is a field, as we had to show. ■

It is now easy to deduce Th. 5.5. If $K, L$ are subfields of $\Omega$ having isomorphic subextensions $K_1, L_1$ and $K$ is normal, then $L_1$ is isomorphic to $K_1$, hence conjugate to $K_1$ in $\Omega$, by Prop. 3.8, and so $L_1 \subseteq K$ by the normality of $K$. It follows that $K \cap L \neq k$. Conversely, when $K \cap L \neq k$, then it is clear that $K, L$ have isomorphic subextensions $\supset k$; now we can apply Th. 5.5′ to complete the proof. ■

Here is an example to show that normality cannot be omitted. Let $E/k$ be an extension of degree $n$ without proper subextensions (e.g. $E$ corresponds to a maximal subgroup of index $n$ within a Galois extension). Let $E'$ be conjugate to $E$ within $\Omega$ but distinct from it. Then $E \cap E' = k$ but $[EE' : k] \leqslant n(n-1)$, hence $E$ and $E'$ are not linearly disjoint.

Linear disjointness may be used to study field extensions in prime characteristic $p$, and here it is useful to define separability in a wider setting. We recall that a commutative ring $A$ is said to be *reduced* if it has no nilpotent element $\neq 0$, i.e. if, given $x \in A, x^n = 0$ for some $n > 0$ implies $x = 0$. Now a commutative $k$-algebra $A$ is called *separable* if $A \otimes_k E$ is reduced for all field

extensions $E$ of $k$. For example, any purely transcendental field extension is separable, since $k(x) \otimes E = E(x)$ for an indeterminate $x$, and similarly for several indeterminates. It is clear that a $k$-algebra is separable iff all its finitely generated subalgebras are separable. Our first concern is to show that for fields the terminology agrees with that introduced earlier.

PROPOSITION 5.6 *An algebraic field extension $E/k$ is separable if and only if $E$ is separable as $k$-algebra.*

*Proof.* By the remark just made we can take $E/k$ to be finitely generated and hence finite. If $E/k$ is separable as field extension, it is simple (Th. 3.9.2 of A.2) and so has the form $E = k(\alpha)$, where $\alpha$ has a minimal polynomial $f$ with distinct zeros in any field. Let $F$ be an extension field of $k$ and suppose that $f = p_1 \cdots p_r$ is the complete factorization of $f$ over $F$. Then $K \otimes F \cong A_1 \times \cdots \times A_r$, where $A_i \cong F[x]/(p_i)$ is a field; hence $E \otimes F$ is reduced and so $E$ is separable as $k$-algebra. On the other hand, if $E/k$ is not separable as field extension, then $k_s$, the separable closure of $k$ in $E$ is distinct from $E$. For any $\alpha \in E \setminus k_s$ the minimal polynomial $f$ of $\alpha$ over $k_s$ has the form $f = (x - \alpha)^q$, where $q = p^s > 1$. Moreover, we have

$$E \otimes_k E = E \otimes_k k_s \otimes_{k_s} E,$$

and $\alpha^q \in k_s$, hence $(\alpha \otimes 1 - 1 \otimes \alpha)^q = \alpha^q \otimes 1 - 1 \otimes \alpha^q = 0$ in $(E \otimes k_s) \otimes E$, while $\alpha \otimes 1 - 1 \otimes \alpha \neq 0$. So in this case $E$ is not separable as $k$-algebra. ■

We recall that in a commutative ring the set $\mathfrak{N}$ of all nilpotent elements is an ideal, the nilradical, which may also be expressed as the intersection of all prime ideals (A.2, Prop. 9.2.5). As a result we have

THEOREM 5.7 *A commutative ring $A$ is reduced if and only if it is a subring of a direct product of fields. If $A$ is a finite-dimensional $k$-algebra and is reduced, it is a direct product of fields.*

*Proof.* Let $A$ be reduced; by the result quoted, $\bigcap \mathfrak{p}_\lambda = 0$, where $\mathfrak{p}_\lambda$ ranges over all prime ideals. The quotient algebra $A/\mathfrak{p}_\lambda$ is an integral domain and so has a field of fractions $E_\lambda$. The natural mappings $f_\lambda : A \to E_\lambda$ can be combined to a homomorphism $f : A \to \prod E_\lambda$ whose kernel is $\bigcap \mathfrak{p}_\lambda = 0$, so $A$ has been embedded in a direct product of fields. The converse is clear.

If $A$ is a finite-dimensional $k$-algebra which is reduced, then it is Artinian with zero radical and so can be expressed as a direct product of fields, by Wedderburn's theorem (A.2, Th. 5.3.4). ■

COROLLARY 5.8 *If $A, B$ are separable $k$-algebras, then so is $A \otimes_k B$.*

*Proof.* We first show that if $A$ is separable and $B$ is reduced, then $A \otimes B$ is

again reduced. By Th. 5.7 we have an embedding $f: B \to \prod E_\lambda$, where the $E_\lambda$ are fields; let $f_\lambda: B \to E_\lambda$ be the projection on $E_\lambda$. Then $1 \otimes f_\lambda: A \otimes B \to A \otimes E_\lambda$ is a homomorphism into a reduced $k$-algebra $A \otimes E_\lambda$. Combining these mappings we obtain a mapping $\varphi: A \otimes B \to \prod(A \otimes E_\lambda)$, which is an embedding in a reduced $k$-algebra; its kernel is $\bigcap(A \otimes \ker f_\lambda) = 0$, and so $A \otimes B$ is reduced.

Now if $A$ and $B$ are separable, then $B \otimes E$ is reduced for any field $E$, hence $A \otimes B \otimes E$ is reduced, by what we have shown, and this means that $A \otimes B$ is separable, as claimed. ■

In characteristic 0 every field extension is separable, as algebra; this is clear by writing it as an algebraic extension of a purely transcendental extension. Now let $A$ be a reduced $k$-algebra, where $k$ is of characteristic 0. Then $A \otimes E$ is reduced for any field $E$ over $k$, by Cor. 5.8, hence we obtain

COROLLARY 5.9 *Let $k$ be a field of characteristic 0. Then every reduced $k$-algebra is separable; in particular every field extension is separable.* ■

In testing for separability it is enough to take $p$-radical field extensions. This is made precise in

THEOREM 5.10 *Let $A$ be any commutative $k$-algebra, where $k$ is a field and denote by $\hat{k}$ the perfect closure of $k$ and put $k^{p^{-1}} = \{x \in \hat{k} \mid x^p \in k\}$. Then the following conditions are equivalent:*

(a) *$A$ is separable,*

(b) *$A \otimes \hat{k}$ is reduced,*

(c) *$A \otimes k^{p^{-1}}$ is reduced,*

(d) *if a family $(a_i)$ of elements of $A$ is linearly independent over $k$, then so is the family $(a_i^p)$.*

*Proof.* Since $\hat{k} \supseteq k^{p^{-1}}$, it follows that (a) $\Rightarrow$ (b) $\Rightarrow$ (c). To prove that (c) $\Rightarrow$ (d), take any linearly independent family $a_i \in A$ and assume that $\sum \lambda_i a_i^p = 0$. In $k^{p^{-1}}$ there exists $\mu_i$ such that $\mu_i^p = \lambda_i$ and we have $(\sum \mu_i \otimes a_i)^p = \sum \lambda_i a_i^p = 0$, hence $\sum \mu_i \otimes a_i = 0$ by (c), and so $\mu_i = 0$. It follows that $\lambda_i = \mu_i^p = 0$, i.e. (d).

(d) $\Rightarrow$ (a). Let $(u_i)$ be a $k$-basis of $A$ and suppose that $E/k$ is an extension. To prove that $A$ is separable we must show that $A \otimes E$ is reduced, and it will be enough to show that $x^p = 0$ implies $x = 0$, for any $x \in A \otimes E$. We can write $x = \sum \lambda_i \otimes u_i$ $(\lambda_i \in E)$; hence $0 = x^p = \sum \lambda_i^p \otimes u_i^p$; by (d) the $u_i^p$ are linearly independent, so $\lambda_i^p = 0$, hence $\lambda_i = 0$ and so $x = 0$, and this shows $A$ to be separable. ■

We can also write down an intrinsic condition for separability. Let $E = k(x_1, \ldots, x_n)$ be a finitely generated field extension; we shall say that $E$ is *separably generated* over $k$ if there is a transcendence basis $u_1, \ldots, u_r$ such that

$E$ is separably algebraic over $k(u_1,\dots,u_r)$. The $u$'s are then called a *separating transcendence basis* for $E/k$. To find conditions for such a basis to exist we can of course restrict ourselves to prime characteristic $p$.

THEOREM 5.11 (Mac Lane's criterion) *Let $k \subseteq E \subseteq \Omega$, where $\Omega$ is algebraically closed, and write $\hat{k}$ for the perfect closure of $k$ in $\Omega$. Then the following conditions are equivalent:*

(a) *$E/k$ is separable,*

(b) *$E$ is linearly disjoint from $\hat{k}$ over $k$,*

(c) *$E$ is linearly disjoint from $k^{p^{-1}}$ over $k$ ($p = \operatorname{char} k$),*

(d) *every subfield of $E$ containing $k$ and finitely generated over $k$ is separably generated.*

*Proof.* (a)⇒(b)⇒(c) is again clear. To prove (c)⇒(d), let $F = k(x_1,\dots,x_n) \subseteq E$; then $F$ is linearly disjoint from $k^{p^{-1}}$, because $E$ is. We use induction on $n$, the case $n = 0$ being trivial; likewise, if the $x_i$ are algebraically independent over $k$, there is nothing to prove. Otherwise choose a polynomial of least degree

$$f(x_1,\dots,x_n) = 0. \tag{7}$$

If $f$ involves $x_1$ and not merely as $p$th power, i.e. $\partial f/\partial x_1 \neq 0$, this is a separable equation for $x_1$ over $k(x_2,\dots,x_n)$, thus $F$ is separable over $k(x_2,\dots,x_n)$, the latter is separably generated over $k$ by induction on $n$, and now the result follows by the transitivity of separability. A similar argument holds if $f$ involves some other $x_i$ to a power not divisible by $p$. This only leaves the case where $f$ is a polynomial in $x_1^p,\dots,x_n^p$. Then we can write $f$ as $f = \sum a_\lambda M_\lambda^p$, where each $M_\lambda$ is a monomial in the $x_i$ and $a_\lambda \in k$. By the minimality of $\deg f$ the monomials $M_\lambda$ are linearly independent over $k$, so they are also linearly independent over $k^{p^{-1}}$, by (c), but we have a relation

$$\sum b_\lambda \otimes M_\lambda = 0,$$

where $b_\lambda^p = a_\lambda$. So this must be trivial, i.e. $b_\lambda = 0$, so $a_\lambda = 0$ for all $\lambda$, which is a contradiction. Hence this case cannot occur.

(d)⇒(a). To prove (a) we can limit ourselves to finitely generated subfields, so let $E$ be separably generated over $k$, say $E$ is separably algebraic over $k(u_1,\dots,u_r) = k(u)$, where the $u$'s are algebraically independent over $k$. We claim that $k(u)$ and $k^{p^{-1}}$ are linearly disjoint over $k$. For the distinct monomials $M_\lambda$ in the $u$'s form a basis for $k[u_1,\dots,u_r]$ and they are linearly independent over $k^{p^{-1}}$, because if

$$\sum c_\lambda M_\lambda = 0, \qquad c_\lambda^p \in k, \tag{8}$$

then $\sum c_\lambda^p M_\lambda^p = 0$. Here the $M_\lambda^p$ are distinct and hence linearly independent over $k$, so (8) must be trivial. Next $E$ and $k^{p^{-1}}(u)$ are linearly disjoint over $k(u)$, for

the first is separable and the second $p$-radical over $k(u)$ (Prop. 5.3). Now we can apply Prop. 5.2 with $k^{p^{-1}}$, $k(u)$, $E$ for $K$, $E$, $F$ respectively, and conclude that $E$ and $k^{p^{-1}}$ are linearly disjoint over $k$; thus $E \otimes k^{p^{-1}}$ is a field and hence by Th. 5.10, $E$ is separable, so (a) holds. ■

The following immediate consequences are often useful.

COROLLARY 5.12 *Any finitely generated subextension of a separably generated extension is again separably generated.* ■

COROLLARY 5.13 (F. K. Schmidt) *If $k$ is perfect, then any finitely generated extension is separably generated.* ■

In conclusion we examine briefly a special class of separable extensions which is of importance in algebraic geometry. A field extension $E/k$ is said to be *regular* if $E \otimes F$ is an integral domain for any field extension $F/k$. It is again easy to check that $E/k$ is regular iff this is so for every finitely generated subextension. Our object is to give some alternative characterizations, but to do so we shall need a lemma. We recall Gauss's lemma (Lemma 3.7.1 of A.2), which states that for any integral domain $A$, a prime element of $A$ remains prime in $A[x]$. Hence for any UFD $A$ with field of fractions $K$, a polynomial $f \in A[x]$ is irreducible over $A$ iff it is irreducible over $K$ and primitive in $A[x]$; this follows because every factorization over $K[x]$ can be pulled down to $A[x]$, i.e. $A[x]$ is inert in $K[x]$ (Th. 3.7.2 of A.2).

LEMMA 5.14 *Let $E/k$ be an extension, where $k$ is relatively algebraically closed in $E$. If $y_1, \ldots, y_n$ are indeterminates over $E$ and $f$ is a polynomial in $x$ with coefficients in $k(y_1, \ldots, y_n)$ which is irreducible over this field, then it is irreducible over $E(y_1, \ldots, y_n)$.*

*Proof* (H. Flanders) We have to show that $f \in k(y_1, \ldots, y_n)[x]$ is irreducible over $E(y_1, \ldots, y_n)$. By the remark just made, we may multiply $f$ by a unit of $k(y_1, \ldots, y_n)$ so that $f$ becomes an irreducible element of $k[y_1, \ldots y_n, x]$. We shall show that $f$ is still irreducible in $E[y_1, \ldots, y_n, x]$. For convenience we denote $x$ by $y_0$ and write $Y = \{y_0, \ldots, y_n\}$. Suppose that $f = gh$ in $E[Y]$; choose an integer $N$ greater than the degree of $f$ and consider the homomorphism $\varphi$: $E[Y] \to E[t]$ defined by $y_i \mapsto t^{N^i}$. We have $f^\varphi = g^\varphi h^\varphi$ and on multiplying $g$ by a unit in $E$ and $h$ by its inverse, we may assume that the leading coefficients of $g^\varphi$ and $h^\varphi$ lie in $k$. By taking a complete factorization of $f^\varphi$ over an algebraic closure of $E$ we see that the coefficients of $g^\varphi$ and $h^\varphi$ are algebraic over $k$ and lie in $E$, hence $g^\varphi, h^\varphi \in k[t]$, by the hypothesis on $k$. But $\varphi$ is injective on polynomials of degree less than $N$, because $y_0^{\alpha_0} \cdots y_n^{\alpha_n} \mapsto t^A$, where

$A = \alpha_0 + \alpha_1 N + \cdots + \alpha_n N^n$. It follows that $g$ and $h$ have coefficients in $k$, so one of $g, h$ must be a unit, and we have proved that $f$ is irreducible in $E[Y]$. By Gauss's lemma it follows that $f$ is irreducible in $E(y_1, \ldots, y_n)[x]$. ■

For any field $E$ let us denote its algebraic closure by $E^a$. If $k \subseteq E$, we may clearly take $k^a$ to be a subfield of $E^a$.

THEOREM 5.15 *For any extension $E/k$ the following conditions are equivalent:*

(a) *$E/k$ is regular,*

(b) *$E$ and $k^a$ are linearly disjoint in $E^a$,*

(c) *$k$ is relatively algebraically closed in $E$ and any finitely generated subextension of $E/k$ is separably generated.*

*Proof.* (a)⇒(b). Let $E/k$ be regular; then $E \otimes_k k^a$ is an integral domain. If $F$ is its field of fractions and $F^a$, an algebraic closure, then it is clear that $E$ and $k^a$ are linearly disjoint in $F^a$, and hence in an algebraic closure of $E$, so (b) holds.

(b)⇒(c). When $E$ and $k^a$ are linearly disjoint, it is clear that $k$ is relatively algebraically closed in $E$, and since $k^a$ contains a perfect closure $\hat{k}$ of $k$, Mac Lane's criterion is satisfied, hence any finitely generated subextension of $E/k$ is separably generated.

(c)⇒(a). To prove that $E/k$ is regular we may assume that $E/k$ is finitely generated; thus we may assume $E$ to be separably generated over $k$. Likewise $F/k$ may be assumed finitely generated, with transcendence basis $Y$ say. If we put $K = k(Y)$, then $F/K$ is a finitely generated algebraic extension, hence finite. Denote by $K_s$ the separable closure of $K$ in $F$; then $K_s = K[\alpha]$ for some $\alpha \in K_s$, and so $K_s = K[x]/(f)$, where $f$ is the minimal polynomial for $\alpha$ over $K$. By Lemma 5.14, $f$ is still irreducible over $E(Y)$, so $K_s \otimes_{k(Y)} E(Y) \cong E(Y)[x]/(f)$ and this is a field, which may be denoted by $E(Y)[\alpha]$. Since $E/k$ is separably generated, it follows that $E(Y)[\alpha]/k(Y)[\alpha]$ is again separably generated. Now $F \subseteq \hat{K}_s$, therefore by Mac Lane's criterion $F \otimes_{K_s} E(Y)[\alpha]$ is a domain. But $F \otimes_k E = F \otimes_{K_s}(K_s \otimes_K (K \otimes_k E)) \subseteq F \otimes_{K_s}(K_s \otimes_K E(Y)) = F \otimes_{K_s} E(Y)[\alpha]$, and it follows that $F \otimes_k E$ is a domain. Therefore $E/k$ is regular, as claimed. ■

The reader is advised to draw a diagram.

For an algebraically closed ground field the conclusion can be simplified:

COROLLARY 5.16 *If $k$ is algebraically closed, then any extension of $k$ is regular.* ■

Comparing Th. 5.11 and 5.15 we see that whereas separability of $E/k$ means that $E$ is linearly disjoint from the $p$-radical elements over $k$, regularity means that $E$ is linearly disjoint from all algebraic elements over $k$. In particular, the notion of regularity is not vacuous even in characteristic zero.

## Exercises

(1) Show that in a finitely generated extension which is separably generated, every generating set contains a separating transcendence basis.

(2) Let $K = k(x, y)$, where $x, y$ are transcendental over $k$ and satisfy $x^2 = y^3 - 1$. Show that if $\operatorname{char} k = 2$, then $K/k(y)$ is not separable, but that $K/k(x)$ is, and hence $K/k$ is separably generated.

If $\operatorname{char} k = 2$ and $\alpha, \beta$ are non-squares in $k$, then $k(x, y)$, where $\alpha x^2 + \beta y^2 = 1$, is not separably generated over $k$.

(3) Let $K/k$ be a function field of one variable, i.e. a finitely generated extension of transcendence degree 1. Show that if $k$ is perfect, then for any $x \in K$ transcendental over $k$ there exists $y \in K$ such that $K = k(x, y)$. If $k$ is imperfect, say $\alpha \notin k^p$, and $K = k(x, y)$, where $x^p + y^p = \alpha$, then there is no pair including $x^p$ which generates $K$.

(4) Show that a root $\alpha$ of an irreducible equation with insoluble group over $\mathbf{Q}$ is such that $\alpha^n$ is irrational for all $n \neq 0$. With such $\alpha$ and a transcendental element $u$ of $\mathbf{C}$ put $E = \mathbf{Q}(u)$, $F = \mathbf{Q}(\alpha u)$. Verify that $E \cap F = \mathbf{Q}$ but that $E, F$ are not linearly disjoint over $\mathbf{Q}$. (Hint: to show that $E \cap F = \mathbf{Q}$ use formal Laurent series in $u$.)

(5) Two extensions $E, F$ are said to be *free* over $k$ if any set of elements of $E$ which is algebraically independent over $k$ is also algebraically independent over $F$. Show that if $E, F$ are free over $k$, then so are $F, E$.

(6) Show that if $E, F$ are linearly disjoint over $k$, then they are free over $k$. Does the converse hold? (Hint: any algebraic extensions of $k$ are free over $k$.)

(7) Show that if $E, F$ are free over $k$ and $E$ is separably generated over $k$, then $EF$ is separably generated over $F$.

(8) Show that if $E/F$ and $F/k$ are regular, then so is $E/k$.

(9) Let $A$ be an integral domain with field of fractions $K$ and suppose that $A$ has an irreducible element $c$ such that $A/(c)$ is not reduced. Show that $A[x]$ has irreducible polynomials which become reducible over $K$.

(10) Let $F = \mathbf{F}_p$, $E = F(c_0, \ldots, c_{p-1}, a)$, where the $c_i$ and $a$ are indeterminates, and define $k = F(c_0^p, \ldots, c_{p-1}^p, a)$, $K = k(\alpha, \beta)$, where $\alpha$ is a root of $x^p = a$ and $\beta$ is a root of $x^p = \sum_0^{p-1} c_i^p a^i$. Show that $K/k$, $E/k$ are both normal but that $K \otimes_k E$ has zero-divisors.

(11) Show that an extension $E/k$ is separable iff $E \otimes_k E$ is reduced.

(12) An extension $E/k$ is called *self-regular* if $E \otimes_k E$ is an integral domain. Show that if $E/k$ is self-regular, then $k$ is relatively algebraically closed in $E$. (Hint: look at the proof of Th. 5.5′.)

(13) Find an example of a self-regular extension which is not regular.

(14) Show that if $E, F$ are free over $k$, then $E \cap F$ is algebraic over $k$.

(15) Let $L/k$ be a field extension, $E, F$ subextensions and $X$ a transcendence basis of $E/k$. Show that $E, F$ are free over $k$ iff $k(X)$ and $F$ are linearly disjoint over $k$.

## 5.6 Infinite algebraic extensions

We recall that a field $K$ is said to be *algebraically closed* if every polynomial equation $f(x) = 0$ over $K$ has a root in $K$. By the remainder theorem this amounts to requiring every polynomial (in one variable) of positive degree over $K$ to have a linear factor, and an induction on the degree shows that then every polynomial in $x$ over $K$ can be written as a product of linear factors. Given a field $k$, an *algebraic closure* $k^a$ is a field containing $k$ which is algebraically closed and algebraic over $k$. In **3.3** of A.2 we saw that every field has an algebraic closure. Below we shall give another existence proof. We begin by establishing the uniqueness.

PROPOSITION 6.1 *Let $k$ be any field. Then any two algebraic closures of $k$ are isomorphic over $k$.*

*Proof.* Let $K, K'$ be algebraic closures of $k$ and consider the algebra $A = K \otimes_k K'$. This is a $k$-algebra containing $K$, via the embedding $a \mapsto a \otimes 1$, and likewise $K'$, and $A$ is generated by them; hence every element is algebraic over $k$. Let $\mathfrak{m}$ be a maximal ideal of $A$ and consider the field $L = A/\mathfrak{m}$. The natural homomorphism $K \to A \to L$ is an embedding, because its kernel is a proper ideal of $K$. Thus $L$ again contains $K$ and $K'$ and is generated by them and is algebraic over $k$. Let $c \in L$ and consider its minimal polynomial $f$ over $k$; by hypothesis $f$ splits into linear factors over $K$: $f(x) = \prod(x - \alpha_i)$, hence $\prod(c - \alpha_i) = 0$ and since $L$ is a field, we have $c = \alpha_i$ for some $i$. Thus $c \in K$ and this shows that $K = L$. By symmetry we also have $K' = L$, so the identity mapping of $L$ is a $k$-isomorphism between $K$ and $K'$. ■

Next we note a property which will facilitate the existence proof of the algebraic closure.

PROPOSITION 6.2 *Let $E/k$ be an algebraic extension. If every polynomial equation over $k$ has a root in $E$, then $E$ is an algebraic closure of $k$.*

*Proof.* We have to show that every polynomial over $k$ splits into linear factors over $E$, and here we may clearly limit ourselves to irreducible polynomials. Let $f$ be an irreducible polynomial over $k$ and denote its splitting field by $F$. Since $F/k$ is normal, we can write $F = L \otimes_k M$, where $L$ is Galois and $M$ $p$-radical over $k$, by Th. 3.10. By the theorem of the primitive element (Th. 3.9.2 of A.2),

$L/k$ is generated by a single element $\alpha$. The minimal polynomial $g$ of $\alpha$ over $k$ is irreducible and so has a zero $\beta$ in $E$; moreover $k(\beta) \cong k(\alpha) = L$, so $E$ contains a subfield isomorphic to $L$. Any element $\gamma$ of $M$ is $p$-radical over $k$ and so is a root of an equation $x^q = c$ over $k$ ($q = p^r$); it follows that there is a unique element $\gamma'$ of $E$ with the same minimal equation as $\gamma$, and the correspondence $\gamma \mapsto \gamma'$ is clearly a homomorphism $M \to E$. Thus $E$ contains a subfield isomorphic to $M$; by Prop. 5.3, $E$ contains a subalgebra isomorphic to $L \otimes M \cong F$, and it follows that $f$ splits over $E$. Now any element algebraic over $E$ is algebraic over $k$, and so by what has been proved, belongs to $E$, hence $E$ is algebraically closed. It is also algebraic over $k$ and hence is an algebraic closure of $k$. ■

THEOREM 6.3 (Steinitz) *Let $k$ be any field. Then $k$ has an algebraic closure and any two algebraic closures of $k$ are isomorphic over $k$.*

*Proof* (E. Artin). Let $F$ be the set of all irreducible polynomials over $k$ of degree at least two, and let $X = \{x_f\}(f \in F)$ be a family of indeterminates indexed by $F$. We form the polynomial ring $R = k[X]$ in these indeterminates and consider the ideal $\mathfrak{a}$ of $R$ generated by the elements $f(x_f)$, $f \in F$. This ideal is proper; for if not, we would have an equation

$$1 = u_1 f_1(x_{f_1}) + \cdots + u_n f_n(x_{f_n}), \qquad \text{where } u_i \in R. \tag{1}$$

If we adjoin to $k$ zeros $\alpha_1, \ldots, \alpha_n$ of $f_1, \ldots, f_n$ respectively, we obtain an algebraic extension $k'$ of $k$, and replacing $x_{f_i}$ in (1) by $\alpha_i$ and $x_f$ by 0 for $f \neq f_i$, we obtain the equation $1 = 0$ in $k'$, a contradiction. Hence $\mathfrak{a}$ is proper; it is therefore contained in a maximal ideal $\mathfrak{m}$ of $R$ (Th. 2.2.11 of A.2). The residue class ring $K = R/\mathfrak{m}$ is a field because $\mathfrak{m}$ was maximal, and the homomorphism $k \to k[X] \to K$ obtained by composing the inclusion with the natural mapping is an embedding of fields. Now in $K$ every irreducible polynomial over $k$ has a zero, the zero of $f$ being the image of $x_f$ in $K$. Hence every polynomial over $k$ has a zero in $K$; moreover, $K$ is generated by the images of the $x_f$, algebraic over $k$, and so $K$ is itself algebraic over $k$. By Prop. 6.2, $K$ is an algebraic closure of $k$, and by Prop. 6.1 it is unique up to $k$-isomorphism. ■

An algebraic closure of $k$ will usually be denoted by $k^a$. Although any two algebraic closures of $k$ are isomorphic, the isomorphism between them is not generally unique, for $k^a$ usually has many $k$-automorphisms. To study these automorphisms let us define an *algebraic Galois extension* as an extension $K/k$, where $k$ is the fixed field of a group $G$ of automorphisms of $K$, acting with finite orbits. It follows that $K/k$ is algebraic, for if $a \in K$ lies in the orbit $\{a_1, \ldots, a_r\}$, then $f = \prod(x - a_i)$ is a polynomial over $k$ with $a$ as zero. As in the finite case, an algebraic extension $K/k$ is said to be *separable* if every element of $K$ is separable over $k$.

Our first task is to extend the characterization of Galois extensions to the infinite case:

THEOREM 6.4 *For any algebraic extension $K/k$, the following conditions are equivalent:*

(a) *$K/k$ is Galois,*

(b) *$K/k$ is normal and separable,*

(c) *$K$ is the minimal splitting field of a family of separable polynomials over $k$.*

*Proof.* (a)$\Rightarrow$(b). Let $f$ be an irreducible monic polynomial with a zero $\alpha\in K$. If the orbit of $\alpha$ under the group of $K/k$ is $\{\alpha_1,\ldots,\alpha_r\}$, we form $g=\prod(x-\alpha_i)$; now $g$ is fixed by all $k$-automorphisms of $K$ and so has coefficients in $k$, hence $f|g$, but $g$ was constructed as a factor of $f$, therefore $f=g$. This shows that $f$ splits over $K$ and has distinct zeros, so $K/k$ is normal and separable.

(b)$\Rightarrow$(c). Let $K$ be generated as $k$-algebra by a family $\{\alpha_i\}$ of separable algebraic elements, and denote the minimal polynomial of $\alpha_i$ over $k$ by $f_i$. If $K/k$ is normal, $f_i$ splits over $K$, hence $K$ is the minimal splitting field of the family $\{f_i\}$ of separable polynomials.

(c)$\Rightarrow$(a). Let $G$ be the group of all $k$-automorphisms of $K$ and suppose that $K$ is the minimal splitting field of a family $\{f_i|i\in I\}$ of separable polynomials. Any element $a$ of $K$ clearly lies in the minimal splitting field $E$ of some finite subfamily $\{f_i|i\in I'\}$. Now $E/k$ is normal and by Th. 3.5.4 of A.2, also separable, hence Galois. So if $a\notin k$, then there exists $\sigma\in\mathrm{Gal}(E/k)$ such that $a^\sigma\neq a$. By Prop. 3.8, $\sigma$ can be extended to a $k$-automorphism $\sigma_1$ of $K$. It follows that the fixed field of $G$ is $k$; moreover, the conjugates of any element satisfy the same equation over $k$ and so are finite in number. Hence $G$ acts with finite orbits and so $K/k$ is Galois. ■

COROLLARY 6.5 *Any separable extension $E/k$ is contained in a Galois extension $K/k$.*

For let $E$ be generated by $X=\{x_i\}$, take the minimal polynomial $f_i$ of $x_i$ over $k$ and let $K$ be a minimal splitting field of the $f_i$ over $k$. Then $K$ is a field containing $E$ and $K/k$ is normal and separable, hence Galois, by Th. 6.4. ■

Let us now investigate how finite Galois theory carries over to the algebraic case. Let $K/k$ be an algebraic Galois extension with group $G$. For any intermediate field $L$ we define

$$L^*=\{\sigma\in G|x^\sigma=x \text{ for all } x\in L\}, \tag{2}$$

and for any subgroup $H$ of $G$ we put

$$H^*=\{x\in K|x^\sigma=x \text{ for all } \sigma\in H\}. \tag{3}$$

As in the finite case we have a Galois connexion

$$\begin{aligned} L_1\subseteq L_2&\Rightarrow L_1^*\supseteq L_2^*,\\ H_1\subseteq H_2&\Rightarrow H_1^*\supseteq H_2^*, \end{aligned} \tag{4}$$

$$L \subseteq L^{**}, \qquad H \subseteq H^{**}, \tag{5}$$

and hence

$$L^{***} = L^*, \qquad H^{***} = H^*. \tag{6}$$

Moreover, $K/L$ is Galois, by Th. 6.4 (b), so $L^* \cong \mathrm{Gal}(K/L)$ and thus $L^{**} = L$. However, in contrast to the finite case, $H^{**}$ need not be equal to $H$, as we shall soon see. Let us therefore denote $H^{**}$ by $\bar{H}$ and call it the *closure* of $H$ in $G$; if $\bar{H} = H$, $H$ is said to be *closed* in $G$. We can now state the fundamental theorem for algebraic Galois extensions.

THEOREM 6.6 *Let $K/k$ be an algebraic Galois extension with group $G$. Then the mappings $L \mapsto L^*, H \mapsto H^*$ define a Galois connexion between the set of all fields between $k$ and $K$ and the set of all closed subgroups of $G$. An extension $L/k$ is normal if and only if $L^*$ is a normal subgroup of $G$, and when this is so, then $\mathrm{Gal}(L/k) \cong G/L^*$. Every closed subgroup $H$ of finite index in $G$ corresponds to a finite extension $L$ such that $[L:k] = (G:H)$. Moreover, if $\mathscr{F}$ denotes the set of all closed subgroups of finite index in $G$, then for any subgroup $H$ of $G$, its closure is given by*

$$\bar{H} = \bigcap_{N \in \mathscr{F}} HN. \tag{7}$$

*Proof.* By the definition of closure we see that $H^{**} = H$ for any closed subgroup of $G$. If $L$ is an intermediate field, then $L^*$ is a closed subgroup by (6), and we have seen that $L^{**} = L$, so we have a Galois connexion.

For any $\sigma \in G$, $(L^\sigma)^* = \sigma^{-1} L^* \sigma$, so $L/k$ is normal (and hence Galois) iff $L^* \lhd G$. When this holds, there is a map $G \to \mathrm{Gal}(L/k)$ defined by $\sigma \mapsto \sigma | L$; clearly this is a homomorphism with kernel $L^*$. As in the proof of Th. 6.4 we see that this homomorphism is surjective, so $G/L^* \cong \mathrm{Gal}(L/k)$.

Suppose now that $L/k$ is a finite extension, where $L \subseteq K$. We can find a finite Galois extension $E/k$ such that $L \subseteq E \subseteq K$. By the previous result, $\mathrm{Gal}(E/k) \cong G/E^*$ and $\mathrm{Gal}(E/L) \cong L^*/E^*$, so by finite Galois theory, $[L:k] = (G/E^*:L^*/E^*) = (G:L^*)$. Conversely, if $H$ is a closed subgroup of finite index in $G$, then the cosets $H\sigma$ give rise to distinct $k$-homomorphisms $H^* \to K$, hence $(G:H) \leqslant [H^*:k]_s$. Similarly, distinct $k$-homomorphisms $H^* \to K$ give rise to distinct cosets of $H^{**} = H$ in $G$, therefore $[H:k]_s \leqslant (G:H^{**}) = (G:H)$.

Finally, let $H$ be an arbitrary subgroup of $G$. If $N \in \mathscr{F}$, then $N$ is closed of finite index in $G$, hence it is open and so $HN$ is also open and hence closed of finite index. Therefore $\bar{H} \subseteq HN$, and it follows that $\bar{H} \subseteq \bigcap HN$. To prove equality, suppose that $\sigma \notin \bar{H}$, i.e. for some $a \in H^*$, $a^\sigma \neq a$. Let $E/k$ be a finite Galois extension such that $k(a) \subseteq E \subseteq K$; then $E^* \in \mathscr{F}$, but $H$ and $E^*$ fix $a$, hence $\sigma \notin HE^* \supseteq \bigcap HN$, and this establishes equality in (7). ■

We remark that the 'closed' subgroups arise indeed as the closed subgroups of a certain topology on $G$ in which the group operations are continuous (i.e. we have a topological group). This topology, called the *profinite* or *Krull topology*, may be defined by taking as a base for the open sets the collection of all cosets $Ha$, where $H \in \mathscr{F}, a \in G$. Since any algebraic extension is the direct limit of its finite extensions, its Galois group is the inverse limit of a system of finite groups, i.e. a *profinite* group (cf. **3.2**). As an example let us show that the topology of a Galois group is always Hausdorff. This means that distinct points have disjoint neighbourhoods, and it will follow if we show that $\bigcap \mathscr{F} = 1$. For if $a \neq b$, then $\mathscr{F}$ contains $H$ such that $ab^{-1} \notin H$, hence $Ha \cap Hb = \varnothing$ and we have found disjoint neighbourhoods for $a, b$. Now $\bigcap \mathscr{F} = 1$ just expresses the fact that $1 = K^*$ is closed (Th. 6.6). We also remark that $G$ is compact, as inverse limit of finite groups. On the other hand, not every subgroup of finite index need be closed (cf. Ex. (8)).

Any profinite group of automorphisms of a field gives rise to a Galois group, under suitable conditions, by the following result:

PROPOSITION 6.7 *Let $K$ be a field with a profinite group $G$ of automorphisms and fixed field $k$. If the stabilizer of any $x \in K$, $S(x) = \{\sigma \in G | x^\sigma = x\}$ is an open subgroup of $G$ and $\bigcap S(x) = 1$, then $G = \mathrm{Gal}(K/k)$.*

*Proof.* For any finite set of elements $a_1, \ldots, a_n \in K$, the subgroup $H = S(a_1) \cap \cdots \cap S(a_n)$ is again open, hence so is the intersection $N$ of all its conjugates. The finite group $G/N$ acts faithfully on $E = k(a_1^G, \ldots, a_n^G)$ and has fixed field $k$. Hence $E/k$ is a finite extension with group $G/N$. Now $K$ is the union of all such fields, $E$, and the intersection of all such subgroups $N$ is 1; hence $\mathrm{Gal}(K/k) \cong \varprojlim \mathrm{Gal}(E/k) \cong \varprojlim G/N \cong G$. ■

As an illustration of algebraic Galois extensions let us take the field $\mathbf{F}_p$ and form its algebraic closure $\mathbf{F}_p^a$. We know from **3.8** of A.2 that any finite extension of $\mathbf{F}_p$ has the form $\mathbf{F}_q$, where $q = p^n$ for some $n \geqslant 1$, and this field may be described as the minimal splitting field of $x^q - x$. Since any algebraic extension is determined by its finite subextensions, every subextension of $\mathbf{F}_p^a$ is the minimal splitting field of a family $\{x^{p^n} - x | n \in I\}$ for some family $I$ of positive integers. From the results of **3.8** of A.2 it is clear that we do not change the field by assuming $I$ to be closed under taking factors and LCMs, and we then have a bijection between such sets of integers and algebraic extensions of $\mathbf{F}_p$. To describe this situation concisely we introduce the notion of a supernatural number. Let $p_1, p_2, \ldots$ be all the prime numbers in some order. By a *supernatural number* we understand a formal product

$$N = \prod p_i^{\alpha_i} \qquad (\alpha_i \in \mathbf{N}_0 \text{ or } \alpha_i = \infty). \tag{8}$$

When all the $\alpha_i$ are finite and almost all are 0, this reduces to an ordinary

integer. Given two supernatural numbers, $N$ given by (8) and $P = \prod p_i^{\beta_i}$, we write $N|P$ to mean: $\alpha_i \leqslant \beta_i$ for all $i$. It is easy to see that any supernatural number is completely determined by its finite divisors. Hence an algebraic extension $E$ of $\mathbf{F}_p$ corresponds to a unique supernatural number $N$ such that $E$ has a finite subfield of degree $n$ over $\mathbf{F}_p$ iff $n|N$, and in this way the algebraic extensions of $\mathbf{F}_p$ are described by supernatural numbers; conversely, to each supernatural number there corresponds such an extension.

Let $G = \mathrm{Gal}(\mathbf{F}_p^a/\mathbf{F}_p)$; to give an example of a non-closed subgroup of $G$, consider the subgroup $H$ generated by the Frobenius endomorphism $x \mapsto x^p$. Clearly $H^* = \mathbf{F}_p$, hence $H^{**} = G$; but for any infinite algebraic extension $E/\mathbf{F}_p$ such that $E \neq \mathbf{F}_p^a$, Th. 6.4 shows that there is $\sigma \in G$ such that $\sigma \neq 1$ and $\sigma$ fixes $E$. It follows that $\sigma \notin H$ and so $H \neq G$. We observe that by Th. 6.6, $\bar{H} = G$; this may be expressed by saying that $H$ is *dense* in $G$.

## Exercises

(1) A group is called *locally cyclic* if every finitely generated subgroup is cyclic. Show that a locally cyclic torsion group is of the form $A = \oplus A_p$, where $A_p$, the $p$-primary component, is either cyclic of order $p^m$ (for some $m \geqslant 0$) or of type $\mathbf{Z}(p^\infty)$. Deduce that each locally cyclic torsion group can be described up to isomorphism by the supernatural number whose factors are the orders of the cyclic subgroups.

(2) Let $K$ be any field. Show that the torsion subgroup of $K^\times$ is locally cyclic of type $N$, where $N$ is divisible by 2 if $\mathrm{char}\, K = 0$, and $N$ is of the form

$$N = \lim_{m|M} (p^m - 1),$$

(where $M$ is supernatural) if $\mathrm{char}\, K = p \neq 0$. Verify that all supernatural numbers $M$ can occur.

(3) Show that any union of a tower of abelian extensions is abelian.

(4) Let $p$ be a prime and $k = k_0 \subset k_1 \subset \cdots$ a tower of fields such that $k_n/k$ is cyclic of degree $p^n$. Show that $K = \bigcup k_n$ is an abelian extension of $k$ and that all finite quotients of $\mathrm{Gal}(K/k)$ have $p$-power order. Verify that $\mathrm{Gal}(K/k) = \mathbf{Z}_p$, the additive group of $p$-adic integers.

(5) Show that $\mathrm{Gal}(\mathbf{F}_p^a/\mathbf{F}_p) = \hat{\mathbf{Z}} = \varprojlim \mathbf{Z}/n$, and deduce that the dual of this group is $\mathrm{Hom}(\hat{\mathbf{Z}}, \mathbf{Q}/\mathbf{Z}) \cong \mathbf{Q}/\mathbf{Z}$.

(6) Show that $\hat{\mathbf{Z}} = \lim \mathbf{Z}/n$ is a topological ring, and $\hat{\mathbf{Z}} \cong \prod \mathbf{Z}_p$ is a ring homomorphism. Verify that every closed subgroup of $\hat{\mathbf{Z}}$ is an ideal.

(7) Let $E/k$ be a Galois extension with group $G$. Given $u_1, \ldots, u_r \in E$, if

$$\sum b_i (u_i^\alpha - u_i^\beta) = 0,$$

for fixed $b_i \in E$ not all 0, a fixed $\alpha \in G$ and all $\beta \in G$, then $1, u_1, \ldots, u_r$ are linearly dependent over $k$.

(8) Let $E$ be the extension of $\mathbf{Q}$ obtained by adjoining all square roots. Show that $G = \mathrm{Gal}(E/\mathbf{Q})$ is a 2-group, uncountable, hence with an uncountable basis, as vector space over $\mathbf{F}_2$. Thus $G$ has uncountably many subgroups of index 2, but $\mathbf{Q}$ has only countably many quadratic extensions. Deduce that $G$ has non-closed subgroups of finite index.

## 5.7 Galois cohomology

Let $E$ be a field with an automorphism $\sigma$ and let $V$ be a vector space over $E$. By a *$\sigma$-semilinear* endomorphism of $V$ we understand a mapping $\alpha_\sigma : V \to V$ such that

$$(x + y)\alpha_\sigma = x\alpha_\sigma + y\alpha_\sigma \qquad x, y \in V, \tag{1}$$

$$(\lambda x)\alpha_\sigma = \lambda^\sigma x \alpha_\sigma \qquad \lambda \in E. \tag{2}$$

To give an example, taking $k$ to be a subfield fixed by $\sigma$, let $V_0$ be a vector space over $k$ and put $V = E \otimes_k V_0$. In terms of a basis $(u_i)$ of $V_0$ we can write every element of $V$ uniquely as $x = \sum \xi_i u_i$, where $\xi_i \in E$. Now the mapping $\alpha_\sigma$ defined by

$$x\alpha_\sigma = \sum \xi_i^\sigma u_i \tag{3}$$

is $\sigma$-semilinear, as is easily verified. Suppose that $E/k$ is a Galois extension with group $G$, and for some $k$-space $V_0$, the mappings $\alpha_\sigma$ are defined on $V = E \otimes_k V_0$ by (3). Then $V_0$ can be recovered as the set of elements of $V$ fixed by all the $\alpha_\sigma$; for clearly, by (3), $x\alpha_\sigma = x$ holds iff $\xi_i^\sigma = \xi_i$ and this holds for all $\sigma \in G$ iff $\xi_i \in k$. This process of passing from $V$ to $V_0$ by means of the $\alpha_\sigma$ is called *Galois descent*. A vector space $V$ over $E$ which is of the form $E \otimes_k V_0$ for some $k$-space $V_0$ is called a space with a *$k$-form* $V_0$ and any $k$-subspace fixed by $G$ is said to be *$k$-rational*. In such a vector space we always have a family of semilinear mappings $\alpha_\sigma$ defined as in (3), and moreover, as is easily seen,

$$\alpha_{\sigma\tau} = \alpha_\sigma \alpha_\tau, \qquad \alpha_1 = 1. \tag{4}$$

Conversely, any such family of mappings on an $E$-space $V$ defines a $k$-form on $V$. For the proof we recall that if $a_1, \ldots, a_r$ is a basis of a finite Galois extension $E/k$, then the matrix $(a_i^\sigma)(\sigma \in G)$ is invertible (cf. the proof of Prop. 3.9.5, A.2).

PROPOSITION 7.1 *Let $E/k$ be a finite Galois extension with group $G$. Given a vector space $V$ over $E$ and a family of mappings $\alpha_\sigma (\sigma \in G)$ of $V$ into itself, satisfying* (1), (2), (4), *denote by $V_0$ the subset of $V$ fixed by all the $\alpha_\sigma$. Then $V_0$ is a $k$-form of $V$.*

*Proof.* It is clear that $V_0$ is a $k$-subspace of $V$ and the mapping

$$\lambda \otimes x \mapsto \lambda x \tag{5}$$

of $E \otimes V_0$ into $V$ is $k$-linear; we claim that it is an isomorphism. To find its kernel, take a basis $(\gamma_i)$ of $E$ over $k$. Any element of $E \otimes V_0$ can be written as $\sum \gamma_i \otimes x_i (x_i \in V_0)$ and if $\sum \gamma_i x_i = 0$, then $\sum \gamma_i^\sigma x_i = 0$ for all $\sigma \in G$. Since the matrix $(\gamma_i^\sigma)$ is invertible (cf. A.2, **3.9**), it follows that $x_i = 0$ and so the kernel is 0. Now let $f$ be a linear form on $V$ which vanishes on $V_0$ and take $x \in V$. For any $\lambda \in E$, $y_\lambda = \sum_\sigma (\lambda x)\alpha_\sigma$ belongs to $V_0$, hence $f(y_\lambda) = 0$, i.e. $\sum \lambda^\sigma f(x\alpha_\sigma) = 0$. By Dedekind's lemma (Lemma 3.5.1 of A.2), $f(x\alpha_\sigma) = 0$ for all $\sigma \in G$. Taking $\sigma = 1$ we find that $f(x) = 0$, and so $f = 0$. This shows the mapping (5) to be an isomorphism. ■

When $E/k$ is a finite Galois extension with group $G$, then $G$ acts on $E^\times$ and we may consider the cohomology groups $H^n(G, E^\times)$. More generally we can consider the action of $G$ on matrices over $E$, by defining, for $A = (a_{ij})$, $A^\sigma = (a_{ij}^\sigma)$, and so obtain $H^n(G, \mathbf{GL}_r(E))$. Since $\mathbf{GL}_r(E)$ is non-abelian for $r > 1$, this may no longer be a group, but we can formally define cocycles and coboundaries as before, and we shall find that $H^1(G, \mathbf{GL}_r(E))$ is trivial, i.e. every 1-cocycle is a coboundary.

PROPOSITION 7.2 (Speiser 1919) *Let $E/k$ be a finite Galois extension with group $G$. Given a family $U_\sigma \in \mathbf{GL}_r(E), \sigma \in G$, there exists $P \in \mathbf{GL}_r(E)$ such that*

$$U_\sigma = (P^\sigma)^{-1} P \tag{6}$$

*if and only if*

$$U_{\sigma\tau} = U_\sigma^\tau U_\tau \qquad \sigma, \tau \in G. \tag{7}$$

The equations (7) are often called *Noether's equations*, at least in the scalar case.

*Proof.* If (6) holds, then $U_\sigma^\tau U_\tau = (P^{\sigma\tau})^{-1} P^\tau (P^\tau)^{-1} P = U_{\sigma\tau}$. Conversely, assume (7) and consider the elements of $E^r$ as rows, with the $k$-form defined by $x^\sigma = (x_1^\sigma, \ldots, x_r^\sigma)$. The mappings

$$x \mapsto x^\sigma U_\sigma \tag{8}$$

form a family of semilinear mappings satisfying (1), (2), (4), as is easily verified. Hence the subspace $V_0$ fixed by these mappings is a $k$-form, by Prop. 7.1. Take a basis $u_1, \ldots, u_r$ of $V_0$ and let $P$ be the matrix formed by these rows. Then $P$ is invertible and $P^\sigma U_\sigma = P$, because the $u$'s are fixed under (8); thus we have $U_\sigma = (P^\sigma)^{-1} P$. ■

In particular, since $E^\times = \mathbf{GL}_1(E)$, this shows $H^1(G, E^\times)$ to be trivial. Consider the special case of Prop. 7.2 where $G$ is cyclic of order $n$, with generator $\sigma$. If $U_\sigma = C$, then by induction on $i$ we have

$$U_{\sigma^i} = C^{\sigma^{i-1}} \cdots C^\sigma C,$$

and so in particular,

$$I = U_{\sigma^n} = C^{\sigma^{n-1}} \cdots C^{\sigma} C. \tag{9}$$

Now Prop. 7.2 states that any matrix $C$ over $E$ satisfying (9) has the form $C = (P^{\sigma})^{-1}P$. In (9) let us take $r = 1$; then we have an element of norm 1 and Prop. 7.2 takes the following form. It was first stated as Theorem 90 of Hilbert's Zahlbericht [1897] and is generally known as 'Hilbert's Theorem 90':

COROLLARY 7.3 *Let $E/k$ be a cyclic Galois extension and $\sigma$ a generator of the Galois group. Then $c \in E$ has the form $c = (a^{\sigma})^{-1}a$ for some $a \in E$ if and only if $N(c) = 1$. Moreover, $a$ is unique up to a factor in $k$.*

Only the last part still needs proof, and this is clear, because if $(a^{\sigma})^{-1}a = (a'^{\sigma})^{-1}a'$, then $(a'a^{-1})^{\sigma} = a'a^{-1}$, hence $a'a^{-1} \in k$. ■

In contrast to Prop. 7.2, the groups $H^n(G, E^{\times})$ for $n \neq 1$ are not generally trivial. Thus $H^0(G, E^{\times}) = k^{\times}/N_{E/k}(E^{\times})$, while $H^2(G, E^{\times})$ gives rise to the Brauer group and will be discussed further in Ch. 7.

Corresponding results exist for the action of $G$ on the additive group of $E$; in order to prove them we need another result of independent interest, the normal basis theorem. Given a finite Galois extension $E/k$, any basis of the form $a^{\sigma} (\sigma \in G)$, consisting of all the conjugates of an element $a$ of $E$, is called a *normal basis*. We know that the extension $E/k$ is simple, so there exists $a \in E$ such that the $a^{\sigma}$ are all distinct, but there is nothing so far to tell us when we have a normal basis. In fact every finite Galois extension has a normal basis. To prove this we begin with a preparatory result on the algebraic independence of automorphisms.

Let $E/k$ be a finite Galois extension and take any $k$-basis $e_1, \ldots, e_n$ of $E$. Any element $a \in E$ has the form $a = \sum \alpha_i e_i$, where $\alpha_i \in k$, and so

$$a^{\sigma} = \sum \alpha_i e_i^{\sigma}, \tag{10}$$

where $\sigma$ ranges over $G = \mathrm{Gal}(E/k)$. We saw already that the matrix $(e_i^{\sigma})$ is invertible. Conversely, given any $n$ elements $e_1, \ldots, e_n \in E$ such that $(e_i^{\sigma})$ is invertible, it follows that $e_1, \ldots, e_n$ are linearly independent over $k$, and hence form a basis.

PROPOSITION 7.4 *Let $k$ be an infinite field and $E/k$ a Galois extension with group $G = \{\sigma_1, \ldots, \sigma_n\}$. If $f \in E[x_1, \ldots, x_n]$ is a polynomial such that $f(u^{\sigma_1}, \ldots, u^{\sigma_n}) = 0$ for all $u \in E$, then $f = 0$.*

*Proof.* Let us take a basis $e_1, \ldots, e_n$ of $E/k$ and define a polynomial $g$ by the equations

$$g(y_1, \ldots, y_n) = f(\sum y_i e_i^{\sigma_1}, \ldots, \sum y_i e_i^{\sigma_n}).$$

Then $g(a_1,\dots,a_n)=0$ for all $a_i\in k$, by hypothesis. Since $k$ is infinite, the polynomial $g$ is identically zero. But the matrix $(e_j^{\sigma_i})$ is invertible; if its inverse is $(p_{ij})$ and $x_j=\sum y_i e_i^{\sigma_j}$, then $y_j=\sum x_i p_{ij}$ and

$$f(x_1,\dots,x_n)=g(y_1,\dots,y_n)=0,$$

so $f$ vanishes identically. ■

We can now prove

THEOREM 7.5 (Normal basis theorem) *Any finite Galois extension $E/k$ has a normal basis.*

*Proof.* (i) Assume that $k$ is infinite and let $G=\{\sigma_1,\dots,\sigma_n\}$ be the Galois group. We know that a family of $n$ elements $e_1,\dots,e_n$ of $E$ is a basis iff the matrix $(e_i^{\sigma_j})$ is invertible, hence the sequence $(a^{\sigma_i})$ is a normal basis iff $(a^{\sigma_i\sigma_j})$ is an invertible matrix. Let us put

$$\sigma_i\sigma_j=\sigma_{r(i,j)}, \tag{11}$$

and consider the following polynomial over $k$:

$$f(x_1,\dots,x_n)=\det(x_{r(i,j)}).$$

We claim that $f$ does not vanish identically. For by (11) we see that for any $j$, $r(i,j)=r(i',j)\Rightarrow i=i'$; thus for fixed $j$, the mapping $i\mapsto r(i,j)$ is a permutation of $1,\dots,n$. Hence each row of the matrix of the determinant $f(1,0,\dots,0)$ has one non-zero entry, equal to 1, and likewise each column, so $f(1,0,\dots,0)$ is the determinant of a permutation matrix, hence equal to $\pm 1$. This establishes the claim that $f\neq 0$. Now

$$f(x^{\sigma_1},\dots,x^{\sigma_n})=\det(x^{\sigma_i\sigma_j}),$$

and by Prop. 7.4 there exists $a\in E$ such that $\det(a^{\sigma_i\sigma_j})\neq 0$, hence $(a^{\sigma_i})$ is then a normal basis.

(ii) $k$ is finite. In this case $E/k$ is cyclic, with generator $\sigma$, say. We may regard $\sigma$ as a linear transformation of $E$ as $k$-space. Let $\mu(x)$ be its minimal polynomial; then $\deg\mu\leqslant n=[E:k]$. But by Dedekind's lemma, $1,\sigma,\dots,\sigma^{n-1}$ are linearly independent over $E$, hence also over $k$, and so $\deg\mu=n$. Since $\sigma^n=1$, this must be the minimal equation. It follows that $\sigma$ is a cyclic transformation, i.e. $E$ is spanned by the conjugates of a single element (Th. 1 of **11.3**, A.1). Thus there exists $a\in E$ such that $a,a^{\sigma},\dots,a^{\sigma^{n-1}}$ span $E$, and by counting we see that they form a basis. ■

The conclusion may be stated by saying that $E$, considered as a $G$-module for its additive structure $E^+$ has the form

$$E^+=kG,$$

where $kG$ is the group algebra of $G$ over $k$. This means that $E^+$ is an induced, hence a coinduced module (cf. **4.1**), and since any coinduced module has trivial cohomology, by Cor. 3.3.6, we obtain

THEOREM 7.6 *For any finite Galois extension $E/k$ with group $G$, and any $n \geqslant 0$, $H^n(G, E^+) = 0$.* ■

We note again the case of a cyclic Galois group. In this case the additive cocycles correspond to elements of trace 0 and we obtain

COROLLARY 7.7 *For any cyclic extension $E/k$ with generating automorphism $\sigma$, an element $c \in E$ has the form $c = b - b^\sigma$ for some $b \in E$ if and only if* $\mathrm{tr}(c) = 0$. ■

For example, if $E/k$ is an extension of degree $p$ where $\mathrm{char}\, k = p$, this shows that there exists $b \in E$ such that $b^\sigma = b + 1$.

For infinite algebraic Galois extensions similar results may be obtained, using the topological Galois groups introduced in **5.6**, cf. Serre [1962].

### Exercises

(1) Show that Prop. 7.4 fails when $k$ is finite.

(2) Let $E/k$ be a finite Galois extension. Use Prop. 3.9.5, A.2 to show that there exists $c \in E$ such that $\mathrm{tr}(c) \neq 0$. Deduce that $H^1(G, E^+)$ is trivial.

(3) Show that $H^1(G, E^+) = 0$ by embedding $E^+$ in $\mathbf{GL}_2(E)$.

(4) Verify that $H^0(G, E^\times) = k^\times / N_{E/k}(E^\times)$.

(5) Show that the normal basis of $\mathbf{F}_{q^n}/\mathbf{F}_q$, say $c, c^q, \ldots$, can be chosen so that $c$ is a generator of $\mathbf{F}_{q^n}^\times$. Can it also be chosen to be a non-generator?

## 5.8 Kummer extensions

Abelian extensions are of importance because they include all extensions of finite fields, and in the case of an algebraic number field $K$ it is possible to describe the abelian extensions of $K$ by subgroups of $K^\times$ itself. This is the subject of class field theory, which is beyond the framework of this book (cf. Neukirch [1986]), but the special case of Kummer theory can be described in elementary terms.

We recall that a group $G$ is said to be of *exponent* $m$ if $x^m = 1$ for all $x \in G$ and $m$ is the least positive integer with this property. A Galois extension $E/k$ is said to be of *exponent* $m$ if its group has exponent dividing $m$; this then means

that $\sigma^m = 1$ for all $\sigma \in \mathrm{Gal}(E/k)$. By a *Kummer extension* one understands an abelian extension $E/k$ of finite exponent $m$, where $k$ contains $m$ distinct $m$th roots of 1. It follows that such an extension has characteristic prime to $m$.

Throughout the discussion the ground field $k$ will be fixed, and all extensions lie in a given algebraic closure $k^a$ of $k$. Moreover, we shall assume that $k$ contains all $m$th roots of 1: $\omega_1 = 1, \omega_2, \ldots, \omega_m$ say. Then for any $a \in k$, the roots of $x^m = a$ are $\alpha = \omega_1\alpha,\ \omega_2\alpha, \ldots, \omega_m\alpha$. When $a \neq 0$, these roots are distinct, and each generates the same field over $k$, which may unambiguously be written $k(a^{1/m})$. We shall put

$$k^{\times m} = \{c^m \mid c \in k^\times\},$$

and for any subgroup $P$ of $k^\times$ containing $k^{\times m}$ we shall write $k(P^{1/m})$ or $E_P$ for the subfield of $k^a$ generated by all subfields $k(c^{1/m})$ for $c \in P$. From what has been said, $E_P$ is determined uniquely as subfield of $k^a$, and it is normal over $k$, for if $c \in P$, then $x^m - c$ splits in $E_P$. Since its zeros are distinct, $E_P/k$ is also separable, and therefore a Galois extension. It is in fact a Kummer extension, for it is generated over $k$ by the elements $\alpha$ such that $\alpha^m \in P$. If $\sigma, \tau \in G = \mathrm{Gal}(E_P/k)$, then $\alpha^\sigma = \omega\alpha$, $\alpha^\tau = \omega'\alpha$, hence $\alpha^{\sigma\tau} = \alpha^{\tau\sigma} = \omega\omega'\alpha$ and $\alpha^{\sigma^m} = \omega^m\alpha = \alpha$, so that $\sigma^m = 1$. This shows that $G$ is abelian of exponent dividing $m$.

Given any Kummer extension $E/k$ of exponent $m$, let us write $N = \{a \in k^\times \mid a = \alpha^m \text{ for some } \alpha \in E\}$. In the particular case where $E = E_P$ it is easily verified that $P \subseteq N$; below (in Th. 8.1) we shall see that equality holds here. Let us put $G = \mathrm{Gal}(E/k)$; then there is a natural homomorphism $N \to \hat{G} = \mathrm{Hom}(G, k^\times)$ defined as follows. Given $a \in N$, choose $\alpha \in E$ such that $\alpha^m = a$; then for any $\sigma \in G$ we have $\alpha^\sigma = \omega_\sigma\alpha$, where $\omega_\sigma^m = 1$. Thus the mapping

$$f_a : \sigma \mapsto \omega_\sigma = \alpha^\sigma\alpha^{-1} \tag{1}$$

is a homomorphism $G \to k^\times$, since $\alpha^{\sigma\tau} = (\omega_\sigma\alpha)^\tau = \omega_\sigma\omega_\tau\alpha$. Moreover, $\omega_\sigma$ depends only on $a$, not on $\alpha$: if $\alpha'$ is another root of $x^m = a$, then $\alpha' = \lambda\alpha$, where $\lambda^m = 1$, and so $\lambda \in k$, hence $\alpha'^\sigma = \lambda\alpha^\sigma = \omega_\sigma\lambda\alpha = \omega_\sigma\alpha'$. Thus $f_a \in \hat{G}$; moreover the mapping $a \mapsto f_a$ is a homomorphism, for if $\alpha^m = a$, $\beta^m = b$, then $(\alpha\beta)^m = ab$ and $(\alpha\beta)^\sigma(\alpha\beta)^{-1} = \alpha^\sigma\alpha^{-1}\beta^\sigma\beta^{-1}$, so $f_{ab} = f_a f_b$. This may be summed up by saying that the mapping $G \times P \to k^\times$ given by

$$\sigma, a \mapsto \langle \sigma, a \rangle = \alpha^\sigma\alpha^{-1}, \qquad \text{where } \alpha^m = a, \tag{2}$$

is bimultiplicative, i.e. multiplicative in each argument. This mapping (2) can be used to give the following description of Kummer extensions.

THEOREM 8.1 *Let $k$ be a field containing $m$ distinct $m$th roots of* 1 *(and hence of characteristic prime to $m$), let $P$ be a subgroup of $k^\times$ including $k^{\times m}$ and write $E_P = k(P^{1/m})$. Then $E_P/k$ is a Kummer extension of exponent $m$ and if $\mathrm{Gal}(E_P/k) = G$, we have an exact sequence*

$$1 \to k^\times \to P \xrightarrow{f} \hat{G} \to 1, \tag{3}$$

*where $f$ is the mapping $a \mapsto f_a$ defined by* (1). *The extension $E_P/k$ is finite if and only if the group $P/k^{\times m}$ is finite, and when this is so, then*

$$|G| = [E_P:k] = (P:k^{\times m}). \tag{4}$$

*Moreover, every Kummer extension of exponent $m$ is of this form, so that $P \mapsto E_P$ is a bijection.*

*Proof.* We have already seen that $E_P/k$ is a Kummer extension of exponent $m$, and it is clear that $N$ defined before satisfies $N \supseteq P$. Consider the mapping (2); if we think of $G$ and $P$ as being written additively, then this is a bilinear mapping and it therefore defines a duality. The kernel in $G$ is 1, for if $\langle \sigma, a \rangle = 1$ for all $a \in P$, then $\sigma$ fixes $E_P$ and so $\sigma = 1$. The kernel in $E_P$ is $k^{\times m}$, for if $\langle \sigma, a \rangle = 1$ for all $\sigma$, then $\alpha = a^{1/m}$ satisfies $\alpha^\sigma = \alpha$ for all $\sigma$, and so $\alpha \in k$, i.e. $a \in k^{\times m}$. It follows that the groups $P/k^{\times m}$ and $G$ are dual to each other, in particular, $P/k^{\times m} \cong \hat{G}$ whenever either side is finite, and then

$$(P:k^{\times m}) = |\hat{G}| = |G| = [E_P:k], \tag{5}$$

where the last equality follows by Galois theory and the second by duality theory. We remark that $P/k^{\times m} \cong \hat{G}$ even in the infinite case, provided that $G$ is correctly interpreted, cf. Ex. (5).

Now consider the homomorphism $f: N \to \hat{G}$ defined by (1). If $\chi \in \hat{G}$, then since $\chi_\sigma \in k$, we have

$$\chi_{\sigma\tau} = \chi_\sigma \chi_\tau = \chi_\sigma^\tau \chi_\tau,$$

hence $\chi$ satisfies Noether's equations and so by Cor. 7.3, $\chi_\sigma = \alpha^\sigma \alpha^{-1}$ for some $\alpha \in E_P$. Since $\chi_\sigma^m = 1$, $(\alpha^m)^\sigma = \alpha^m$, i.e. $\alpha^m \in k$, so $f$ is surjective. The kernel consists of all $a \in N$ such that $\alpha^\sigma = \alpha$ for $\alpha = a^{1/m}$ and all $\sigma \in G$, i.e. $a \in k^{\times m}$. This establishes the exact sequence (3), with $N$ in place of $P$. It follows that

$$(N:k^{\times m}) = |\hat{G}|,$$

and for finite extensions this together with (5) shows that $N = P$. In general $E_P$ is the union of its subextensions $E_{P_i}$ for $P_i$ such that $(P_i:k^{\times m})$ is finite, and from this the equality $N = P$ follows in the general case. Finally if $E/k$ is a Kummer extension of exponent $m$ and $N$ is defined as before, then it is clear that $E_N = k(N^{1/m}) = E$, so the correspondence $P \mapsto E_P$ is indeed a bijection. ■

In a field of characteristic $p \neq 0$ there is an analogous theory for extensions of exponent $p^n$. For the case $n > 1$ this requires Witt vectors, and the description would take us too far afield, but we shall briefly outline the case $n = 1$.

Let $E/k$ be a cyclic extension of degree $p$ where $\operatorname{char} k = p$; the element $1 \in k$ satisfies $\operatorname{tr}(1) = p \cdot 1 = 0$, hence by Cor. 7.7 there exists $\alpha \in E$ such that $\alpha^\sigma - \alpha = 1$, where $\sigma$ is a generator of $\operatorname{Gal}(E/k)$. Thus we have $\alpha^\sigma = \alpha + 1$, so the conjugates

of $\alpha$ are $\alpha, \alpha+1, \ldots, \alpha+p-1$ and the minimal equation for $\alpha$ is

$$x^p - x - a = 0, \qquad \text{where} \qquad a = \prod_{i=1}^{p} (\alpha + i). \tag{6}$$

This is to be regarded as the analogue of the binomial equation in characteristic $p$. We shall write $\wp(x) = x^p - x$; then (6) may be written $\wp(x) = a$ and a root of this equation will be denoted by $\wp^{-1}a$. Let $P$ be a subgroup of the additive group $k^+$ containing $\wp k = \{\wp a | a \in k\}$ and put $E_P = k(\wp^{-1}P)$, the field obtained by adjoining all the elements $\wp^{-1}a, a \in P$, to $k$. Writing $G = \mathrm{Gal}(E_p/k)$, we have a mapping $G \times P \to \mathbf{F}_p$ such that

$$\sigma, a \mapsto [\sigma, a\rangle = \alpha^\sigma - \alpha, \qquad \text{where } \wp\alpha = a. \tag{7}$$

It is easily verified that this mapping is mutiplicative in the first and additive in the second argument. Moreover, its kernel in $G$ is 1 and in $P$ it is $\wp k$. We thus obtain the following analogue of Th. 8.1:

THEOREM 8.2 *Given a field $k$ of characteristic $p \neq 0$, and a subgroup $P$ of $k^+$ containing $\wp k$, define $E_P = k(\wp^{-1}P)$. Then $E_P/k$ is an abelian extension of exponent $p$, and if $G = \mathrm{Gal}(E_P/k)$, we have an exact sequence*

$$0 \to \wp k^+ \to P \xrightarrow{g} \hat{G} \to 1, \tag{8}$$

*where $g$ is the mapping $a \mapsto g_a$ and $g_a : \sigma \mapsto [\sigma, a\rangle$. The extension $E_P/k$ is finite if and only if the group $P/\wp k$ is finite, and when this is so, then*

$$|G| = [E_P : k] = (P : \wp k).$$

*Moreover, every abelian extension of exponent $p$ is of this form.*

The proof is exactly analogous to that of Th. 8.1, using Cor. 7.7 in place of Cor. 7.3. ■

Extensions $E_P/k$ of the form described in Th. 8.2 are sometimes called *Artin–Schreier extensions.*

## Exercises

In Exercises (1)–(3) $k$ is any field containing a primitive $m$th root of 1.

(1) Show that when $x^m - a$ and $x^m - b$ are irreducible, they have the same minimal splitting field over $k$ iff $ba^r \in k^{\times m}$ for some $r$ prime to $m$.

(2) Show that the finite Kummer extensions of exponent $m$ are just the minimal splitting fields of $(x^m - a_1) \cdots (x^m - a_r)$. What are the degrees of these extensions?

(3) Show that any field $k$ as above is contained in a unique maximal Kummer extension $K$ of exponent $m$, of degree $(k^\times : k^{\times m})$. Deduce that $\mathbf{Q}$ has infinitely many non-isomorphic quadratic extensions.

(4) Let $\operatorname{char} k = p \neq 0$. Show that $x^p - x = a$ is a normal equation over $k$, for any $a \in k$.

(5) Let $E_P/k$ be an infinite Kummer extension. Show that every homomorphism from $P/k^{\times m}$ into the group $\mathbf{U}_m(k)$ of $m$th roots of 1 in $k$ is realized by the form (2), for a suitable element $\sigma$ of $G$, while a homomorphism $\sigma$ of $G$ into $\mathbf{U}_m(k)$ is realized by (2), for some $a \in k$ iff $\sigma$ is continuous, as homomorphism of $G$ (with the Krull topology) into $\mathbf{U}_m(k)$ with the discrete topology.

(6) Let $E_P/k$ be an infinite Artin–Schreier extension. Show that every homomorphism from $P$ to $\mathbf{F}_p$ is realized by the form (7) for a suitable element $\sigma$ of $G = \operatorname{Gal}(E_P/k)$; a homomorphism $\sigma$ of $G$ into $\mathbf{F}_p$ is realized by the form (7) for some $a \in k$ iff $\sigma$ is continuous.

## Further exercises on Chapter 5

(1) Show that two uncountable algebraically closed fields are isomorphic iff they have the same characteristic and the same cardinal. Does this remain true in the countable case?

(2) Let $\mathbf{F}_2(x, y)$ be the rational function field in two indeterminates over the field $\mathbf{F}_2$ of two elements, and consider the subset $S = \{x, y, xy, x + y\}$. Show that under algebraic dependence the independent subsets are the subsets of $S$ of at most two elements. Show that there is no vector space over $\mathbf{F}_2$ with a subset $\{a, b, c, d\}$ such that the linearly independent subsets are precisely the subsets of cardinal at most 2.

(3) Find the subfields of $\mathbf{C}(x)$ corresponding to the cyclic and the dihedral groups.

(4) (Abhyankar) Let $f, g$ be polynomials over $k$ of degrees $m, n$ respectively. Given a simple transcendental extension $k(x)$, write $u = f(x)$, $v = g(x)$. Show that $u, v$ satisfy an equation of degree $n$ in $u$ and $m$ in $v$, monic in $v$ if $f$ was monic. (Hint: use the resultant to express the fact that $f(t) - u$ and $g(t) - v$ have a common zero.)

(5) Let $k$ be a field of characteristic $p$ with perfect closure $\hat{k}$. Show that an irreducible polynomial $f$ over $k$ can over $\hat{k}$ be written as $f = g^q$, where $q = p^r$, and $g$ is irreducible with distinct zeros.

(6) Let $E/k$ be separable and $F/k$ $p$-radical and form $L = E \otimes_k F$. Verify that $L$ is $p$-radical over $E$; deduce that $L$ has no zero-divisors and hence is a field.

(7) Let $E/k$ be any extension in characteristic $p$, $F = k(\alpha)$ a $p$-radical extension with minimal equation $\alpha^q = a \in k$ and let $\alpha^r = \alpha_1$ be the least power of $\alpha$ in $E$. Show that $E \otimes_k F$ is the $E$-algebra generated by $x, y$ subject to $x^r = \alpha_1 - y, y^{q/r} = 0$. (Hint: prove the case $r = 1$ first.)

(8) Let $k$ be any field. Show that the field of formal Laurent series $k((x))$ has non-trivial derivations over $k(x)$.

(9) Let $E$ be a field of prime characteristic $p$. Given $x_1,\ldots,x_n, y\in E$, $y$ is said to be *p-dependent* on $x_1,\ldots,x_n$ if $y\in E^p(x_1,\ldots,x_n)$, and otherwise *p-independent*. Verify that this is a dependence relation. Show that $E/k$ is separable iff any $p$-independent family in $k$ remains $p$-independent in $E$.

(10) Let $L/k$ be a finite field extension containing subextensions $E/k$, $F/k$ of degrees $r, s$ respectively. Show that if $r, s$ are coprime, then the subfield generated by $E, F$ has degree $rs$ over $k$. Deduce that $E$ and $F$ are linearly disjoint over $k$.

(11) Let $L/k$ be a finite Galois extension and $E/k, F/k$ subextensions. Put $G = \mathrm{Gal}(L/k)$ and let $H, K$ be the subgroups of $G$ fixing $E, F$ respectively. Show that $EF$ has the group $H\cap K$ and that $E, F$ are linearly disjoint over $k$ iff $HK = G$. (Hint: express the condition $[EF:k] = [E\otimes F:k]$ in terms of $G$.)

(12) With the notation of Ex. (11), show that $E, F$ have no non-trivial isomorphic subextensions iff $H$ and $x^{-1}Kx$ generate $G$, for any $x\in G$. Find subgroups $H, K$ of $\mathrm{Sym}_4$ to satisfy this condition, but $HK \neq G$. Deduce the existence of a Galois extension with group $\mathrm{Sym}_4$ and two subfields whose tensor product has zero-divisors.

(13) Show that if $E/k$ is regular, then for any commutative $k$-algebra $A$ there is a natural bijection between the prime ideals of $A$ and those of $A\otimes E$. Show that if $k$ is separably closed in $E$, then for any $F/k$, $F$ is separably closed in $E\otimes F$. Further show that if $E/k$, $F/k$ are regular, then the field of fractions of $E\otimes F$ over $k$ is regular.

(14) Let $k$ be a field of characteristic 0 and $K = k((x))$ the field of formal Laurent series in $x$. Given $f\in K$, find $y = x + c_2x^2 + \cdots \in K$ such that $f = ay^r$ $(a\in k, r\in\mathbf{Z})$. Deduce that for any $f\in K\backslash k$ there is an automorphism of $K$ fixing $k$ but moving $f$.

(15) (G. Cauchon) Let $E$ be a field, $\alpha$ an automorphism of $E$ of infinite order and $F$ the fixed subfield of $\alpha$. Show that for any $n > 1$ there is just one subextension of $E$ of degree $n$ over $F$, namely the fixed field of $\alpha^n$.

(16) (W. Waterhouse) Let $G$ be any profinite group and $F/k$ a Galois extension with a surjective homomorphism $f: G\to \mathrm{Gal}(F/k)$. With a set $X$ indexed by the transversals of open normal subgroups of $G$ form $L = F(X)$ (treating the elements of $X$ as indeterminates) and define an action of $G$ on $L$ by $(N\sigma)^\tau = N\sigma\tau$ for $N\sigma\in X$ and $\tau\in G$, and $a^\sigma = a^{\sigma f}$ for $a\in F$. Verify that this action is faithful and the stabilizer of any element is an open subgroup. Use Prop. 6.7 to show that $G\cong \mathrm{Gal}(L/E)$, where $E$ is the fixed subfield of $G$, and $E\cap F = k$.

(17) Given a finite Galois extension $E/k$, let $f$ be a polynomial over $E$ which is such that its zeros are permuted by any element of $\mathrm{Gal}(E/k)$. Show that there exists $c\in E^\times$ such that $cf$ has coefficients in $k$.

(18) Examine Kummer extensions of finite fields.

(19) Let $k$ be a field of characteristic $p \neq 0$. Show that if $x^p - x - a$ is irreducible over $k$ and has a zero $\alpha$ in an extension field, then $x^p - x - \alpha^{p-1}a$ is irreducible over $k(\alpha)$. Deduce that if $k$ has a separable extension of degree $p$, then it has separable extensions of arbitrarily high degree.

# 6
# Algebras

In **5.3** of A.2 we saw that semisimple Artinian rings can be described quite explicitly as direct products of full matrix rings over skew fields (Wedderburn's theorem). Below in Ch. 9, 10 we shall see what can be said when the Artinian hypothesis is dropped, but in many cases, such as the study of group algebras in finite characteristic, it is important to find out more about the non-semisimple (but Artinian) case. There is now a substantial theory of such algebras which is still developing, and a full treatment is beyond the framework of this book, but some of the basic properties are described here.

One of the main results, the Krull–Schmidt theorem, asserts the uniqueness of decompositions of a module as a direct sum of indecomposables. This is established here for finitely generated modules over Artinian rings, but as we shall see, for projective modules it holds over the somewhat larger class of semiperfect rings. These rings are also of interest because they allow the construction of a projective cover for each finitely generated module.

The rest of the chapter is concerned with conditions for two rings to have equivalent module categories (Morita equivalence), with flat modules and their relation to projective and injective modules, and with conditions for lifting past the radical in a finite-dimensional algebra.

## 6.1 The Krull–Schmidt theorem

A finitely generated module over a general Artinian ring may not be semisimple, i.e. a direct sum of simple modules, but it can always be written as a direct sum of indecomposable modules. Moreover, these indecomposable summands do not depend on the decomposition chosen, but are unique up to isomorphism. This is the content of the Krull–Schmidt theorem; our aim in this section is to prove this result, but we shall do so in a slightly more general form.

We recall that a *local ring* is a ring $R$ in which the set of all non-units forms an ideal $\mathfrak{m}$; this is then the unique maximal ideal and it is easily seen that $R/\mathfrak{m}$ is a skew field, called the *residue class field* of $R$. From the definition it is clear that a ring is local precisely if the sum of any two non-units is a non-unit or, equivalently, for any non-unit $c$, $1-c$ is a unit. The maximal ideal of a local ring is its Jacobson radical; since the latter is nilpotent in any Artinian ring

(A.2, Th. 5.4.5), it follows that an Artinian ring is local iff every element is either nilpotent or a unit. Such rings arise naturally as endomorphism rings of indecomposable modules, as we shall now see. In what follows we shall write our modules as left modules and put module homomorphisms on the right, except when otherwise stated.

LEMMA 1.1 (Fitting's lemma) *Let $R$ be any ring and $M$ an $R$-module of finite composition length. Given any endomorphism $\alpha$ of $M$, there exists a direct decomposition*

$$M = M_0 \oplus M_1, \tag{1}$$

*such that $M_0, M_1$ admit $\alpha$, and $\alpha$ restricted to $M_0$ is nilpotent, while its restriction to $M_1$ is an automorphism.*

*Proof.* We have $M \supseteq M\alpha \supseteq M\alpha^2 \supseteq \cdots$, $0 \subseteq \ker \alpha \subseteq \ker \alpha^2 \subseteq \cdots$; since $M$ has finite length, there exists $n$ (at most equal to the length of $M$) such that $M\alpha^n = M\alpha^{n+1} = \cdots$, $\ker \alpha^n = \ker \alpha^{n+1} = \cdots$. We put $M_1 = M\alpha^n$, $M_0 = \ker \alpha^n$ and then have $M_1\alpha^n = M\alpha^{2n} = M\alpha^n = M_1$. Thus for any $x \in M$, $x\alpha^n = x_1\alpha^n$ for some $x_1 \in M_1$; hence $x \in M_0 + M_1$, so $M = M_0 + M_1$ and this sum is direct, for if $y \in M_0 \cap M_1$, then $y = x\alpha^n$, hence $0 = y\alpha^n = x\alpha^{2n}$. Thus $x \in \ker \alpha^{2n} = \ker \alpha^n$ and $y = x\alpha^n = 0$. This shows the sum to be direct and it establishes the decomposition (1); clearly $\alpha$ is nilpotent on $M_0$ and bijective on $M_1$. ■

If in this lemma $M$ is indecomposable then $M_0$ or $M_1$ is the zero module and we obtain

COROLLARY 1.2 *Let $M$ be an indecomposable $R$-module of finite length. Then any endomorphism of $M$ is either nilpotent or an automorphism, thus $\mathrm{End}_R(M)$ is a local ring.* ■

This result does not hold without some restriction on the module $M$ (cf. Ex. (5)), but we remark that conversely, any module with local endomorphism ring is indecomposable. For if $M$ is decomposable, say $M = M_1 \oplus M_2$, then the projection on $M_1, e_1$ say, is an idempotent of $\mathrm{End}_R(M)$ which for a non-trivial decomposition is neither 0 nor 1. Hence $1 - e_1$ is not a unit and so $\mathrm{End}_R(M)$ cannot be local. Sometimes a module with local endomorphism ring is called 'strongly indecomposable'.

In the sequel we shall state our conclusions for modules with local endomorphism rings. By Cor. 1.2 they will apply to any indecomposable modules of finite length, in particular to any finitely generated indecomposable modules over Artinian rings, but they will also apply in some other cases.

LEMMA 1.3 *Let $R$ be any ring and $V$ an indecomposable $R$-module, such that*

$\mathrm{End}_R(V) = E$ *is a local ring with maximal ideal* $\mathfrak{m}$. *Given any* $R$*-module* $M$ *and homomorphisms* $\alpha: V \to M$, $\beta: M \to V$, *we have* $\alpha\beta \in \mathfrak{m}$ *unless* $V$ *is a direct summand of* $M$.

*Proof.* Suppose that $\alpha\beta \notin \mathfrak{m}$. Then $\alpha\beta$ is a unit in $E$, so there exists $\gamma \in E$ such that $\alpha\beta\gamma = \gamma\alpha\beta = 1$. This shows that $\alpha$ is injective and we have an exact sequence

$$0 \to V \xrightarrow{\alpha} M \to \mathrm{coker}\,\alpha \to 0.$$

The sequence is split by $\beta\gamma$, hence $M \cong V \oplus \mathrm{coker}\,\alpha$. ■

In what follows let us fix a left $R$-module $V$ with local endomorphism ring $E = \mathrm{End}_R(V)$. We denote the maximal ideal of $E$ by $\mathfrak{m}$, put $K = E/\mathfrak{m}$ and write $x \mapsto \bar{x}$ for the natural homomorphism $E \to K$. Given any left $R$-module $M$, we can consider $\mathrm{Hom}_R(V, M)$ as left $E$-module in a natural way. We shall write $\overline{(V, M)} = \mathrm{Hom}_R(V, M)/\mathfrak{m}\,\mathrm{Hom}_R(V, M)$; this is a left $E$-module which is annihilated by $\mathfrak{m}$, so it can be defined as a left vector space over $K$ in a natural way. Similarly we can consider $\mathrm{Hom}_R(M, V)$ as a right $E$-module and hence define $\overline{(M, V)} = \mathrm{Hom}_R(M, V)/\mathrm{Hom}_R(M, V)\mathfrak{m}$ as a right $K$-space. Next we define a bilinear mapping

$$b: \overline{(V, M)} \times \overline{(M, V)} \to K$$

by the following rule: Given $\alpha \in \overline{(V, M)}$, $\beta \in \overline{(M, V)}$, take homomorphisms $f, g$ such that $\bar{f} = \alpha, \bar{g} = \beta$ and define

$$b(\alpha, \beta) = \overline{fg}. \tag{2}$$

This is a well defined operation on $\alpha, \beta$, for if $\bar{f} = \overline{f'}$, say $f' = f + \sum \lambda_i h_i$, where $\lambda_i \in \mathfrak{m}$, $h_i \in \mathrm{Hom}_R(V, M)$, then $\overline{f'g} = \overline{fg} + \sum \lambda_i \overline{h_i g} = \overline{fg}$, hence $\overline{fg}$ depends only on $\bar{f}$, not on $f$, and similarly for $g$. In this way (2) defines a pairing of the spaces $\overline{(V, M)}$, $\overline{(M, V)}$. We define the *rank* of $b$ in the usual way as the rank of the matrix obtained by taking bases. Thus if $(u_i)$ is a left $K$-basis of $\overline{(V, M)}$ and $(v_j)$ is a right $K$-basis of $\overline{(M, V)}$, then the rank of $b$ is given by the rank of the matrix $(b(u_i, v_j))$. Clearly this is independent of the choice of bases; we shall denote it by $\mu_V(M)$, thus

$$\mu_V(M) = \mathrm{rk}(b(u_i, v_j)). \tag{3}$$

Suppose now that $M$ is expressed as a direct sum: $M = \sum_1^s \oplus M_i$. It is clear that this gives rise to a direct sum decomposition for both $\overline{(V, M)}$ and $\overline{(M, V)}$:

$$\overline{(V, M)} = \oplus \overline{(V, M_i)}, \qquad \overline{(M, V)} = \oplus \overline{(M_i, V)}.$$

Here a homomorphism $f: V \to M_i$ corresponds to a homomorphism $V \to M$

which is obtained by combining $f$ with the canonical injection $M_i \to M$. Similarly a homomorphism $g: M_j \to V$ corresponds to the homomorphism $M \to V$ obtained by following the canonical projection $M \to M_j$ by $g$. It follows that the elements of $\overline{(V,M)}\cdot\overline{(M,V)}$ are $s \times s$ matrices with the members of $\overline{(V,M_i)}\cdot\overline{(M_i,V)}$ along the main diagonal and zeros elsewhere. By Lemma 1.3, $\overline{(V,M_i)}\cdot\overline{(M_i,V)}$ can be non-zero only if $M_i$ has $V$ as a direct summand; in particular it will be zero whenever $M_i$ is indecomposable and not isomorphic to $V$. All these arguments clearly still apply when $M$ is an infinite direct sum, and they allow us to draw the following conclusion:

PROPOSITION 1.4 *Let $R$ be a ring and $V$ an $R$-module with local endomorphism ring. Given an $R$-module $M$ which is expressed as a direct sum of indecomposable modules, the multiplicity of $V$ in this direct sum is independent of the sum chosen, and is equal to $\mu_V(M)$.*

*Proof.* Let the given direct decomposition of $M$ be

$$M = \oplus M_\lambda. \tag{4}$$

Then $\overline{(V,M)}\cdot\overline{(M,V)}$ is a direct sum of terms $\overline{(V,M_\lambda)}\cdot\overline{(M_\lambda,V)}$, by the earlier remarks, and since $M_\lambda$ is indecomposable, we have

$$\overline{(V,M_\lambda)}\cdot\overline{(M_\lambda,V)} = \begin{cases} K & \text{if } M_\lambda \cong V, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that the rank $\mu_\lambda(M)$ is equal to the number of terms $M_\lambda$ in (4) that are isomorphic to $V$, i.e. the multiplicity of $V$ in (4). ■

COROLLARY 1.5 *Given any ring $R$ and $R$-module $M$, let*

$$M = \oplus_I M_\lambda = \oplus_J N_\mu$$

*be two direct decompositions into indecomposable modules. If the components in at least one of these decompositions have local endomorphism rings, then there is a bijection $\lambda \mapsto \lambda'$ from $I$ to $J$ such that $M_\lambda \cong N_{\lambda'}$.*

*Proof.* Suppose that $\mathrm{End}_R(M_\lambda)$ is local for all $\lambda$. By Prop. 1.4, the multiplicity of each $M_\lambda$ is the same in both decompositions and this allows us to construct the desired bijection. ■

For Artinian rings the conclusion may be stated as follows:

THEOREM 1.6 (Krull–Schmidt theorem) *Let $R$ be any Artinian ring. Any finitely generated $R$-module $M$ has a finite direct decomposition*

$$M = M_1 \oplus \cdots \oplus M_r, \tag{5}$$

*where the terms $M_i$ are indecomposable, and this decomposition is unique up to isomorphism and the order of the terms; thus if also $M = N_1 \oplus \cdots \oplus N_s$, where each $N_j$ is indecomposable, then $s = r$ and there is a permutation $i \mapsto i'$ of $1, \ldots, r$ such that $M_i \cong N_{i'}$.*

*Proof.* Any finitely generated $R$-module over an Artinian ring $R$ is Artinian (A.2, Th. 2.2.8) and so has finite length (A.2, Th. 5.4.8). Hence we can form a direct decomposition (5) with a maximum number of terms, and all terms are then indecomposable. By Cor. 1.2 each endomorphism ring $\mathrm{End}_R(M_i)$ is local, so we can apply Cor. 1.5 to conclude that the decompositions have isomorphic terms (possibly after reordering). ■

Theorem 1.6 was stated for finite groups by Wedderburn in 1909, and first completely proved by Remak in 1911. It was later extended to abelian groups with operators by Krull in 1928 and to general groups with operators by O. Yu. Schmidt in 1928. In 1950 Azumaya noted that the finiteness condition could be replaced by the condition that the endomorphism rings of the components be local (Cor. 1.5). The above presentation is based on a lecture by J. A. Green.

## Exercises

(1) Show that a ring $R$ with Jacobson radical $J$ is local iff $R/J$ is a skew field.

(2) Show that in any ring an idempotent with a 1-sided inverse equals 1. Deduce that a ring $R$ in which for each $a \in R$ either $a$ or $1 - a$ has a 1-sided inverse must be local (here '$a$ has a 1-sided inverse' is taken to mean: either $ax = 1$ or $xa = 1$ has a solution).

(3) Give an example of a non-Artinian non-local ring in which every element is either nilpotent or a unit. (Hint: try the commutative case.)

(4) Show that a ring $R$ is indecomposable as module over itself iff $R$ contains no idempotent $\neq 0, 1$.

(5) Use Ex. (4) to give an example of an indecomposable module whose endomorphism ring is not local.

(6) Let $V$ be an indecomposable module which is not injective and let $I$ be its injective hull. Show that $\overline{(V, I)} \neq 0$ but $\overline{(V, I)} \cdot \overline{(I, V)} = 0$.

(7) Let $M = M_1 \oplus \cdots \oplus M_r$ and suppose that $V$ is an indecomposable module which is a direct summand of $M$. Show that $V$ is isomorphic to a direct summand of $M_i$ for some $i, 1 \leqslant i \leqslant r$.

## 6.2 The projective cover of a module

We have seen in **3.3** that every module has an injective hull. Dually one can define the projective cover of a module, but this does not exist for all modules. Later, in **6.3**, we shall meet general conditions for its existence, but for the moment we shall show that its existence is assured over Artinian rings. Throughout this section we shall limit ourselves to finitely generated modules; we recall that over an Artinian ring every finitely generated module has finite composition length (A.2, Th. 2.2.8 and Th. 5.4.8).

LEMMA 2.1 *Let $R$ be any ring with Jacobson radical $J$ and let $M$ be a finitely generated left $R$-module. Then*

$$JM \subseteq \bigcap M_\lambda, \tag{1}$$

*where $M_\lambda$ ranges over all maximal submodules of $M$. Moreover, if $R/J$ is semisimple, then equality holds in* (1) *and we have*

$$M/JM \cong \oplus S_\mu, \tag{2}$$

*where the $S_\mu$ are simple modules, quotients of $M$.*

*Proof.* Let $M_1$ be a maximal submodule of $M$. If $M_1 \not\supseteq JM$, then $M_1 + JM = M$, hence by Nakayama's lemma (Cor. 5.4.7 of A.2), $M_1 = M$, which is a contradiction. Thus $JM \subseteq M_1$ and (1) follows. Assume now that $\bar{R} = R/J$ is semisimple; then $\bar{M} = M/JM$ may be regarded as an $\bar{R}$-module and hence is semisimple, so we obtain (2), for a family $\{S_\mu\}$ of simple modules. If we combine the isomorphism (2) with the projection on $S_\mu$, we obtain a homomorphism $f_\mu: \bar{M} \to S_\mu$ whose kernel is a maximal submodule of $\bar{M}$, and so is of the form $N_\mu/JM$, where $N_\mu$ is a maximal submodule of $M$. Hence $\bigcap N_\mu/JM = 0$, i.e. $\bigcap N_\mu = JM$, and since the $N_\mu$ form a subfamily of the $M_\lambda$, we have equality in (1). ■

A module homomorphism $f: M \to N$ is called *essential* if it is surjective but its restriction to any proper submodule of $M$ fails to be surjective. By a *projective cover* of a module $M$ we shall understand a projective module $P$ with an essential homomorphism $P \to M$. Here the following lemma is useful for testing for essentiality:

LEMMA 2.2 *Let $R$ be any ring. Given an $R$-module $M$, a finitely generated projective $R$-module $P$ and a surjective homomorphism $\alpha: P \to M$, if $\ker \alpha \subseteq JP$, then $\alpha$ is essential. If $R/J$ is semisimple, then this sufficient condition is also necessary.*

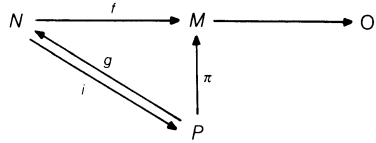
*Proof.* Let $P'$ be any submodule of $P$ such that $P'\alpha = M$. Then for any $x \in P$ there exists $x' \in P'$ such that $x\alpha = x'\alpha'$, i.e. $x \in P' + \ker \alpha$. Thus $P' + \ker \alpha = P$ and

by Nakayama's lemma, $P' = P$; this shows $\alpha$ to be essential. Suppose now that $R/J$ is semisimple and $\alpha$ is essential. Let $P_1$ be any maximal submodule of $P$; if $P_1 \not\supseteq \ker\alpha$, then $P_1 + \ker\alpha = P$, hence $P_1\alpha = P\alpha = M$, which contradicts the essentiality of $\alpha$. Therefore $\ker\alpha \subseteq P_1$ and now $\ker\alpha \subseteq \bigcap P_\lambda = JP$, by Lemma 2.1. ■

Our first task is to prove the existence of projective covers in the Artinian case:

PROPOSITION 2.3 *Let $R$ be a left Artinian ring. Then any finitely generated left $R$-module has a projective cover.*

*Proof.* Let $M$ be a finitely generated projective $R$-module; we can find a finitely generated projective module $P$ mapping onto $M$. Let us choose $P$ of shortest length; we claim that in this case the homomorphism $\pi: P \to M$ is essential. For let $N$ be a minimal submodule of $P$ such that $\pi|N$ is surjective. Since $P$ is projective, there is a map $g: P \to N$ such that $gf = \pi$, where $f = i\pi$.




Since $\pi$ maps $N$ onto $M$, $f$ maps $\mathrm{im}(g|N)$ onto $M$ and by the minimality of $N$ it follows that $\mathrm{im}(g|N) = N$. Thus $g|N$ is a surjective endomorphism of the module $N$ of finite length, therefore it is an automorphism and so $N \cap \ker g = 0$. Given $x \in P$, we have $xg \in N$ and since $g|N$ is an automorphism, we have $xg = yg^2$ for some $y \in P$. Now $x - yg \in \ker g$ and $x = yg + (x - yg)$; this shows that $P = N \oplus \ker g$. Therefore $N$ is projective, but this contradicts the minimality of $P$ unless $N = P$. ■

When a projective cover exists, it must be unique; this can be proved quite generally.

PROPOSITION 2.4 *If $P, Q$ are projective covers of a module $M$ (over any ring $R$), with essential homomorphisms $\alpha: P \to M, \beta: Q \to M$, then there is an isomorphism $\theta: P \to Q$ such that $\alpha = \theta\beta$.*

*Proof.* Since $P$ is a projective module and $\beta$ is surjective, there exists $\theta: P \to Q$ such that $\alpha = \theta\beta$. This map $\theta$ must be surjective, because $\beta$ is essential, and so $P$ splits over $\ker\theta$, say $P = Q_1 \oplus \ker\theta$, where $Q_1 \cong Q$. But then $Q_1\alpha = Q_1\theta\beta = M$, hence $Q_1 = P$ and so $\ker\theta = 0$. Thus $\theta$ is an isomorphism, as claimed. ■

We shall denote the projective cover of a module $M$ by $\mathbf{P}(M)$, bearing in mind that this operation $\mathbf{P}$ is not defined for all modules.

There is a second operation closely related to **P** which is defined for all modules. For any module $M$ we define its *top* as

$$\mathbf{T}(M) = M/JM, \tag{3}$$

where $J = \mathbf{J}(R)$ is the Jacobson radical. If $M$ is finitely generated and $\mathbf{T}(M) = 0$, then $M = 0$; this is just the content of Nakayama's lemma. When $M$ is finitely generated, then the natural homomorphism $\tau: M \to \mathbf{T}(M)$ is essential, for it is clearly surjective, and if $N \subset M$, then $N$ is contained in a maximal submodule $N_0$ of $M$. We have $N_0 \supseteq JM$, hence the natural map $v: M \to M/N_0$ can be factored as $M \to \mathbf{T}(M) \to M/N_0$, where $v|N_0 = 0$, and this contradicts the fact that $\tau|N_0$ is surjective. This shows $\tau$ to be essential. When $R$ is Artinian, or more generally, when $R/J$ is semisimple, then $\mathbf{T}(M)$ is semisimple, by Lemma 2.1.

We note that the projective cover of a module, when it exists, may be obtained by taking the projective cover of its top:

PROPOSITION 2.5 *For any ring $R$ and any $R$-module $M$ with a projective cover, we have* $\mathbf{P}(M) \cong \mathbf{P}(\mathbf{T}(M))$.

For the composition of two essential maps is clearly again essential and so we have the essential map

$$\mathbf{P}(M) \to M \to \mathbf{T}(M), \tag{4}$$

hence the result. ■

Similarly when forming the top, it does not matter whether we start from a given module or its projective cover.

PROPOSITION 2.6 *Let $R$ be a ring such that $R/J$ is semisimple, and let $M$ be a finitely generated $R$-module with a projective cover. Then*

$$\mathbf{T}(\mathbf{P}(M) \cong \mathbf{T}(M). \tag{5}$$

*Proof.* Write $P = \mathbf{P}(M)$ and let $N$ be the kernel of the essential map $\alpha: P \to M$. By Lemma 2.1, $N \subseteq JP$ and since $\alpha$ is surjective, it maps $JP$ onto $JM$. Therefore $\mathbf{T}(P) = P/JP \cong P/N/JP/N \cong M/JM = \mathbf{T}(M)$. ■

In particular this shows that the projective cover of a simple module has a simple top.

## Exercises

(1) Verify that **T** is a functor. Under what circumstances is **P** a functor?

(2) Show that the only finitely generated **Z**-modules with a projective cover are the free modules.

(3) Show that $\mathbf{T}(\mathbf{T}(M)) = \mathbf{T}(M)$, $\mathbf{P}(\mathbf{P}(M)) = \mathbf{P}(M)$.

(4) Show that if $\mathbf{P}(M)$ is indecomposable, then so is $M$. Does the converse hold?

(5) Show that if $\alpha: P \to M$ is a projective cover and $\beta: Q \to M$ is surjective, where $Q$ is a projective module, then $Q = P_0 \oplus P_1$ where $P_1 \cong P$ and $\beta | P_1$ corresponds to $\alpha$, while $\beta | P_0 = 0$. Deduce Prop. 2.4.

## 6.3 Semiperfect rings

Any Artinian ring is of finite composition length as left module over itself and so can be decomposed into a direct sum of indecomposable left ideals. Such decompositions can be described by the corresponding decompositions of 1 into a sum of idempotents. Let us recall that two idempotents $e$, $f$ of a ring $R$ are called *orthogonal* if $ef = fe = 0$; an idempotent $e$ is *primitive* if $e \neq 0$ and $e$ cannot be written as a sum of two non-zero orthogonal idempotents.

PROPOSITION 3.1 *Let $R$ be any ring. Any decomposition of $R$ as a direct sum of left ideals*

$$R = \mathfrak{a}_1 \oplus \cdots \oplus \mathfrak{a}_r \tag{1}$$

*corresponds to a decomposition of* 1 *as a sum of pairwise orthogonal idempotents*:

$$1 = e_1 + \cdots + e_r, \qquad e_i^2 = e_i, \qquad e_i e_j = 0 \quad \text{for } i \neq j, \tag{2}$$

*where $\mathfrak{a}_i = Re_i$, and for any idempotent $e$, $Re$ is indecomposable if and only if $e$ is primitive.*

*Proof.* Given (1), we write 1 as $\sum e_i$, where $e_i \in \mathfrak{a}_i$. Then $e_k = \sum_i e_k e_i$; by the directness we see that $e_k e_i = 0$ for $i \neq k$ and $e_k^2 = e_k$, so (2) follows. Conversely, given (2), put $\mathfrak{a}_i = Re_i$; then any $x \in R$ can be written as $x = \sum x e_i$, $xe_i \in \mathfrak{a}_i$, hence $R = \sum \mathfrak{a}_i$, and this decomposition is direct, for if $\sum x_i e_i = 0$, then right multiplication by $e_k$ gives $x_k e_k = 0$.

For any idempotent $e$, $Re$ is a direct summand of $R$: $R = Re + R(1 - e)$; now the above correspondence shows that any direct decomposition of $Re$ corresponds to writing $e$ as a sum of two orthogonal idempotents. It follows that $Re$ is indecomposable iff $e$ is primitive. ■

In the Artinian case one can construct decompositions (1) by writing the semisimple ring $R/J$ as a direct sum of simple left ideals and then 'lifting' this decomposition to $R$. The essential step is to lift an idempotent from $R/J$ to $R$; below we shall see how to do this for Artinian rings and then go on to describe a somewhat larger class of rings for which this is possible.

Let $R$ be any ring and $\mathfrak{N}$ an ideal of $R$; an element $u \in R$ such that $u^2 \equiv u \pmod{\mathfrak{N}}$ is called an idempotent mod $\mathfrak{N}$, and we say that $u$ can be *lifted* to $R$

if there exists $e \in R$ such that $e^2 = e$ and $e \equiv u \pmod{\mathfrak{N}}$. For such a lifting to be possible one usually has to assume that $\mathfrak{N} \subseteq \mathbf{J}(R)$, but this by itself is not enough. For example let $R$ be the ring of rational numbers with denominators prime to 6; then $\mathbf{J}(R) = 6R$ and 3, 4 are idempotents mod $6R$, which however cannot be lifted to $R$. The next result gives a sufficient condition.

LEMMA 3.2 *Let $R$ be a ring and $\mathfrak{N}$ a nil ideal in $R$. Then idempotents* mod $\mathfrak{N}$ *can be lifted to $R$.*

We recall that a *nil ideal* is an ideal consisting of nilpotent elements.

*Proof.* Let $u$ be an idempotent mod $\mathfrak{N}$, say $(u - u^2)^m = 0$. We have

$$1 = [u + (1-u)]^{2m} = \sum \binom{2m}{i} u^{2m-i}(1-u)^i.$$

On the right the first $m$ terms are divisible by $u^m$, while each term after the first $m$ is divisible by $(1-u)^m$, so on denoting the sum of the first $m$ terms by $e$, we can write $1 = e + (1-u)^m g$, where $g$ is a polynomial in $u$. Now $u(1-u) \in \mathfrak{N}$, so $e = u^{2m} + 2mu^{2m-1}(1-u) + \cdots + \equiv u \pmod{\mathfrak{N}}$, and $e(1-e) = e(1-u)^m g = 0$, hence $e$ is an idempotent. ■

In an Artinian ring $R$ the Jacobson radical $\mathbf{J}(R)$ is nilpotent, so in this case any idempotent mod $\mathbf{J}(R)$ can be lifted to $R$, by Lemma 3.2, but there are other cases where this is possible and it is convenient to make a definition at this point (due to H. Bass (1960)):

DEFINITION A ring $R$ is said to be *semiperfect* if idempotents (mod $\mathbf{J}(R)$) can be lifted to $R$ and $R/\mathbf{J}(R)$ is semisimple.

It is clear from the definition that this notion is left–right symmetric. Moreover, since for any ring $R$, $R/\mathbf{J}(R)$ has zero radical, it is semisimple iff it is right (or equivalently, left) Artinian, by Th. 5.4.5, A.2. In particular, this establishes

THEOREM 3.3 *Every left (or right) Artinian ring is semiperfect.* ■

To describe semiperfect rings we shall need to know when two idempotents generate isomorphic ideals. Let us call two idempotents $e, f$ in a ring $R$ *conjugate* if $R$ contains a unit $u$ such that $f = u^{-1}eu$. If there exist $a \in eRf$, $b \in fRe$ such that $ab = e$, $ba = f$, we shall call $e$ and $f$ *isomorphic*. For example, conjugate idempotents are isomorphic, for if $f = u^{-1}eu$, we can take $a = eu$, $b = u^{-1}e$ and then find that $a = euf$, $b = fu^{-1}e$ and $ab = e$, $ba = f$. The next lemma clarifies the relation between these concepts.

LEMMA 3.4 *Let $e, f$ be any idempotents in a ring $R$. Then* (i) *$e, f$ are isomorphic if and only if $eR \cong fR$ or equivalently, $Re \cong Rf$,* (ii) *$e, f$ are conjugate if and only if $e$ is isomorphic to $f$ and $1-e$ is isomorphic to $1-f$.*

*Proof.* (i) Assume that $Re \cong Rf$, say $\theta: Re \to Rf$ is an isomorphism and suppose that $\theta$ maps $e$ to $a$, while $\theta^{-1}$ maps $f$ to $b$. Since $a \in Rf$, we have $af = a$ and $e(e\theta) = e\theta$, i.e. $ea = a$. Hence $eaf = af = a$ and similarly $fbe = b$. Further, $e = e\theta\theta^{-1} = a\theta^{-1} = (af)\theta^{-1} = a \cdot f\theta^{-1} = ab$, and similarly $f = ba$, so $e$ is isomorphic to $f$. Conversely, if $e, f$ are isomorphic, say $ab = e$, $ba = f$, where $a \in eRf$, $b \in fRe$, then $x \mapsto xa$ is a homomorphism from $Re$ to $Rf$ with inverse $y \mapsto yb$. Hence $Re \cong Rf$ iff $e, f$ are isomorphic, and by symmetry this is equivalent to $eR \cong fR$.

(ii) If $e, f$ are conjugate, say $f = u^{-1}eu$, then as we saw, $e$ and $f$ are isomorphic. Further, $1 - f = u^{-1}(1-e)u$, so $1-e$, $1-f$ are also isomorphic. Conversely, if $e, f$ are isomorphic and $1-e$, $1-f$ are isomorphic, say $ab = e$, $ba = f$, where $a = eaf$, $b = fbe$, and $a'b' = 1-e$, $b'a' = 1-f$, $a' = (1-e)a'(1-f)$, $b' = (1-f)b'(1-e)$, let us put $A = a + a'$, $B = b + b'$. Then $b'(1-e) = b'$, so $b'e = 0$ and similarly $ea' = 0$ and hence $BeA = bea + b'ea + bea' + b'ea' = bea = f$. For the same reasons $B(1-e)A = 1-f$, and hence $BA = 1$. By symmetry $AfB = e$, $A(1-f)B = 1-e$, and so $AB = 1$. This shows $e, f$ to be conjugate, as claimed. ∎

We shall usually want to lift orthogonal families of idempotents; this can be accomplished without further hypotheses on the ring:

PROPOSITION 3.5 *Let $R$ be any ring and $e, f$ idempotents in $R$. Write $J = \mathbf{J}(R)$ and denote the natural homomorphism $R \to \bar{R} = R/J$ by $x \mapsto \bar{x}$. Then*

(i) *$e = 0$ or $1$ if and only if $\bar{e} = 0$ or $1$ respectively;*

(ii) *$e$ is isomorphic to $f$ if and only if $\bar{e}$ is isomorphic to $\bar{f}$; in particular, if $\bar{e} = \bar{f}$, then $e$ is conjugate to $f$;*

(iii) *if $ef \equiv fe \equiv 0 \pmod{J}$, then there exists an idempotent $f_1$ such that $f_1 \equiv f \pmod{J}$ and $ef_1 = f_1 e = 0$.*

*Proof.* (i) If $e \in J$, then $1-e$ is a unit and since $e(1-e) = 0$, we have $e = 0$. Similarly if $1 - e \in J$ and the converse is clear.

(ii) Let $a, b \in R$ be such that $a \equiv eaf$, $b \equiv fbe$, $ab \equiv e$, $ba \equiv f \pmod{J}$ and put $a_1 = eaf$, $b_1 = fbe$. Then $a_1 b_1 = e - z$, where $z \in eJe$. Let $z'$ be the quasi-inverse of $z$ (i.e. $z + z' = zz' = z'z$) and put $z'' = ez'e$; then $z''$ is a quasi-inverse of $z$ in $eRe$, and it follows that $a_1 b_1 (e - z'') = e$. Putting $a_2 = a_1$, $b_2 = b_1(e - z'')$, we have $a_2 \equiv a$, $b_2 \equiv b \pmod{J}$ and $a_2 b_2 = e$. Next write $b_2 a_2 = f - y$; then $y \in fJf$ and since $(b_2 a_2)^2 = b_2 e a_2 = b_2 a_2$, it follows that $f - y = (f-y)^2 = f^2 - fy - yf + y^2 = f - 2y + y^2$. Thus $y(y-1) = 0$, and since $y \in J$, we find that $y = 0$ and so $b_2 a_2 = f$. Thus $e$ is isomorphic to $f$; the rest is clear.

(iii) If $ef \equiv fe \equiv 0 \pmod{J}$, then $1 - fe$ is invertible. Put

$$f_0 = (1 - fe)^{-1} f (1 - fe);$$

then $f_0$ is an idempotent conjugate to $f$. Moreover, $f_0 \equiv f \pmod{J}$ and clearly $f_0 e = 0$. Writing $f_1 = (1 - e) f_0$, we have $f_1 = f_0 - e f_0 \equiv f_0 - ef \equiv f_0 \equiv f \pmod{J}$ and $f_1 e = 0 = e f_1$; moreover, $f_1^2 = (1 - e) f_0 (1 - e) f_0 = (1 - e) f_0^2 = f_1$, so $f_1$ is the required idempotent. ■

We shall use this result to lift decompositions of the form (2), again without further hypotheses:

PROPOSITION 3.6 *In any ring $R$, let $e_1, \ldots, e_r$ be a set of idempotents such that $e_i e_j \equiv 0 \pmod{J}$ for $i \neq j$. Then there exist idempotents $e_i'$ such that $e_i' \equiv e_i \pmod{J}$ and $e_i' e_j' = 0$ for $i \neq j$. Moreover, if*

$$1 = e_1 + \cdots + e_r, \qquad e_i e_j \equiv 0 \pmod{J}, \qquad i \neq j,$$

*then there exist idempotents $e_i'$ such that $e_i' \equiv e_i \pmod{J}$, $1 = \sum e_i'$ and $e_i' e_j' = 0$ for $i \neq j$.*

*Proof.* For $n = 1$ there is nothing to prove, so we assume that $n > 1$ and use induction on $n$. This means that we may assume $e_i e_j = 0$ for $i \neq j$, $i, j > 1$. Put $e = e_2 + \cdots + e_r$; then $e$ is again idempotent and $e_1 e \equiv e e_1 \equiv 0 \pmod{J}$, By Prop. 3.5 there exists an idempotent $e_1'$ such that $e_1' \equiv e_1 \pmod{J}$ and $e e_1' = e_1' e = 0$. It follows that $e_1', e_2, \ldots, e_r$ are pairwise orthogonal idempotents.

Assume now that (3) holds, and choose the $e_i'$ as in the first part. Then $e = e_1' + \cdots + e_r'$ is an idempotent such that $e \equiv 1 \pmod{J}$, hence $e = 1$ by Prop. 3.5, (i). ■

The first part remains true (with the same proof) for a countable set of idempotents, but it ceases to hold for uncountable sets (Zelinsky 1954).

Prop. 3.6 shows that every semiperfect ring $R$ can be written as $R = \sum R e_i$, where the $e_i$ are primitive and so the $Re_i$ are indecomposable. To establish the uniqueness of such decompositions we shall want to apply the Krull–Schmidt theorem and we need to check that the endomorphism ring of an indecomposable left ideal $Re$ is local. Here we shall need a couple of elementary lemmas:

LEMMA 3.7 *A semiperfect ring in which 1 is a primitive idempotent is a local ring.*

*Proof.* The semisimple ring $R/J$ can be written as a direct sum of simple left ideals, and this decomposition can be lifted to $R$. Since 1 is primitive, 1 (mod $J$) is primitive, hence there is a single summand and so $R/J$ is a skew field. This means that $R$ is a local ring. ■

LEMMA 3.8 *Let $R$ be any ring, $e$ an idempotent in $R$ and $M$ a left $R$-module. Then there is an isomorphism of left $eRe$-modules:*

$$\mathrm{Hom}_R(Re, M) \cong eM, \tag{4}$$

*In particular, taking $M = Re$, we obtain a ring isomorphism*

$$\mathrm{End}_R(Re) \cong eRe. \tag{5}$$

*Proof.* Each homomorphism $\alpha: Re \to M$ is completely determined by its effect on $e$. If $e\alpha = u$, then $x\alpha = (xe)\alpha = x(e\alpha) = xu$; in particular, $u = e\alpha = eu \in eM$. Conversely, for any $u \in eM$ the mapping $\alpha_u: xe \mapsto xu$ is a homomorphism from $Re$ to $M$, and the correspondence $u \mapsto \alpha_u$ is additive, as is easily checked. It is surjective, by what we have seen, and if $\alpha_u = 0$, then $u = eu = 0$, so it is injective and hence an isomorphism of abelian groups, or more generally, left $eRe$-modules. This establishes (4). If $M = Re$, both sides acquire a multiplicative structure, which is again compatible with the isomorphism, so we have a ring isomorphism (5). ■

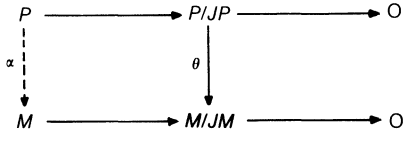
Over a semiperfect ring $R$, the top of any finitely generated $R$-module $M$ can be written

$$\mathbf{T}(M) = \oplus S_\mu,$$

where each $S_\mu$ is a simple quotient of $M$, by Lemma 2.1. We shall use this remark to show that projective covers exist over a semiperfect ring.

THEOREM 3.9 *Any finitely generated module over a semiperfect ring $R$ has a projective cover. More precisely, the projective module $P$ is a projective cover for $M$ if and only if $\mathbf{T}(P) \cong \mathbf{T}(M)$.*

*Proof.* Let $P$ be a projective cover for $M$, say $P/N \cong M$, where $N \subseteq JP$, by Lemma 2.1. Then $JM \cong JP/N$, hence $\mathbf{T}(M) = M/JM \cong P/N/JP/N \cong P/JP = \mathbf{T}(P)$, so the condition is necessary. Now let $M$ be any finitely generated left $R$-module and put $\bar{R} = R/J$. As we have just seen, $\mathbf{T}(M)$ is a finite direct sum of simple $R$-modules, which may also be regarded as simple $\bar{R}$-modules; further, any simple $\bar{R}$-module has the form $\bar{R}\bar{e}$, where $\bar{e}$ is a primitive idempotent in $\bar{R}$. By definition of $R$, $\bar{e}$ lifts to an idempotent $e$ in $R$, which is again primitive, by Prop. 3.5. Write $P_i = Re_i$ for the corresponding indecomposable projective and put $P = \oplus P_i$. Then $\mathbf{T}(P) \cong \oplus \bar{R}\bar{e}_i \cong \mathbf{T}(M)$. More generally, given any projective module $P$ and an isomorphism $\theta: \mathbf{T}(P) \to \mathbf{T}(M)$, we have a diagram



Since $P$ is projective, there exists $\alpha$ to make the diagram commutative. Given $x \in M$, there exists $a \in P$ such that $\bar{a}\theta = \bar{x}$, hence $P\alpha + JM = M$, and so by Nakayama's lemma, $P\alpha = M$. Since $\theta$ is injective, $\ker \alpha \subseteq JP$; therefore by Lemma 2.2, $P$ is a projective cover for $M$. ■

This result and its proof allows us to view finitely generated modules over a semiperfect ring in a new light. With every such module $M$ we associate on the one hand its top $T$ and on the other its projective cover $P$. We have essential mappings

$$P \to M, \qquad M \to T,$$

and $P, T$ are the largest resp. smallest modules for which such essential mappings exist, for a given $M$. We conclude with an analogue of the Krull–Schmidt theorem for projective modules over semiperfect rings.

THEOREM 3.10 *Let $R$ be a semiperfect ring. Every finitely generated projective left $R$-module $P$ can be written as a direct sum*

$$P = P_1 \oplus \cdots \oplus P_r, \tag{6}$$

*where each $P_i$ is isomorphic to an indecomposable left ideal which is a direct summand of $R$, and the $P_i$ are unique up to order and isomorphism.*

*Proof.* We have seen that $R$ has a direct decomposition into indecomposable left ideals, e.g. by lifting a direct decomposition of $R/J$. Hence for any $n \geqslant 1$, $R^n$ can be written as a direct sum of indecomposable modules isomorphic to left ideals; by Lemmas 3.8 and 3.7 each such left ideal has as endomorphism ring a local ring, therefore Cor. 1.5 can be applied to establish the uniqueness of this decomposition, up to isomorphism and order. Now if $P$ is any finitely generated projective left $R$-module, we have $P \oplus P' \cong R^n$ for some $n \geqslant 1$, and so obtain a decomposition $P \oplus P' \cong \oplus Q_i$, where each $Q_i$ is isomorphic to an indecomposable left ideal. If we now take a direct decomposition of $P$ with maximal number of terms and apply Prop. 1.4, we obtain a decomposition of the required form (6). ■

It can be shown that semiperfect rings form the precise class of rings for which every finitely generated module has a projective cover. A ring over which every left module has a projective cover is said to be *left perfect*. Such a ring $R$ is characterized by the fact that $R/J$ is semisimple and $J$ is *right vanishing*, i.e. for any sequence $\{a_v\}$ in $J$ there exists $n$ such that $a_1 \cdots a_n = 0$ (cf. e.g. Faith [1981]).

## Exercises

(1) Show that an idempotent $e$ in a ring $R$ is primitive iff $eRe$ is non-trivial and has no idempotents $\neq 0, e$.

(2) Let $R$ be a ring and $\mathfrak{a}$ a minimal left ideal. Show that either $\mathfrak{a}^2 = 0$ and $\mathfrak{a}R$ is a nilpotent two-sided ideal in $R$, or $R\mathfrak{a} = Re$ for some idempotent $e$, and hence $\mathfrak{a}$ is a direct summand in $R$.

(3) Find conditions on idempotents $e, f$ for $Re = Rf$.

(4) Show that if $Re \cong Rf$ and $e$ is central, then $f = fe = ef$. If $f$ is also central, deduce that $e = f$.

(5) Let $R$ be a local ring and $P$ a finitely generated projective left $R$-module. Show by lifting a basis of $\mathbf{T}(P)$ that $P$ is free.

(6) Let $R$ be a semiperfect ring such that $R/J$ is simple. Show that $R$ is a full matrix ring over a local ring. If further, $R$ is an integral domain, deduce that it must be a local ring.

(7) Let $e$ be a central idempotent in a ring $R$. Show that if $e_1$ is an idempotent such that $e_1 \equiv e \pmod{J}$, then $e_1 = e$.

(8) Show that if $1 = \sum e_i = \sum f_i$ are two decompositions of 1 into orthogonal families of idempotents such that $e_i \equiv f_i \pmod{J}$, then $u = \sum e_i f_i$ is a unit and $u^{-1} e_i u = f_i$.

(9) Show that if $R$ is semiperfect, then so is $R_n$ for all $n \geqslant 1$.

(10) Show that a commutative Artinian ring is a direct product of completely primary rings (i.e. every non-unit is nilpotent). Give a counterexample in the non-commutative case, using triangular matrix rings.

## 6.4 Equivalence of module categories

A natural question asks when two rings $A$, $B$ have equivalent module categories. We recall from A.2, **4.3** that two rings $A$, $B$ are said to be *Morita equivalent*, $A \sim B$, or simply *equivalent* if there is a category equivalence $\mathscr{M}_A \cong \mathscr{M}_B$. There we also saw that any ring $A$ is equivalent to $A_n$ for all $n \geqslant 1$; however there are other cases and in this section and the next we shall find explicit conditions for $A$ and $B$ to be Morita equivalent. We begin by describing the notion of a generator which plays an important role in what follows.

By a *generator* in an abelian category $\mathscr{A}$ one understands an object $P$ in $\mathscr{A}$ such that $h^P = \mathscr{A}(P, -)$ is faithful. An equivalent condition is that every $\mathscr{A}$-object is a quotient of a copower of $P$. For if $P$ is a generator and for $X \in Ob\,\mathscr{A}$ we put

$$S = \coprod P_f,$$

where $P_f \cong P$ and $f$ runs over $\mathscr{A}(P, X)$, with natural injection $i_f : P_f \to S$; then the family of maps $f : P_f \to X$ gives rise to a map $F : S \to X$ such that $f = i_f F$. To establish that $X$ is a quotient of $S$ we show that $F$ is epic. Let $g : X \to \operatorname{coker} F$ be the natural map. Then $Fg = 0$, hence $fg = 0$ for all $f \in \mathscr{A}(P, X)$ and since $P$

is a generator, it follows that $g = 0$. Thus coker $F = 0$ and $F$ is epic, as claimed.

Conversely, assume that $X$ is a quotient of $S = \coprod P_\lambda$, where $P_\lambda \cong P$, $F: S \to X$ and write $i_\lambda: P_\lambda \to S$ for the natural injection. Given $f: X \to Y$, $f \neq 0$, we have $Ff \neq 0$ because $F$ is epic, so $i_\lambda Ff \neq 0$ for some $\lambda$, but $i_\lambda F \in \mathscr{A}(P_\lambda, X)$. This shows $P$ to be a generator.

It is clear that in the category $\mathscr{M}_R$ of all right $R$-modules, $R$ is a generator. The existence of a generator gives rise to a useful criterion for a natural isomorphism of functors.

THEOREM 4.1 *Let $\mathscr{A}$, $\mathscr{B}$ be abelian categories with coproducts and let $F, G$ be functors from $\mathscr{A}$ to $\mathscr{B}$ which are right exact and preserve coproducts. If there is a natural transformation*

$$t: F \to G, \tag{1}$$

*such that for some generator $P$ of $\mathscr{A}$, the map*

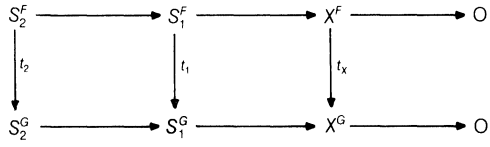
$$t_P: P^F \to P^G \tag{2}$$

*is an isomorphism, then $t$ is a natural isomorphism.*

*Proof.* For any $X \in Ob\,\mathscr{A}$ we have a short exact sequence

$$S_2 \to S_1 \to X \to 0,$$

where $S_1, S_2$ are copowers of $P$. Applying $F$ and $G$ we have the following commutative diagram with exact rows (by right exactness of $F$ and $G$):




By hypothesis $S_2^F = (\coprod P_\lambda)^F = \coprod P_\lambda^F$, hence $t_2 = t_{S_2}$ is an isomorphism, and likewise $t_1$. By the 5-lemma, $t_X$ is an isomorphism, as asserted. ■

Let us now consider right exact functors $\mathscr{M}_A \to \mathscr{M}_B$ which preserve direct sums (coproducts). An example of such a functor is $\otimes_A M$, where $M$ is an $(A, B)$-bimodule. Our next result shows that this is essentially the only case:

LEMMA 4.2 (Eilenberg–Watts) *For any functor $S: \mathscr{M}_A \to \mathscr{M}_B$ between module categories the following assertions are equivalent:*

(a) *$S$ has a right adjoint $T: \mathscr{M}_B \to \mathscr{M}_A$;*

(b) *$S$ is right exact and preserves direct sums;*

(c) *$S = \otimes_A P$ for some $(A, B)$-bimodule $P$.*

*Moreover, when (a)–(c) hold, then the right adjoint $T$ is given by*

$$Y^T = \mathrm{Hom}_B(P, Y),$$

*where $P = A^S$, and $T$ is unique up to natural isomorphism.*

*Proof.* (a)⇒(b) follows by Th. 3.2.7, since $S$ is a left adjoint. (b)⇒(c). Given a functor $S$, right exact and preserving direct sums, put $P = A^S$. Each $A$-endomorphism of $A_A$ induces a $B$-endomorphism of $P$. But the $A$-endomorphisms of $A_A$ are just the left multiplications by elements of $A$ (A.2, Th. 5.1.3), thus $P$ is an $(A, B)$-bimodule. Now consider the functors $S$ and $\otimes_A P$: both are right exact and preserve direct sums, so to show their isomorphism we need only, by Th. 4.1, find a natural transformation between them which, for the generator $A$ of $\mathscr{M}_A$ is an isomorphism. Given $X \in \mathscr{M}_A$, we have a map

$$X \xrightarrow{\cong} \mathrm{Hom}_A(A, X) \xrightarrow{S} \mathrm{Hom}_B(A^S, X^S) \cong \mathrm{Hom}_B(P, X^S). \tag{3}$$

The result is a map $f_X: X \to \mathrm{Hom}_B(P, X^S)$ which is an $A$-module homomorphism. By adjoint associativity we have

$$f_X \in \mathrm{Hom}_A(X, \mathrm{Hom}_B(P, X^S)) \cong \mathrm{Hom}_B(X \otimes P, X^S).$$

Each step in (3) is natural in $X$, so $f_X$ is a natural transformation from $X \otimes P$ to $X^S$. For $X = A$ it clearly reduces to the identity, hence it is a natural isomorphism, as claimed.

(c)⇒(a). Given (c), we define a functor $T: \mathscr{M}_B \to \mathscr{M}_A$ by

$$Y \mapsto \mathrm{Hom}_B(P, Y).$$

Then $\mathrm{Hom}_B(X^S, Y) \cong \mathrm{Hom}_B(X \otimes P, Y) \cong \mathrm{Hom}_A(X, \mathrm{Hom}_B(P, Y)) \cong \mathrm{Hom}_B(X, Y^T)$, where each step is natural in $X$ and $Y$. Thus $T$ is the required right adjoint of $S$.

When (a)–(c) hold, we have

$$Y^T = \mathrm{Hom}_A(A, Y^T) \cong \mathrm{Hom}_B(A^S, Y) \cong \mathrm{Hom}_B(P, Y),$$

where $P = A^S$, and here $P$ is clearly unique up to isomorphism. ■

We shall introduce a preordering of module categories on the basis of the next result:

PROPOSITION 4.3 *For any rings $A$, $B$ the following are equivalent:*

(a) *there exists a functor $S: \mathscr{M}_A \to \mathscr{M}_B$ with right adjoint $T: \mathscr{M}_B \to \mathscr{M}_A$ such that $T$ has a right adjoint and $ST \cong 1$;*

(b) *there exist functors $S: \mathscr{M}_A \to \mathscr{M}_B$, $T: \mathscr{M}_B \to \mathscr{M}_A$, both right exact and preserving direct sums, such that $ST \cong 1$;*

(c) *there exist modules ${}_AP_{B,B}Q_A$ such that $P \otimes_B Q \cong A$, as bimodules.*

If one (and hence all) of (a)–(c) holds, we shall call $\mathscr{M}_A$ a *quotient category* of $\mathscr{M}_B$ and write $A \prec B$. The functor $S$ is called the *section functor* and $T$ the *retraction functor.*

*Proof.* (a)⇒(b) follows by Lemma 4.2 and (b)⇒(a) will follow if we show $T$ to be the right adjoint of $S$. This follows because we have the natural transformation

$$\mathrm{Hom}_B(X^S, Y) \xrightarrow{T} \mathrm{Hom}_A(X^{ST}, Y^T) \cong \mathrm{Hom}_A(X, Y^T),$$

which is an isomorphism for $X = A$.

Moreover, if (b) holds, then by Lemma 4.2, $S \cong \otimes_A P$, $T \cong \otimes_B Q$, for some ${}_AP_B, {}_BQ_A$, and $ST \cong 1$ means that $P \otimes Q \cong A$, so (c) holds. Conversely, this condition clearly implies (a). ■

We list some properties of quotient categories:

PROPOSITION 4.4 *Let $A$, $B$ be rings such that $A \prec B$, with modules ${}_AP_B, {}_BQ_A$ satisfying $P \otimes Q \cong A$. Then*

(i) $Q \cong \mathrm{Hom}_B(P, B)$, $P \cong \mathrm{Hom}_B(Q, B)$;

(ii) $A \cong \mathrm{End}_B(P) \cong \mathrm{End}_B(Q)$;

(iii) $P_B$ *and* ${}_BQ$ *are projective*;

(iv) ${}_AP$ *and* $Q_A$ *are generators*;

(v) *we have the following lattice homomorphisms with right inverses ('retractions'):*

$\mathrm{Lat}(A_A) \to \mathrm{Lat}(P_B)$ *with 2-sided ideals of $A$ corresponding to $(A, B)$-submodules of $P$,*

$\mathrm{Lat}({}_AA) \to \mathrm{Lat}({}_BQ)$ *with 2-sided ideals of $A$ corresponding to $(B, A)$-submodules of $Q$.*

*Proof.* In each case it is enough to prove the first part; the second then follows by symmetry.

(i) Write $S = \otimes P$, $T = \otimes Q$; we have

$$\mathrm{Hom}_B(P, B) \cong \mathrm{Hom}_B(A^S, B) \cong \mathrm{Hom}_A(A, B^T) \cong Q.$$

(ii) We have the bimodule isomorphisms

$$\mathrm{End}_B(P) \cong \mathrm{Hom}_B(A^S, P) \cong \mathrm{Hom}_A(A, P^T) \cong \mathrm{Hom}_A(A, A) \cong A.$$

Each term has a natural multiplication; all these correspond and give a ring isomorphism.

(iii) We have

$$\mathrm{Hom}_B(P, Y) \cong \mathrm{Hom}_A(A, Y^T) \cong Y^T,$$

and by hypothesis the functor $T: Y \mapsto Y^T$ is right exact. Hence $\mathrm{Hom}_B(P, -)$ is right exact and so $P_B$ is projective.

(iv) We have $P \oplus P' \cong {}^IB$, where ${}^IB$ stands for a direct sum of $|I|$ copies of

*B*. Hence

$$ {}^{I}Q \cong {}^{I}B \otimes Q \cong (P \otimes Q) \oplus (P' \otimes Q) \cong A \oplus (P' \otimes Q). $$

This shows that $Q_A$ is a generator, because $A$ is one.

(v) It is clear that the functor $S$ induces a map from $\mathrm{Lat}({}_AA)$ to $\mathrm{Lat}({}_BP)$ which is order-preserving and has a right inverse, induced by $T$; further, $A$-bimodules (= ideals) of $A$ correspond to $(A, B)$-bimodules in $P$. ■

Later, in Th. 5.4, we shall find that the modules $P_B$ and ${}_BQ$ are actually finitely generated. For the moment we note that Prop. 4.3 leads to a criterion for Morita equivalence:

THEOREM 4.5 *For any rings $A$, $B$ the following conditions are equivalent*:

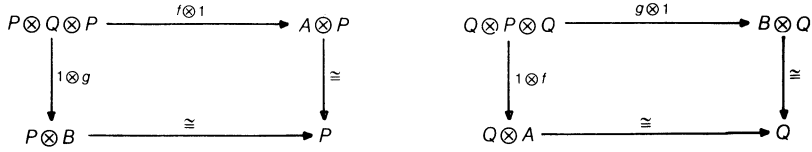
(a) $\mathscr{M}_A \cong \mathscr{M}_B$,

(a$^0$) ${}_A\mathscr{M} \cong {}_B\mathscr{M}$,

(b) *there exist bimodules ${}_AP_B$, ${}_BQ_A$ with bimodule isomorphisms*

$$ P \otimes_A Q \cong A, \qquad Q \otimes_B P \cong B. $$

*Moreover, when (b) holds and $f: P \otimes Q \to A$, $g: Q \otimes P \to B$ are bimodule isomorphisms, these may be chosen so as to make the following diagrams commutative*:




*Proof*. The equivalence of (a), (b) is clear by the proof of Prop. 4.3, and now the equivalence of (a$^0$), (b) follows by the evident symmetry of (b). Let us pick isomorphisms $f: P \otimes Q \to A$, $g: Q \otimes P \to B$; then all the arrows in the diagrams are isomorphisms. Consider the first diagram. If we take $p \in P$ and move it (anticlockwise) round the square we obtain $\theta p \in P$, where $\theta$ is an $(A, B)$-automorphisn of $P$. Now $\mathrm{End}_B(P) = \mathrm{End}_A(A) \cong A^0$, so $\theta$ is left multiplication by a unit $u$ in $A$; since $\theta$ is also an $A$-automorphism, $u$ lies in the centre of $A$. If we replace $f$ by $uf$, the first square becomes commutative. We complete the proof by showing that with this choice of $f, g$ the second square also commutes. For brevity write $f(p \otimes q) = (p, q)$, $g(q \otimes p) = [q, p]$; we have adjusted $f$ so that

$$ (p, q)p' = p[q, p'], \tag{4} $$

and we must show that

$$ [q, p]q' = q(p, q'). \tag{5} $$

Given $p, p' \in P$, $q, q' \in Q$, we have by the left and right $B$-linearity of $g$,

$$[[q,p]q',p'] = [q,p][q',p'] = [q,p[q',p']].$$

By (4) and the fact that $g$ is $A$-balanced, this is

$$[q,p[q',p']] = [q,(p,q')p'] = [q(p,q'),p'].$$

Thus if $c = [q,p]q' - q(p,q')$, then $c \in Q$ and $[c,p'] = 0$ for all $p' \in P$, Let us define $h: A \to Q$ by $h(a) = ca$; then $(h \otimes 1)g: A \otimes P \to Q \otimes P \to B$ and this map is zero because $[c,p'] = 0$. But $g$ is an isomorphism, so $h \otimes 1 = h^T = 0$ and $T$ is an equivalence, hence $h = 0$. Thus $c = h(1) = 0$, as we had to show. ■

To illustrate the result we take $B = A_n$ for some $n > 1$. Then we may choose $P = A^n$, $Q = {}^nA$ and it is clear that $P \otimes Q \cong A$, $Q \otimes P \cong A_n$.

As a first consequence we see how to sharpen Prop. 4.4 for Morita equivalent rings:

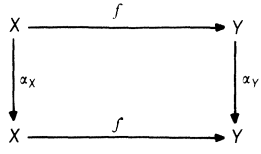
COROLLARY 4.6 *If $A$, $B$ are Morita equivalent rings, with bimodules $P, Q$ satisfying $P \otimes Q \cong A$, $Q \otimes P \cong B$, then* $\mathrm{Lat}(A_A) \cong \mathrm{Lat}(P_B)$, $\mathrm{Lat}({}_AA) \cong \mathrm{Lat}({}_BQ)$, $\mathrm{Lat}({}_BB) \cong \mathrm{Lat}({}_AP)$, $\mathrm{Lat}(B_B) \cong \mathrm{Lat}(Q_A)$. *Moreover,*

$$\mathrm{Lat}({}_AA_A) \cong \mathrm{Lat}({}_AP_B) \cong \mathrm{Lat}({}_BB_B),$$

*in other words: $A$ and $B$ have isomorphic ideal lattices.* ■

By a *Morita invariant* we understand a property of rings which is preserved by Morita equivalence. For example, being simple is a Morita invariant, by Cor. 4.6.

As another example of a Morita invariant we have the centre. It is convenient to define this notion in the wider context of abelian categories. In any abelian category $\mathscr{A}$ let $\mathrm{Nat}(I)$ be the set of all natural transformations from the identity functor to itself. Since functors and natural transformations themselves form a category $\mathscr{A}^{\mathscr{A}}$, it follows that the set $\mathrm{Nat}(I)$, as the set of all 'endomorphisms' of $I$ is a ring; it is called the *centre* of the category $\mathscr{A}$. Explicitly, $\alpha \in \mathrm{Nat}(I)$ means that for each $X \in Ob\,\mathscr{A}$ there is a map $\alpha_X: X \to X$ such that any map $f: X \to Y$ in $\mathscr{A}$ gives rise to a commutative square




In the particular case where $\mathscr{A} = \mathscr{M}_A$, $\mathrm{Nat}(I)$ consists of all $A$-endomorphisms which commute with all $A$-homomorphisms. Writing $C$ for the centre of the

ring $A$, we have for each $c \in C$ an element $\mu_c$ of $\mathrm{Nat}(I)$, defined as

$$x\mu_c = xc, \qquad \text{for } x \in X,\ X \in \mathscr{M}_A.$$

It is clear that the map $\mu: c \mapsto \mu_c$ defines a ring homomorphism

$$C \to \mathrm{Nat}(I).$$

We assert that this is an isomorphism: if $\mu_c = 0$, take $X = A$; then $0 = 1\mu_c = c$, so $c = 0$ and $\mu$ is injective. To prove surjectivity, let $f \in \mathrm{Nat}(I)$, say $1f_A = c$. By naturality, $ac = ca$, hence $c \in C$ and so $g = f - \mu_c \in \mathrm{Nat}(I)$. Now on $A$, $g = 0$ by definition, and given $x \in X$, where $X \in M_A$, we define $\varphi: A \to X$ by $1 \mapsto x$. Then commutativity shows that $g(x) = x \cdot g(1) = 0$, so $g = 0$ and $f = \mu_c$. The result may be stated as follows:

THEOREM 4.7 *The centre of the ring $A$ is isomorphic to the centre of the category $\mathscr{M}_A$. Hence Morita equivalent rings have isomorphic centres.* ■

This result shows for example that two commutative rings are equivalent iff they are isomorphic; more generally, a ring $R$ is equivalent to a commutative ring iff it is equivalent to its centre.

The property of being finitely generated can be expressed categorically: $M$ is finitely generated iff $M$ cannot be expressed as the union of a chain of proper submodules. This means that any module corresponding to a finitely generated module under a category equivalence is again finitely generated. However, the cardinal of a minimal generating set may well be different for the two modules, and this fact can be utilized to turn any problem on finitely generated modules into a problem on cyclic modules. In order to show this clearly we examine the equivalence $A \sim A_n$ in greater detail.

Fix $n \geqslant 1$ and write $P = A^n$, $Q = {}^nA$. We have the functors

$$M \mapsto M^n = M \otimes_A P \qquad (M \in \mathscr{M}_A), \tag{6}$$

and

$$N \mapsto N^\natural = N \otimes_{A_n} Q \qquad (N \in \mathscr{M}_{A_n}). \tag{7}$$

Here $N^\natural$ may also be defined as $\mathrm{Hom}_{A_n}(P, N)$. It is easily checked that

$$(M^n)^\natural \cong M, \qquad (N^\natural)^n \cong N,$$

and this provides an explicit form for the equivalence $A \sim A_n$. Given any finitely generated right $A$-module $M$, with generating set $u_1, \ldots, u_n$ say, we can apply (6) and pass to the right $A_n$-module $M^n$, which is generated by the single element $(u_1, \ldots, u_n)$. We state the result as

THEOREM 4.8 *For any ring $A$, any finitely generated $A$-module $M$ corresponds to a cyclic $A_n$-module under the category equivalence* (6)*, for suitable $n$. In fact it is enough to take $n$ equal to the cardinal of a generating set of $M$.* ■

For example, if $A$ is a principal ideal ring, any submodule of an $n$-generator module over $A$ can be generated by $n$ elements (as follows from Prop. 9.2.3 below). Applying Th. 4.8, we see that any submodule of a cyclic $A_n$-module is cyclic, so $A_n$ is again a principal ideal ring. In the opposite direction, if $A_n$ is a principal ideal ring, then any submodule of a cyclic module is cyclic. It follows that any submodule of an $n$-generator $A$-module can be generated by $n$ elements. This can happen for some $n > 1$ for certain rings which are not principal (cf. Webber (1970)).

### Exercises

(1) Show that a skew field $K$ is Morita equivalent only to $K_n$, $n = 1, 2, \ldots$.

(2) Verify directly that the centre of $A$ is $\mathrm{Hom}_{A-A}(A, A)$.

(3) Verify that for a ring to be Noetherian, or Artinian is a Morita invariant.

(4) Show that $A \prec B \Leftrightarrow A^0 \prec B^0$.

(5) Show that any non-trivial ring without IBN has a simple homomorphic image without IBN. Verify that a simple ring without IBN is Morita equivalent to only a finite number of rings (up to isomorphism).

## 6.5 The Morita context

We have seen in Th. 4.5 that a Morita equivalence between two rings $A$ and $B$ is determined by two bimodules $P$, $Q$, but in practice one wants to know for which pairs of rings $A$, $B$ it is true that $A \sim B$. The first step is to find conditions on $P$ and $Q$ for Th. 4.5 to apply. Here we need a general property of modules.

For any right $A$-module $M$ we define its *dual* as $M^* = \mathrm{Hom}_A(M, A)$; this is a left $A$-module in a natural way. Let us write the image of $x \in M$ under $f \in M^*$ as $(f, x)$ and put

$$\tau(M) = \tau_A(M) = \{\sum(f, x) | f \in M^*, x \in M\}.$$

This is a 2-sided ideal in $A$, called the *trace ideal* of $M$. For example, if $F$ is a non-zero free $A$-module, then $\tau(F) = A$. The modules for which $\tau = A$ are of particular interest; as we see from the next result, they are just the generators of the category ${}_A\mathscr{M}$.

LEMMA 5.1 *Let $A$ be any ring. For any left $A$-module $M$ the following are equivalent:*

(a) *$M$ is a generator,*

(b) $\tau_A(M) = A$,

(c) $M^n = A \oplus N$ *for some integer $n$ and some* ${}_AN$.

*Proof.* (a)⇒(b). Assume that $\tau(M) = \mathfrak{a} \neq A$; then the natural homomorphism $\pi: A \to A/\mathfrak{a}$ is non-zero, hence by (a) the induced map

$$M^* = \mathrm{Hom}_A(M, A) \to \mathrm{Hom}_A(M, A/\mathfrak{a}) \tag{1}$$

is non-zero. But every $f \in M^*$ maps $M$ into $\mathfrak{a}$ by assumption, and this means that (1) is zero, a contradiction. Hence $\tau(M) = A$ and (*b*) holds.

(b)⇒(c). By hypothesis $\tau(M) = A$, hence there exist $f_1, \ldots, f_n \in M^*$, $u_1, \ldots, u_n \in M$ such that $\sum(f_i, u_i) = 1$. We define a homomorphism $\varphi: M^n \to A$ by the rule $(x_1, \ldots, x_n) \mapsto \sum(f_i, x_i)$. Its image in $A$ is a left ideal containing $1 = \sum(f_i, u_i)$, hence $\varphi$ is surjective and if $\ker \varphi = N$, we have the exact sequence

$$0 \to N \to M^n \to A \to 0.$$

Since $A$ is projective, this sequence splits and (c) follows.

(c)⇒(a). Given a map $f: X \to Y$ of $A$-modules, if the induced map $\mathrm{Hom}(M, X) \to \mathrm{Hom}(M, Y)$ is zero, then this also holds for the map $\mathrm{Hom}(M^n, X) \to \mathrm{Hom}(M^n, Y)$, i.e. $\mathrm{Hom}(A \oplus N, X) \to \mathrm{Hom}(A \oplus N, Y)$. But the restriction to the first summand is just the original map $f: X \to Y$ (because $\mathrm{Hom}(A, X) \cong X$), so $f = 0$. This shows $h^M$ to be faithful, so (a) holds. ■

We now consider the following situation: Given two rings $A, B$ and bimodules ${}_AP_B, {}_BQ_A$, assume that we have two bimodule homomorphisms:

$$\tau: P \otimes Q \to A, \qquad f, x \mapsto (f, x), \tag{2}$$

$$\mu: Q \otimes P \to B, \qquad x, f \mapsto [x, f], \tag{3}$$

such that

$$[x, f]y = x(f, y), \qquad f, g \in P, x, y \in Q, \tag{4}$$

$$g[x, f] = (g, x)f. \tag{5}$$

These rules may be expressed symbolically by saying that $\begin{pmatrix} A & P \\ Q & B \end{pmatrix}$ is a ring under the usual matrix multiplication. This sums up the module laws and (2), (3), while (4), (5) are instances of the associative law. The 6-tuple $(A, B, P, Q, \tau, \mu)$ is called a *Morita context*. We remark that $\mathrm{im}\,\tau$ is an ideal in $A$ and $\mathrm{im}\,\mu$ is an ideal in $B$.

Starting from any module $E_A$ we obtain a Morita context as follows. We put

$$E^* = \mathrm{Hom}_A(E, A), \qquad B = \mathrm{End}_A(E),$$

and regard $E$ as a $(B, A)$-bimodule and $E^*$ as an $(A, B)$-bimodule in the natural way. Further we have a natural map $\tau: E^* \otimes E \to A$ given by evaluation, as in (2). To find $\mu$, we use (4): for any $x \in E$, $f \in E^*$ we define $[x, f]$ by its effect on $E$:

$$[x, f]y = x(f, y) \qquad (y \in E).$$

This is an $A$-balanced biadditive map of $E \times E^*$ into $B$ and so defines a homomorphism $\mu: E \otimes E^* \to B$ which is easily verified to be a $B$-bimodule homomorphism. Further, (4) holds by the definition of $\mu$ and the definition of $E^*$ as $B$-module shows that for all $y \in E$, $(g[x, f], y) = (g, [x, f]y) = (g, x(f, y)) = (g, x)(f, y) = ((g, x)f, y)$; hence $g[x, f] = (g, x)f$, and this proves (5).

We thus have a Morita context $(A, B, E^*, E, \tau, \mu)$ starting from $E_A$; this is called the Morita context *derived from* $E_A$. To give an example, if $A$ is a simple Artinian ring, say $A \cong K_n$, where $K$ is a skew field (by Wedderburn's theorem, Th. 5.3.2 of A.2) and $E = K^n$ is a simple right $A$-module, then $E^* = {}^nK$ and the derived Morita context is of the form $(K_n, K, {}^nK, K^n, \tau, \mu)$.

For general Morita contexts we shall be interested in the case where $\tau, \mu$ are isomorphisms. In that case we have a Morita equivalence between $A$ and $B$, by Th. 4.5, with functors

$$S: \mathcal{M}_A \to \mathcal{M}_B, \qquad M \mapsto M \otimes_A P,$$
$$T: \mathcal{M}_B \to \mathcal{M}_A, \qquad N \mapsto N \otimes_B Q,$$

which are mutually inverse, because $P \otimes Q \cong B$, $Q \otimes P \cong A$. Under $S, A$ corresponds to $P$ and right ideals of $A$ correspond to $B$-submodules of $P$, with 2-sided ideals corresponding to $(A, B)$-submodules. Similarly, $Q$ corresponds to $B$ and under $T, A$ corresponds to $Q$ and $P$ to $B$ (cf. Prop. 4.4). Moreover, we have the isomorphisms

$$Q \cong \mathrm{Hom}_A(P, A) \cong \mathrm{Hom}_B(P, B), \qquad P \cong \mathrm{Hom}_A(Q, A) \cong \mathrm{Hom}_B(Q, B),$$
$$A \cong \mathrm{End}_B(Q) \cong \mathrm{End}_B(P)^0, \qquad B \cong \mathrm{End}_A(P) \cong \mathrm{End}_A(Q)^0.$$

The following lemma shows that to verify that $\mu$ is an isomorphism it is enough to check surjectivity.

LEMMA 5.2 *In any Morita context $(A, B, P, Q, \tau, \mu)$, if $\mu$ is surjective, then it is an isomorphism; similarly for $\tau$.*

*Proof.* Since $\mu$ is surjective, we have

$$\sum [x_i, f_i] = 1 \qquad \text{for some } x_i \in Q, f_i \in P.$$

Now assume that $\sum [y_\lambda, g_\lambda] = 0$; then

$$\sum y_\lambda \otimes g_\lambda = \sum y_\lambda \otimes g_\lambda [x_i, f_i] = \sum y_\lambda \otimes (g_\lambda, x_i) f_i = \sum y_\lambda (g_\lambda, x_i) \otimes f_i$$
$$= \sum [y_\lambda, g_\lambda] x_i \otimes f_i = 0.$$

This shows $\mu$ to be injective, and hence an isomorphism. ■

In the special case of a derived Morita context we can give an explicit criterion for $\mu$ to be surjective. We recall the dual basis lemma from A.2 (Cor. 4.5.6). In the finitely generated case this states that ${}_AP$ is a direct summand of $A^n$ iff there

exist $u_1, \ldots, u_n \in P$, $f_1, \ldots, f_n \in P^*$ (the 'projective coordinate system') such that

$$x = \sum (f_i, x) u_i \qquad (x \in P). \tag{6}$$

Similarly for a right module this equation takes the form $x = \sum u_i (f_i, x)$.

LEMMA 5.3 *Given any module $Q_A$, let $(A, B, P, Q, \tau, \mu)$ be the derived Morita context. Then $\mu: Q \otimes P \to B$ is an isomorphism if and only if $Q_A$ is finitely generated projective.*

*Proof.* By the dual basis lemma just quoted, $Q_A$ is finitely generated projective iff there is a finite projective coordinate system:

$$x = \sum u_i (f_i, x) \qquad (x \in Q).$$

Bearing in mind that $Q^* = P$, we can by (4) write this as $x = \sum [u_i, f_i] x$ for all $x \in Q$, i.e. $\sum [u_i, f_i] = 1$. But this is just the condition for $\mu$ to be surjective, and by Lemma 5.2, for $\mu$ to be an isomorphism. ■

We can now state a condition on any module $Q_A$ for its derived Morita context to define an equivalence:

THEOREM 5.4 *Let $A$ be a ring, $Q_A$ any module and $(A, B, P, Q, \tau, \mu)$ the Morita context derived from $Q$. Then this context defines a Morita equivalence between $A$ and $B$ if and only if $Q_A$ is a finitely generated projective generator.*

A finitely generated projective generator is also called a *progenerator*.

*Proof.* If we have a Morita equivalence, then $Q_A$ corresponds to $B_B$ in the equivalence, so $Q$ is a finitely generated projective generator because $B$ is. Conversely, assume this condition for $B$. By Lemma 5.3 the map $\mu: Q \otimes P \to B$ is an isomorphism. Since $Q$ is a generator, the trace ideal $\operatorname{im} \tau$ is $A$ (Lemma 5.1), so $\tau$ is surjective and hence an isomorphism by Lemma 5.2. Thus $P \otimes Q \cong A$, and now Th. 4.5 shows that we have indeed a Morita equivalence. ■

This result shows that every Morita equivalence can be obtained from a particular Morita context. Given $A \sim B$, to find $Q$ we need a finitely generated projective $A$-module; this is a direct summand of $A^n$ for some $n \geqslant 1$, and it may be specified by an idempotent $e$ in $A_n \cong \operatorname{End}_A(A^n)$. Suppose first that $n = 1$; this means that $Q_A = eA$, where $e$ is an idempotent in $A$. By Lemma 3.8 (or rather, its left–right dual) we have $P = \operatorname{Hom}_A(eA, A) = Ae$, $B = \operatorname{End}_A(eA) = eAe$. Now the condition for $Q$ to be a generator reads: the natural map $P \otimes Q \to A$ is surjective, i.e. $AeA = A$. The translation to $A_n$ is now clear. We choose $n \geqslant 1$ and take an idempotent $e$ in $A_n$ such that $A_n e A_n = A_n$. Then we have $B = eA_n e$, and all rings Morita equivalent to $A$ are obtained in this way, with the appropriate Morita context $(A, eA_n e, A^n e, e^n A, \tau, \mu)$.

An important particular case is obtained by starting from a commutative

ring $K$ say. If $Q$ is any finitely generated projective $K$-module, then $A = \mathrm{End}_K(Q)$ is a $K$-algebra and for any $K$-algebra $R$ we have

$$R \sim \mathrm{End}_R(R \otimes Q) \cong R \otimes_K A,$$

as is easily checked. Thus $A \sim K$ and tensoring with $A$ (over $K$) converts any $K$-algebra into one Morita equivalent to it. Such algebras are called *Brauer equivalent*; in the special case where $K$ is a field, we have $A = K_n$ and $R \otimes K_n \cong R_n$, while the general case leads to the study of Azumaya algebras (Azumaya (1950), Auslander and Goldman (1960), cf. also **10.7** below).

We conclude this section by describing another important Morita invariant, the trace group. Let $K$ be any commutative ring and consider a $K$-algebra $A$. With $A$ we associate a $K$-module, its *trace group*, defined as

$$\mathbf{T}(A) = A/C, \qquad \text{where} \qquad C = \{\sum(xy - yx) | x, y \in A\}. \tag{7}$$

The natural map $A \to \mathbf{T}(A)$ is called the *trace function* and is written $\mathrm{tr}(x)$. Clearly it has the following properties:

**T.1** $\mathrm{tr}: A \to \mathbf{T}(A)$ *is* $K$*-linear,*

**T.2** $\mathrm{tr}(xy) = \mathrm{tr}(yx)$ *for all* $x, y \in A$.

Moreover, any linear map $\alpha: A \to M$ into a $K$-module such that $\alpha(xy) = \alpha(yx)$ can be written as $\alpha(x) = \alpha'(\mathrm{tr}(x))$ for a unique $\alpha': \mathbf{T}(A) \to M$. Thus tr is the universal mapping satisfying **T.1–2**. Let us show that **T** is a Morita invariant:

PROPOSITION 5.5 *For any rings* $A, B$, *if* $A \sim B$, *then* $\mathbf{T}(A) \cong \mathbf{T}(B)$.

*Proof.* Let $(A, B, P, Q, (\ ), [\ ])$ be the Morita context providing the equivalence and consider the map $A \to \mathbf{T}(B)$ given by

$$\alpha: \sum(f_i, x_i) \mapsto \mathrm{tr}(\sum[x_i, f_i]), \qquad \text{where } f_i \in P, x_i \in Q.$$

To show that this is well defined we must check that the map $f, x \mapsto \mathrm{tr}([x, f])$ is bilinear and $B$-balanced. The bilinearity is clear, and we have for $b \in B$,

$$\mathrm{tr}([x, fb]) = \mathrm{tr}([x, f]b) = \mathrm{tr}(b[x, f]) = \mathrm{tr}([bx, f]),$$

hence the result. Moreover, $\alpha(aa') = \alpha(a'a)$, because $\alpha(\sum(f_i, x_i)a) = \mathrm{tr}(\sum[x_i a, f_i]) = \mathrm{tr}(\sum[x_i, af_i]) = \alpha(\sum a(f_i, x_i))$; hence $\alpha$ induces a map $\alpha': \mathbf{T}(A) \to \mathbf{T}(B)$. By symmetry there is a map $\beta': \mathbf{T}(B) \to \mathbf{T}(A)$ and these two maps are easily seen to be mutually inverse. ■

As an illustration, if $A$ is an algebra with centre $K$, then of the two Morita invariants $\mathbf{C}(A)$, $\mathbf{T}(A)$ the centre is just $K$, whereas $\mathbf{T}(A)$ is a $K$-module which in general is larger than $K$ and so will tell us more about $A$. We remark that in (7) $C$ is merely a $K$-module and not an ideal; this means that $\mathbf{T}(A)$ will often be non-zero even if $A$ is simple. However, for some rings $\mathbf{T}(A) = 0$ (cf. **9.3**,

Ex. (10)). The trace function introduced here was first defined by Hattori and independently by Stallings in 1965.

The notion of Morita equivalence was introduced by K. Morita in 1958. The above account follows the author's notes (Cohn [1966]), which in turn were based on the notes of H. Bass [1962].

### Exercises

(1) Given a Morita context $(A, B, P, Q, \tau, \mu)$, if $P \otimes Q \cong A$ and $A \cong \mathrm{End}_B(Q)$, show that ${}_BQ$ is finitely generated projective.

(2) Let $K$ be a commutative ring and $P$ a finitely generated faithful projective module ($P$ is *faithful* if $Pa = 0 \Rightarrow a = 0$). Show by using the dual basis lemma that $\tau(P) = K$ and so $P$ is a generator.

(3) If $(A, B, P, Q, \tau, \mu)$ is a Morita context defining a Morita equivalence and $R = \begin{pmatrix} A & P \\ Q & B \end{pmatrix}$, show that $A \prec R, B \prec R$.

(4) Let $K$ be any ring, $E, F$ any $K$-modules and define $P = \mathrm{Hom}(E, F)$, $Q = \mathrm{Hom}(F, E)$, $A = \mathrm{End}(E)$, $B = \mathrm{End}(F)$. Show that together with the natural maps $P \otimes Q \to A$, $Q \otimes P \to B$ this defines a Morita context.

(5) Show that the Morita context of Ex. (4) defines a Morita equivalence between $A$ and $B$ iff there exist integers $m, n \geqslant 1$ and $K$-modules $E', F'$ such that $E \oplus E' \cong F^n$, $F \oplus F' \cong E^m$ (K. Hirata).

## 6.6 Projective, injective and flat modules

We now take a closer look at flat modules and their relation to projective and injective modules.

Let $R$ be any ring. Any left $R$-module $M$ has a presentation

$$G \xrightarrow{\alpha} F \xrightarrow{\beta} M \to 0, \tag{1}$$

where $F, G$ are free $R$-modules. Explicitly, let $F$ have the basis $(f_\lambda)$ and $G$ the basis $(g_i)$; then $\alpha$ is described by a matrix $A = (a_{i\lambda})$, called the *presentation matrix* of $M$, where

$$g_i\alpha = \sum a_{i\lambda} f_\lambda. \tag{2}$$

In general $F, G$ will not be finitely generated and $A$ may have infinitely many rows and columns, but each row has only finitely many non-zero entries; we say that $A$ is *row-finite*. We note that $M$ is determined up to isomorphism by $A$ as the cokernel of the corresponding map $\alpha$, by (1). Moreover, every row-finite matrix $A$ defines a module in this way. If $F, G$ have finite ranks $m, n$ respectively, then the presentation matrix is $m \times n$.

It is clear that $F$ can be taken to be of finite rank iff $M$ is finitely generated. If $G$ can be taken to be of finite rank, $M$ is said to be *finitely related* and the module $M$ is called *finitely presented* if $F, G$ can both be taken to be of finite rank. Given two presentations

$$0 \to K_i \to F_i \to M \to 0 \qquad (i = 1, 2), \tag{3}$$

where $F_1, F_2$ are free, we have by Schanuel's lemma (Lemma 3.4.2), $F_1 \oplus K_2 \cong F_2 \oplus K_1$, hence if $M$ has a presentation with $F_1$ of finite rank and another with $K_2$ finitely generated, then $F_2, K_1$ are also finitely generated and $M$ is finitely presented.

Let us examine the presentation matrix of a projective module.

PROPOSITION 6.1 *Let $M$ be an $R$-module with the presentation* (1). *Then $M$ is projective if and only if there exists $\alpha' : F \to G$ such that $\alpha\alpha'\alpha = \alpha$. Thus a matrix $A$ represents a projective module if and only if there is a matrix $A'$ such that $AA'A = A$.*

*Proof.* Assume that $M$ is projective; then $F \cong M \oplus \ker\beta \cong M \oplus \operatorname{im}\alpha$, so $\operatorname{im}\alpha$ is also projective. Therefore the projection $F \to \operatorname{im}\alpha$ can be lifted to a map $\alpha' : F \to G$, whose composition with $\alpha$ is the projection onto $\operatorname{im}\alpha$, hence $\alpha\alpha'\alpha = \alpha$.

Conversely, if $\alpha'$ satisfies $\alpha\alpha'\alpha = \alpha$, then $\alpha\alpha'$ is an idempotent endomorphism of $G$, hence $G = G_1 \oplus G_2$, where $G_1 = \operatorname{im}\alpha\alpha'$, $G_2 = \ker\alpha\alpha'$, and writing $\alpha_1 = \alpha | G_1$, we have the exact sequence (1), with $G, \alpha$ replaced by $G_1, \alpha_1$. Since $\alpha_1\alpha' = 1$, the sequence splits and so $M$ is projective. The final assertion follows by rewriting the result in terms of matrices. ■

Next we give some conditions for a module to be flat.

THEOREM 6.2 *Let $R$ be a ring. For any right $R$-module $U$ the following conditions are equivalent:*

(a) *$U$ is flat,*

(b) $\operatorname{Tor}_1^R(U, -) = 0$,

(c) *for any free left $R$-module $F$ and any submodule $G$ of $F$, the induced map $U \otimes G \to U \otimes F$ is injective,*

(d) *given $uc = 0$, where $u \in U^n, c \in {}^nR$, there exists $A \in {}^mR^n$ and $v \in U^m$ such that $u = vA$ and $Ac = 0$,*

(e) *for any finitely generated left ideal $\mathfrak{a}$ of $R$ the map $U \otimes \mathfrak{a} \to U$ induced by the inclusion $\mathfrak{a} \subseteq R$ is injective, so that $U \otimes \mathfrak{a} \cong U\mathfrak{a}$.*

Condition (d) may be expressed loosely by saying that any relation in $U$ is a consequence of relations in $R$; this explains the name 'flat', if one thinks of relations in a module as a kind of generalized torsion.

*Proof.* It is clear from the definition that (a) and (b) are equivalent. When $U$ is flat, the induced sequence

$$0 \to U \otimes G \to U \otimes F$$

is injective, so (a)$\Rightarrow$(c). To show that (c)$\Rightarrow$(d), consider the exact sequence

$$0 \to K \xrightarrow{\alpha} R^n \xrightarrow{\beta} R,$$

where $\beta:(x_i) \mapsto \sum x_i c_i$ and $K = \ker \beta$. By (c) the induced sequence

$$0 \to U \otimes K \to U^n \to U$$

is exact; since $\sum u_i c_i = 0$, we have $u_i = \sum v_h a_{hi}$ for some $v_h \in U$, $a_{hi} \in K$, and so $\sum a_{hi} c_i = 0$, which proves (d).

(d)$\Rightarrow$(e). Let $\mathfrak{a} = Rc_1 + \cdots + Rc_n$; then every element of $U \otimes \mathfrak{a}$ has the form $\sum u_i \otimes c_i$ for some $u_i \in U$. If $\sum u_i c_i = 0$, then in vector notation, $u = vA$, where $Ac = 0$; hence in $U \otimes \mathfrak{a}$ we have $u \otimes c = vA \otimes c = v \otimes Ac = 0$, and this shows the mapping $U \otimes \mathfrak{a} \to U$ to be injective.

(e)$\Rightarrow$(a). By (e) we have, for any finitely generated left ideal $\mathfrak{a}$, an exact sequence

$$0 \to U \otimes \mathfrak{a} \to U \to U \otimes (R/\mathfrak{a});$$

hence $\operatorname{Tor}_1^R(U, C) = 0$ for any finitely related cyclic $R$-module $C$. By induction, using the exact homology sequence, we find that $\operatorname{Tor}_1^R(U, A) = 0$ for any finitely presented left $R$-module $A$. Consider now the left $R$-module $M$ and take a resolution

$$0 \to K \to F \to M \to 0,$$

where $F$ is free. Tensoring up, we find a sequence

$$0 \to U \otimes K \to U \otimes F \to U \otimes M \to 0, \tag{4}$$

which is exact, except possibly at $U \otimes K$, and we must show that it is exact. Take $\sum u_i \otimes k_i \in U \otimes K$ mapping to zero. This element will still map to zero if in (1) we discard from $F$ the free generators not occurring in the $k_i$ and replace $K$ by the submodule generated by the $k_i$. We now have a resolution for a finitely presented module, which has been proved to be exact. Hence $\sum u_i \otimes k_i = 0$, and it follows that the original sequence was exact. Hence $U$ is flat, and (a) holds. ■

There is another characterization of flat modules, which clarifies their relation to projective modules.

PROPOSITION 6.3 *Let $R$ be any ring and $U$ an $R$-module with a presentation*

$$0 \to K \xrightarrow{\alpha} F \xrightarrow{\beta} U \to 0, \tag{5}$$

*where $F$ is free. Then the following conditions are equivalent:*

(a) *$U$ is flat,*

(b) *given $x_1, \ldots, x_n \in K$, there exists $\gamma: F \to K$ such that $x_i \alpha\gamma = x_i$ for $i = 1, \ldots, n$,*

(c) *given $x \in K$, there exists $\gamma: F \to K$ such that $x\alpha\gamma = x$.*

*Proof.* (a)$\Rightarrow$(b). We first prove (c). Let $(f_\lambda)$ be a basis of $F$; by suitably renumbering the $f$'s we may write $x = f_1 c_1 + \cdots + f_r c_r$. Indicating the action of $\beta$ by a bar, we have $\bar{x} = \sum \bar{f}_i c_i = 0$, hence by Th. 6.2 (d) there exist $g_1, \ldots, g_m \in F$ and $a_{hi} \in R$ such that $\bar{f}_i = \sum \bar{g}_h a_{hi}$, $\sum a_{hi} c_i = 0$. Hence $f'_i = f_i - \sum g_h a_{hi} \in K$ and $x = \sum f_i c_i = \sum (f'_i + \sum g_h a_{hi}) c_i = \sum f'_i c_i$. Now define $\gamma: F \to K$ by

$$f_\lambda \gamma = \begin{cases} f'_\lambda & \text{for } \lambda = 1, \ldots, r, \\ 0 & \text{otherwise.} \end{cases}$$

Then $x\alpha\gamma = (\sum f_i c_i)\alpha\gamma = \sum f'_i c_i = x$, as required.

If we now apply this argument to the exact sequence of right $R_n$-modules

$$0 \to K^n \to F^n \to U^n \to 0,$$

and observe that $U^n$ is flat whenever $U$ is, we obtain (b).

(b)$\Rightarrow$(c) holds trivially, and to prove (c)$\Rightarrow$(a) we shall verify (d) of Th. 6.2. Let $(f_\lambda)$ be a basis of $F$ and consider a relation in $U$, which by suitable numbering of the $f$'s may be written $\sum \bar{f}_i c_i = 0$. Then $x = \sum f_i c_i \in K$ and by hypothesis there exists $\gamma: F \to K$ such that $x\alpha\gamma = x$. Write $f_i \gamma = f_i + \sum f_h a_{hi}$; then since $x\alpha\gamma = x$, we have $\sum f_i c_i = \sum (f_i \gamma) c_i = \sum f_i c_i + \sum f_h a_{hi} c_i$, and so $\sum a_{hi} c_i = 0$, because the $f_\lambda$ are free. Moreover, since $f_\lambda \gamma \in K$, we have $\bar{f}_i = \sum \bar{f}_h a_{hi}$, so condition (d) of Th. 6.2 is satisfied, therefore $U$ is flat. ■

If $U$ is finitely related, we can take $K$ in (5) to be finitely generated, and by taking a generating set of $K$ we obtain a mapping $\gamma: F \to K$ which splits (5). Hence $U$ is then projective; since any projective module is flat, this proves

COROLLARY 6.4 *A finitely related module (over any ring) is flat if and only if it is projective.* ■

This shows that for example, over a Noetherian ring, every finitely generated flat module is projective. We can now also characterize rings over which every (left or right) module is flat:

THEOREM 6.5 *For any ring $R$ the following are equivalent:*

(a) *every right $R$-module is flat,*

($a^0$) *every left $R$-module is flat,*

(b) *given $a \in R$, there exists $a' \in R$ such that $aa'a = a$.*

*Proof.* We shall prove the implications $(a^0) \Rightarrow (b) \Rightarrow (a)$; the theorem then follows by symmetry.

Let us apply Prop. 6.3 to the exact sequence of left $R$-modules

$$0 \to Ra \to R \to R/Ra \to 0. \tag{6}$$

By hypothesis $R/Ra$ is flat, so there exists a mapping $\gamma: R \to Ra$ such that $a\gamma = a$. Let $1\gamma = a'a$; then $a = a\gamma = (a \cdot 1)\gamma = a \cdot a'a$ and (b) is proved. Conversely when (b) holds, the argument just given shows that (6) satisfies the condition (c) of Prop. 6.3; more generally, this holds for any left ideal $\mathfrak{a}$ in place of $Ra$; thus if $u \in \mathfrak{a}$ and $uu'u = u$, we can define $\gamma: R \to \mathfrak{a}$ by $x \mapsto xu'u$. Hence $R/\mathfrak{a}$ is flat and so $\mathrm{Tor}_1^R(U, R/\mathfrak{a}) = 0$ for any right $R$-module $U$, so the mapping $U \otimes \mathfrak{a} \to U$ induced by the inclusion $\mathfrak{a} \subseteq R$ is injective. By Th. 6.2(c) it follows that $U$ is flat and so (a) holds. ■

A ring satisfying condition (b) of this theorem is called (*von Neumann*) *regular* or sometimes *absolutely flat*. Clearly every semisimple ring is regular. However, the converse does not hold; for example, if $K$ is any field and $I$ a set, then the direct power $K^I$ is a regular ring, but it is not semisimple, unless $I$ is finite.

A link between flat and injective modules is provided by

PROPOSITION 6.6 *Let $R$ be any ring. A left $R$-module $V$ is flat if and only if $\hat{V} = \mathrm{Hom}_{\mathbf{Z}}(V, \mathbf{K})$ is injective, as right $R$-module, where $\mathbf{K} = \mathbf{Q}/\mathbf{Z}$.*

*Proof.* For any right $R$-module $U$ we have, by adjoint associativity, the natural isomorphism

$$\mathrm{Hom}_R(U, \mathrm{Hom}_{\mathbf{Z}}(V, \mathbf{K})) \cong \mathrm{Hom}_{\mathbf{Z}}(U \otimes V, \mathbf{K}). \tag{7}$$

Now assume that $V$ is flat; then $- \otimes V$ is an exact functor, hence the right-hand side of (7) is exact as a functor of $U$, hence so is the left, and this means that $\hat{V}$ is injective (by definition). Conversely, when $\hat{V}$ is injective, the two sides of (7) are exact as functors of $U$. But the functor $\mathrm{Hom}_{\mathbf{Z}}(-, \mathbf{K})$ is faithful, and so preserves inexact sequences, by Prop. 3.2.6, hence $U \otimes V$ must be exact in $U$, so $V$ is flat, as claimed. ■

For example, $\hat{R}$ is always injective; this module has another property that is sometimes useful. We recall that a *cogenerator* $U$ is defined dually to the term 'generator' by the condition that $h_U$ is faithful; explicitly, for $U \in {}_R\mathscr{M}$, this means that for any ${}_RM$ and $0 \neq x \in M$ there is a homomorphism $\varphi: M \to U$ such that $x\varphi \neq 0$. For example, $\mathbf{K} = \mathbf{Q}/\mathbf{Z}$ is a cogenerator for $\mathbf{Z}$. For, given any abelian group $A$ and $0 \neq x \in A$, let $p$ be a prime factor of the order of $x$ (or any prime if $x$ has infinite order). Then $x\mathbf{Z}/px\mathbf{Z}$ is of order $p$ and can be embedded in $\mathbf{K}$; since $\mathbf{K}$ is injective (i.e. divisible in this case, cf. **3.3**), this embedding can be extended to a homomorphism of $A/px\mathbf{Z}$ into $\mathbf{K}$; combined with the natural

mapping $A \to A/pxZ$ this gives a homomorphism $A \to \mathbf{K}$ which does not kill $x$. In the general case we obtain a cogenerator by forming a coinduced extension, as follows.

THEOREM 6.7 *Let $R$ be any ring and $f: \mathbf{Z} \to R$ the canonical map. Then $\hat{R} = \mathrm{Hom}_{\mathbf{Z}}(R, \mathbf{K})$ is an injective cogenerator.*

*Proof.* For any right $R$-module $M$ we have, by adjoint associativity,

$$\mathrm{Hom}_R(M, \hat{R}) \cong \mathrm{Hom}_{\mathbf{Z}}(M \otimes R, \mathbf{K}) = \mathrm{Hom}_{\mathbf{Z}}(M, \mathbf{K}),$$

and this is non-zero whenever $M \neq 0$, because $\mathbf{K}$ is a cogenerator for $\mathbf{Z}$. It follows that there is a non-zero homomorphism from $M$ to $\hat{R}$. Hence for any $x \in M$, $x \neq 0$, we have a non-zero homomorphism from $xR$ to $\hat{R}$; since $\hat{R}$ is injective, by Prop. 6.6, this can be extended to a homomorphism from $M$ to $\hat{R}$. This shows $\hat{R}$ to be a cogenerator for $R$, and by Prop. 6.6, it is injective. ■

From the definition it is clear that the class of projective modules admits direct sums, while that of injective modules admits direct products. In the Noetherian case one can say a little more.

THEOREM 6.8 *Let $R$ be a ring. Then the direct sum of any family of injective left $R$-modules is injective if and only if $R$ is left Noetherian.*

*Proof.* Suppose that $R$ is left Noetherian, let $\{E_\lambda\}$ be any family of injective left $R$-modules and put $E = \oplus_I E_\lambda$. We have to show that any homomorphism $f: \mathfrak{a} \to E$ from a left ideal $\mathfrak{a}$ of $R$ into $E$ extends to a homomorphism of $R$ into $E$ (by Th. 3.3.4). Since $R$ is left Noetherian, $\mathfrak{a}$ is finitely generated, by $u_1, \ldots, u_r$ say and the images $u_1 f, \ldots, u_r f$ have only finitely many non-zero components and so lie in a submodule $E' = \oplus_{I_0} E_\lambda$, where $I_0$ is a finite subset of the index set $I$. Thus $f$ maps $\mathfrak{a}$ into $E'$; as a finite direct sum of injective modules $E'$ is again injective, so $f$ extends to a homomorphism of $R$ into $E'$ and combining this with the inclusion $E' \subseteq E$ we obtain the desired extension of $f$.

Conversely, assume that any direct sum of injective modules is injective and consider an ascending chain of left ideals in $R$:

$$\mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots \tag{8}$$

Write $\mathfrak{a} = \bigcup \mathfrak{a}_i$, let $I_n$ be the injective hull of $R/\mathfrak{a}_n$, put $I = \oplus I_n$ and define $f: \mathfrak{a} \to I$ as $xf = \sum x f_n$, where $f_n: \mathfrak{a} \to I_n$ is the homomorphism induced by the natural mapping $\mathfrak{a} \to \mathfrak{a}/\mathfrak{a}_n$. Since $x \in \mathfrak{a}$, we have $x \in \mathfrak{a}_r$ for some $r = r(x)$ and hence $x f_n = 0$ for all $n \geqslant r$; hence only finitely many of the $x f_n$ are non-zero and $f$ is well defined. By hypothesis $I$ is injective, so there is a homomorphism $f': R \to I$ extending $f$. Let $1f' = c \in I$; then $c \in I_1 + \cdots + I_s$ for some $s$ and so $f'$, followed by the projection on $I_n$ is zero for $n > s$. Hence the same is true of $f$, so any $x$

in $\mathfrak{a}$ must lie in $\mathfrak{a}_s$, i.e. $\mathfrak{a} = \mathfrak{a}_s$. Thus (8) breaks off and this shows $R$ to be left Noetherian. ■

The corresponding problems for projective and flat modules have been solved by Chase (1960). A ring $R$ is said to be *left coherent* if every finitely generated left ideal in $R$ is finitely related. Now Chase proves (i) the direct product of any family of flat left $R$-modules is flat iff $R$ is left coherent, and (ii) the direct product of any family of projective left $R$-modules is projective iff $R$ is right perfect and left coherent.

For commutative Noetherian rings we can describe the indecomposable injective modules in terms of the prime ideals of the ring. We recall from A.2, **9.8** that a prime ideal $\mathfrak{p}$ of a commutative ring $R$ is meet-irreducible (this is easily verified directly), hence $E(R/\mathfrak{p})$, the injective hull of $R/\mathfrak{p}$, is indecomposable. Further we recall that a maximal annihilator of a module $M$ is a prime ideal; the set of prime ideals which form annihilators of elements of $M$ is just $\mathrm{Ass}(M)$, and this consists of a single element $\mathfrak{p}$ precisely when 0 is primary in $M$.

THEOREM 6.9 *Let $R$ be a commutative Noetherian ring. Then* (i) *for any prime ideal $\mathfrak{p}$ of $R$, $E(R/\mathfrak{p})$ is an indecomposable injective module,* (ii) *any indecomposable injective module $E$ is isomorphic to $E(R/\mathfrak{p})$, for a unique prime ideal $\mathfrak{p}$.*

*Proof.* Let $\mathfrak{p}$ be a prime ideal in $R$. Then $\mathfrak{p}$ is meet-irreducible in $R$, hence $E(R/\mathfrak{p})$ is indecomposable injective. If $E$ is indecomposable injective, then 0 is meet-irreducible in any submodule of $E$, hence 0 is primary and so $\mathrm{Ass}(E)$ consists of a single element $\mathfrak{p}$. Clearly $R/\mathfrak{p}$ is embedded in $E$, hence $E(R/\mathfrak{p}) \cong E$, and $\mathfrak{p}$ is clearly unique. ■

Although this result has been extended for certain non-commutative rings, there is no corresponding description in the general case. However, there is a connexion with divisible modules which is quite general and which we shall now explain.

We have seen in A.2, **4.5** that over an integral domain every injective module is divisible. Let us examine more closely how the hypothesis that $R$ is a domain was used. Given $u \in M$, $0 \neq a \in R$, we need to be sure that the mapping $ar \mapsto ur$ from $aR$ to $M$ is well defined, and this will be the case if

$$ax = 0 \Rightarrow ux = 0 \qquad \text{for all } x \in R. \tag{9}$$

Let $R$ be any ring and $M$ a right $R$-module such that for any $u \in M$ and $a \in R$, the equation $u = va$ has a solution $v$ whenever (9) holds. In that case $M$ is said to be 1-*divisible*. Over an integral domain this reduces to the notion of a divisible module and the proof of Prop. 4.5.8 of A.2 can be adapted to obtain

PROPOSITION 6.10 *Over any ring $R$, an injective module is 1-divisible; over a principal ideal domain the converse holds too.* ■

In general, being 1-divisible is of course not sufficient for injectivity, but a necessary and sufficient condition is now easily obtained. We recall that for any index set $I$, $M^I$ denotes the direct product of $I$ copies of $M$, while ${}^IM$ denotes the direct sum of $I$ copies. We visualize the elements of $M^I$ and ${}^IM$ as rows and columns respectively; thus for any $u \in M^I$, $x \in {}^IR$ we can form $ux = \sum u_i x_i$, because almost all components of $x$ are zero.

A right $R$-module $M$ is called *fully divisible* if it satisfies the following generalization of (9):

Given any set $I$, if $u \in M^I$, $a \in R^I$ are such that

$$ax = 0 \Rightarrow ux = 0 \qquad \text{for all } x \in {}^IR, \tag{10}$$

then there exists $v \in M$ such that $u = va$.

Now the characterization of injective modules can be stated as follows.

THEOREM 6.11 *A module $M$ over a ring $R$ is injective if and only if it is fully divisible.*

*Proof.* The necessity follows easily: given $u \in M^I$, $a \in R^I$ satisfying (10), let $\mathfrak{a}$ be the right ideal generated by the components of $a = (a_i)$ and define a homomorphism $\mathfrak{a} \to M$ by

$$\sum a_i x_i \mapsto \sum u_i x_i.$$

By (10) this is well defined, clearly it is a homomorphism, so it can be extended to a homomorphism $R \to M$, because $M$ is injective. If $1 \mapsto v$ in this mapping, then $a_i \mapsto va_i = u_i$, and this shows $M$ to be fully divisible.

Conversely, if $M$ is fully divisible, let $\mathfrak{a}$ be a right ideal of $R$ and $(a_i)$ a generating set of $\mathfrak{a}$. Suppose we have a homomorphism $f: \mathfrak{a} \mapsto M$, and let $a_i \mapsto u_i$ in this homomorphism. If the $x_i$ are such that $\sum a_i x_i = 0$, then $\sum u_i x_i = \sum (a_i f) x_i = (\sum a_i x_i) f = 0$, hence (10) holds and by full divisibility there exists $v \in M$ such that $u_i = va_i$. Now the map $x \mapsto vx$ of $R$ into $M$ extends $f$, because $a_i \mapsto va_i = u_i$, hence $\sum a_i x_i \mapsto \sum u_i x_i$ for any family $x = (x_i) \in {}^IR$. By Baer's criterion it follows that $M$ is injective. ■

Let us call $M$ *finitely divisible* if for any integer $n$ and any $u \in M^n$, $A \in R_n$ satisfying the condition

$$Ax = 0 \Rightarrow ux = 0 \qquad \text{for all } x \in {}^nR, \tag{11}$$

there exists $v \in M^n$ such that $u = vA$. Essentially this states that $M^n$ is 1-divisible, as $R_n$-module, for all $n$. Then we have

COROLLARY 6.12 *A right R-module M over a right Noetherian ring R is injective if and only if it is finitely divisible.*

The necessity is clear; to prove the sufficiency, we observe that by Morita equivalence we have for all $n$,

$$M \cong \mathrm{Hom}_R(R, M) \cong \mathrm{Hom}_{R_n}({}^nR, {}^nM).$$

Let $\mathfrak{a}$ be a right ideal of $R$, generated by $n$ elements say. Then $\mathfrak{a}$ corresponds to a submodule $T$ of ${}^nR$ generated by a single element (namely the column with the generating set of $\mathfrak{a}$ as components). As in the proof of the theorem, any homomorphism $T \to {}^nM$ extends to a homomorphism ${}^nR \to {}^nM$. Thus in the commutative diagram

$$\begin{array}{ccccc}
\mathrm{Hom}_R(R, M) & \longrightarrow & \mathrm{Hom}_R(\mathfrak{a}, M) & \longrightarrow & 0 \\
\downarrow \cong & & \downarrow \cong & & \\
\mathrm{Hom}_{R_n}({}^nR, {}^nM) & \longrightarrow & \mathrm{Hom}_{R_n}(T, {}^nM) & \longrightarrow & 0
\end{array}$$

the bottom row is exact; hence so is the top row. Now the result follows again by Baer's criterion. ■

## Exercises

(1) Show that a module is flat whenever every finitely generated submodule is contained in a flat submodule.

(2) Show that a flat module over an integral domain is torsion-free, and that the converse holds over a principal ideal domain.

(3) Show that a direct sum of modules is flat iff each term is flat.

(4) Use Prop. 6.6 to show that $M_R$ is flat iff the canonical map $M \otimes \mathfrak{a} \to M$ is injective for every finitely generated left ideal $\mathfrak{a}$ of $R$. Hence obtain another proof that (a)⇔(d) in Th. 6.2.

(5) Show that every finitely related projective module is a direct sum of a free module and a finitely presented module.

(6) Show that every finitely generated projective module is finitely presented.

(7) Show that every simple module over a commutative regular ring is injective.

(8) Let $R$ be a commutative integral domain with field of fractions $K$. Show that $K$ is flat as $R$-module, but not free, unless $K = R$.

(9) Show that every right Noetherian regular ring is semisimple.

(10) Let $E$ be an injective cogenerator and for any module $M$ define $I^0(M) = E^{\mathrm{Hom}(M,E)}$. Show that $M$ may be embedded in $I^0(M)$, and that this module is injective.

(11) Let $M$ be a flat module. Show that if $M$ has a resolution of finite length by finitely generated projective modules, then $M$ is projective.

(12) Show that every finitely generated projective module over $\mathbf{Z}/p^n$ is free.

(13) Let $R$ be any ring and $V$ a left $R$-module generated by a family $(f_\lambda)$ with defining relations $\sum a_{i\lambda} f_\lambda = 0$. Let $U$ be a right $R$-module. Show that, given any relation

$$\sum x_\lambda \otimes f_\lambda = 0 \qquad \text{in } U \otimes V,$$

where $x_\lambda \in U$ and almost all $x_\lambda$ vanish, there exist $y_i \in U$, almost all 0, such that $x_\lambda = \sum y_i a_{i\lambda}$.

(14) Let $M$ be an $R$-module with a presentation (3) in which $F_1$ is free and $K_1$ finitely generated, and another where $F_2$ is finitely generated (but not necessarily free). Show that $M$ is finitely presented. Give an example to show that if $F_1$ is finitely generated free and $K_2$ is finitely generated (but $F_2$ is not free), $M$ need not be finitely presented.

(15) Show that a presentation matrix $A$ represents a projective module iff there exists a matrix $C$ such that $AC = I$. Show that $A$ represents a flat module iff for any finite set of rows of $A$ there exists $C$ such that the corresponding rows of $AC$ have the form $I$.

(16) Show that every non-zero projective module has a maximal (proper) submodule.

## 6.7 Hochschild cohomology and separable algebras

Let $K$ be a commutative ring and $A$ a $K$-algebra; when we speak of an $A$-bimodule $M$, it will be understood that the two $K$-actions on $M$ agree, i.e. $\alpha m = m\alpha$ for all $m \in M, \alpha \in K$. It is then clear that an $A$-bimodule is the same as a right $(A^0 \otimes A)$-module, when $A^0$ is the opposite ring of $A$. We shall write $A^e$ for $A^0 \otimes A$ and call it the *enveloping algebra* of $A$. We shall see in Ch. 7 that for a finite-dimensional simple algebra the enveloping algebra is a full matrix algebra over the centre.

The free right $A^e$-module on one free generator is $A \otimes A$, with multiplication rule

$$(x \otimes y)(a \otimes b) \mapsto ax \otimes yb.$$

Similarly $A$ itself is a right $A^e$-module with multiplication rule $x(a \otimes b) \mapsto axb$, and the multiplication mapping $\mu: A \otimes A \to A$ defined by $(x \otimes y)\mu = xy$ is an $A^e$-module homomorphism. Its kernel is again denoted by $\Omega$, as in **2.2**, so that

we have an exact sequence

$$0 \to \Omega \to A \otimes A \xrightarrow{\mu} A \to 0. \tag{1}$$

If this sequence of $A^e$-modules splits, $A$ is said to be *separable* over $K$. We shall soon see that this is consistent with the use of this term in **5.5**. We begin by giving some equivalent conditions for separability:

PROPOSITION 7.1 *Let $K$ be a commutative ring and $A$ a $K$-algebra, with enveloping algebra $A^e$. Then the following conditions are equivalent:*

(a) *$A$ is separable,*

(b) *$A$ is projective as $A^e$-module,*

(c) *there exists $e \in A^e$ such that $e\mu = 1$ and $ae = ea$ for all $a \in A$.*

*Proof.* (a)$\Leftrightarrow$(b) is clear from the definition of projective module. Now assume that $A$ is separable and let $\lambda: A \to A \otimes A$ be a section, i.e. an $A^e$-module homomorphism such that $\lambda\mu = 1_A$. The image of $1 \in A$ under $\lambda$ is of the form $e = \sum p_i \otimes q_i$ and since $\lambda\mu = 1$ we have $e\mu = \sum p_i q_i = 1$. Further, for any $a \in A$, $ae = a \cdot 1\lambda = (a \cdot 1)\lambda = (1 \cdot a)\lambda = 1\lambda \cdot a = ea$, and (c) follows. If (c) holds, and $e = \sum p_i \otimes q_i$, then $\lambda: x \mapsto \sum p_i x q_i$ defines a splitting of (1) and it follows that $A$ is separable, i.e. (a). ■

The element $e$ in (c) is called a *separability idempotent* of $A$. The projective dimension of $A$ as $A^e$-module is sometimes called the *bidimension*, written bidim $A$; thus bidim $A = 0$ iff $A$ is separable.

Let $A$ be a $K$-algebra; any $A$-bimodule $M$ can be regarded as a right $A^e$-module in the usual way: $m(a \otimes b) = amb$, or also as left $A^e$-module, by the rule $(a \otimes b)m = bma$. We define the $n$th *homology group* of $M$ as the derived functor of $M \mapsto A^e \otimes M$, namely

$$H_n(A, M) = \operatorname{Tor}_n^{A^e}(A, M),$$

where $M$ is regarded as left $A^e$-module. Secondly we define the $n$th *cohomology group* of $M$ as the derived function of $M \mapsto \operatorname{Hom}_{A^e}(A, M)$:

$$H^n(A, M) = \operatorname{Ext}_{A^e}^n(A, M),$$

where $M$ is regarded as right $A^e$-module. These groups were first introduced (when $K$ is a field) by Hochschild in 1945 and are called the *Hochschild groups* of $M$ and $A$.

To compute these groups we construct the standard resolution for algebras as follows:

$$\cdots \to S_n \to S_{n-1} \xrightarrow{d} \cdots \xrightarrow{d} S_0 \xrightarrow{\varepsilon} A \to 0, \tag{2}$$

where $S_n$ is the $(n+2)$-fold tensor product of $A$ with itself over $K$, defined as $A$-bimodule in the natural way; thus the $A^e$-module structure is given by

$$(x_0 \otimes x_1 \otimes \cdots \otimes x_{n+1})(a \otimes b) = ax_0 \otimes x_1 \otimes \cdots \otimes x_{n+1}b.$$

The map $\varepsilon$ is the multiplication and the differential $d: S_n \to S_{n-1}$ is given by

$$(x_0 \otimes \cdots \otimes x_{n+1})d = \sum(-1)^i x_0 \otimes \cdots \otimes x_i x_{i+1} \otimes \cdots \otimes x_{n+1}.$$

It is clearly an $A^e$-homomorphism, and $d^2 = 0$ can be checked as in **4.1**, using the associativity of $A$. To show that (2) is acyclic, we define a homotopy $h: S_n \to S_{n+1}$ by $u \mapsto 1 \otimes u (u \in S_n)$, together with $\eta: A \to S_0$ given by $x \mapsto 1 \otimes x$. Then it is easily verified that $dh + hd = 1$, $\eta\varepsilon = 1_A, h_0 d + \varepsilon\eta = 1_{S_0}$.

Now $S_0 = A \otimes A$ is free of rank 1, while $S_n = A \otimes S_{n-2} \otimes A = S_{n-2} \otimes A^e$ is $A^e$-projective whenever $S_{n-2}$ is $K$-projective, e.g. when $K$ is a field. Thus when $A$ is $K$-projective, (2) is a projective resolution of $A$ and may be used to compute the Hochschild groups. We note that $H^0(A, M) = \mathrm{Hom}_{A^e}(A, M) = \{u \in M \mid au = ua$ for all $a \in A\}$; this group is also denoted by $M^A$.

The 1-cocycles are functions from $A$ to $M$ such that

$$f(xy) = xf(y) + f(x)y,$$

i.e. derivations, while coboundaries are inner derivations:

$$f(x) = xa - ax.$$

The 2-cocycles are functions $f: A^2 \to M$ such that

$$f(xy, z) + f(x, y)z = f(x, yz) + xf(y, z),$$

while a coboundary is of the form

$$f(x, y) = xg(y) - g(xy) + g(x)y.$$

To interpret $H^2(A, N)$, let us consider algebra extensions of $A$, where $A$ is $K$-projective, i.e. short exact sequences

$$0 \to N \xrightarrow{\alpha} B \xrightarrow{\beta} A \to 0, \tag{3}$$

where $B$ is a $K$-algebra with $A$ as quotient and $N$ as kernel. For simplicity we shall take $N$ to be contained in $B$, so that $\alpha$ is the inclusion mapping. Let $\gamma: A \to B$ be a $K$-linear mapping left inverse to $\beta$ (which exists since $A$ is $K$-projective). If $\gamma$ is a homomorphism, $B$ is the direct sum of $N$ and the subalgebra $A\gamma$ and we shall say that the sequence (3) *splits*. In general the failure of $\gamma$ to be a homomorphism is measured by

$$f(a, b) = (ab)\gamma - (a\gamma)(b\gamma). \tag{4}$$

Since $\beta$ is an algebra homomorphism inverse to $\gamma$, we have $f(a, b)\beta = ab - ab = 0$, hence $f(a, b) \in N$. Let us assume further that $N^2 = 0$; then $N$, which is a

$B$-bimodule, may be regarded as a $(B/N)$-bimodule, i.e. an $A$-bimodule. In that case $B$ is completely determined by $A, N$ and the function $f: A^2 \to N$, as the $K$-module $A \oplus N$ with multiplication

$$(a, u)(b, v) = (ab, ub + av + f(a, b)) \qquad (a, b \in A, u, v \in N). \tag{5}$$

As one might expect, $f$ satisfies a factor set condition, derived from the associative law. Write (4) as $(ab)\gamma = (a\gamma)(b\gamma) + f(a, b)$; then $(ab \cdot c)\gamma = (ab)\gamma \cdot c\gamma + f(ab, c) = a\gamma \cdot b\gamma \cdot c\gamma + f(a, b)c + f(ab, c)$. Similarly $(a \cdot bc)\gamma = a\gamma \cdot b\gamma \cdot c\gamma + af(b, c) + f(a, bc)$, hence

$$f(a, b)c + f(ab, c) = af(b, c) + f(a, bc);$$

this shows $f$ to be a cocycle. The extension splits iff there is a section $\gamma$ for $\beta$ which is a homomorphism. This means that there is a mapping $g: A \to N$ such that $\gamma$, defined by $a\gamma = (a, -g(a))$ is a homomorphism, i.e. by (5),

$$a\gamma \cdot b\gamma = (ab, -g(a)b - ag(b) + f(a, b)) = (ab)\gamma = (ab, -g(ab)),$$

whence

$$f(a, b) = ag(b) - g(ab) + g(a)b.$$

This is just the condition for $f$ to be a coboundary, and so we get

PROPOSITION 7.2 *Let $A$ be a $K$-algebra which is projective as $K$-module, and let $N$ be an $A$-bimodule. Then the set of isomorphism classes of algebra extensions of $A$ by $N$ as ideal with zero multiplication is in a natural bijection with $H^2(A, N)$.* ■

In a similar way the first cohomology group describes the module extensions. Given two right $A$-modules $U, V$, we can form $\mathrm{Hom}_K(U, V)$, which inherits the right $A$-module structure from $V$ and a left $A$-module structure from $U$. Now a verification, as in **3.3**, shows that we have an $A$-bimodule structure. We now have the following result:

THEOREM 7.3 *Let $A$ be a $K$-algebra and $U, V$ right $A$-modules, where $A$ and $U$ are projective as $K$-modules. Then*

$$H^n(A, \mathrm{Hom}_K(U, V)) \cong \mathrm{Ext}_A^n(U, V). \tag{6}$$

*Proof.* We start from the situation $({}_AA_A, {}_KU_A, {}_KV_A)$. As we have just seen, there is a natural $A$-bimodule structure on $\mathrm{Hom}_K(U, V)$ and as is easily verified, we have a natural isomorphism (essentially by adjoint associativity)

$$\mathrm{Hom}_{A^e}(A, \mathrm{Hom}_K(U, V)) \to \mathrm{Hom}_A(U \otimes_A A, V) \cong \mathrm{Hom}_A(U, V). \tag{7}$$

Let $X$ be a free resolution of $A$ as $A^e$-module; then as free $A$-bimodule $X_n$ has

the form $A \otimes_K F_n \otimes_K A$, where $F_n$ is a free $K$-module. Now the tensor product (over $K$) of free $K$-modules is free, hence the tensor product of projective $K$-modules is projective, so $X_n = (A \otimes_K F_n) \otimes_K A$ is right $A$-projective and $U \otimes_A X_n = (U \otimes_K F_n) \otimes_K A$ is right $A$-projective. By (7) we obtain an isomorphism of complexes

$$\mathrm{Hom}_{A^e}(X, \mathrm{Hom}_K(U, V)) \cong \mathrm{Hom}_A(U \otimes_A X, V).$$

The complex on the left has cohomology groups $\mathrm{Ext}^n_{A^e}(A, \mathrm{Hom}_K(U, V)) \cong H^n(A, \mathrm{Hom}(U, V))$. On the right, since clearly $\mathrm{Tor}^A_n(U, A) = 0$ for $n \geqslant 1$, $U \otimes_A X$ is a resolution for $U$. As we saw, it is $A$-projective, and so we obtain for its cohomology group $\mathrm{Ext}^n_A(U, V)$ and (6) follows. ■

We observe that the hypotheses of Th. 7.3 on $A$ and $U$ are satisfied when $K$ is a field. The results of Th. 7.3 and Prop. 7.2 can now be combined to establish one of the main theorems on the splitting of algebra extensions.

THEOREM 7.4 (Wedderburn's principal theorem) *Let $B$ be a finite-dimensional algebra over a field $k$ and let $I$ be a nilpotent ideal in $B$ such that bidim $B/I \leqslant 1$. Then $B$ has a subalgebra $A$ which is a complement of $I$ as $k$-space*:

$$B = A \oplus I. \tag{8}$$

*Proof.* If $I^2 = 0$, this follows from Prop. 7.2, because $H^2(B/I, I) = 0$, so we shall assume that $I^2 \neq 0$ and use induction on the dimension of $B$. The algebra $B' = B/I^2$ has lower dimension than $B$ and $B'/I/I^2 \cong B/I$, hence by the induction hypothesis there exists a subalgebra $C$ of $B$ such that $C \supseteq I^2$ and

$$B = C + I, \qquad C \cap I = I^2.$$

Now $C/I^2 = C/(I \cap C) \cong (C + I)/I = B/I$ satisfies the same hypothesis; since $I$ is nilpotent, $I^2 \subset I$, hence $C \subset B$ and applying induction again, we find a subalgebra $A$ of $C$ such that $C = A \oplus I^2$. Now $B = C + I = A + I^2 + I = A + I$ and $A \cap I = A \cap C \cap I = A \cap I^2 = 0$, therefore (8) holds. ■

In particular, taking $I$ to be the Jacobson radical of $B$, we see that $\mathbf{J}(B)$ is complemented whenever $B/\mathbf{J}(B)$ is separable. In that case the result can be proved more explicitly as follows. Let $e = \sum p_i \otimes q_i$ be a separability idempotent for $B/I$. Given the cocycle $f(a, b)$ arising from a section, we put $g(a) = \sum f(a, p_i) q_i$. Then

$$\begin{aligned} ag(b) - g(ab) + g(a)b &= \sum af(b, p_i)q_i - \sum f(ab, p_i)q_i + \sum f(a, p_i)q_i b \\ &= \sum f(a, b)p_i q_i - \sum f(a, bp_i)q_i + \sum f(a, p_i)q_i b. \end{aligned}$$

The last two terms cancel because $be = eb$ and the first reduces to $f(a, b)$ because

$e\mu = 1$; hence the right-hand side is just $f(a, b)$, and this shows $f$ to be a coboundary.

It remains to determine the separable algebras. In the first place we note that every separable algebra over a field is finite-dimensional; this follows from

PROPOSITION 7.5 (Villamayor–Zelinsky 1966) *Let $A$ be a separable $K$-algebra. If $A_K$ is projective, then it is finitely generated over $K$.*

*Proof.* Since $A$ is separable, it is projective as $A^e$-module, hence the same is true of $A^0$. Let us write $A^0 \oplus B = F$, where $F$ is a free $K$-module with a basis $u_i (i \in I)$, and write $\pi: F \to A^0$ for the projection. Let $\alpha_i \in F^*$ be the dual basis and put $\beta_i = \alpha_i | A^0$; then we have

$$x = \sum (\beta_i, x) u_i = \sum (\beta_i, x) \pi(u_i) \qquad \text{for all } x \in A^0. \tag{9}$$

The mapping $(x, y) \mapsto (\beta_i, x)y$ from $A^0 \times A$ to $A$ is bilinear, hence there exists $\varphi_i: A^e \to A$ such that

$$(\varphi_i, x \otimes y) = (\beta_i, x)y. \tag{10}$$

Moreover, for any $z \in A$,

$$(\varphi_i, (x \otimes y)z) = (\varphi_i, x \otimes yz) = (\beta_i, x)yz = (\varphi_i, x \otimes y)z.$$

Hence by linearity we have

$$(\varphi_i, wz) = (\varphi_i, w)z \qquad \text{for all } w \in A^e, z \in A, \tag{11}$$

and for any $w, z$ the two sides of (11) vanish for almost all $i$, because this is true of the $\beta_i$. We claim that there exists a family $w_i \in A^e (i \in I)$ such that for any $w \in A^e$,

$$w = \sum w_i (\varphi_i, w). \tag{12}$$

This is an analogue of the dual basis lemma (Prop. 4.5.5, A.2). We define $w_i = \pi(u_i) \otimes 1_A$; then for any $x \in A^0, y \in A$,

$$\begin{aligned} x \otimes y &= \sum (\beta_i, x) \pi(u_i) \otimes y = \sum \pi(u_i) \otimes (\beta_i, x)y \\ &= \sum w_i (\varphi_i, x \otimes y), \end{aligned}$$

by (10); now (12) follows by linearity.

Now let $e = \sum p_k \otimes q_k$ be a separability idempotent for $A^e$. By definition $e\mu = 1$ and $ex = xe$ for $x \in A$. Hence for any $y \in A$,

$$\begin{aligned} y = 1_A y = e\mu \cdot y = (ey)\mu &= [\sum w_i (\varphi_i, ey)]\mu \qquad \text{by (12)}, \\ &= \sum (w_i \mu)(\beta_i, p_k) q_k y, \end{aligned}$$

because $\mu$ is a bimodule homomorphism. Hence $A$ is generated by the family $(w_i \mu) q_k$, where $i$ ranges over the finite set of indices such that $(\varphi_i, ey) \neq 0$. Since $(\varphi_i, ey) = (\varphi_i, e)y$, this set is contained in the set $I_0 = \{i \in I | (\varphi_i, e) \neq 0\}$ which is

finite and independent of $y$. Thus we have found a finite generating set for $A$. ■

For the rest of this section we shall confine ourselves to algebras over a field. Our aim will be to show that an algebra over a field $k$ is separable iff it is semisimple and remains so under all extensions of $k$. We begin with some generalities.

PROPOSITION 7.6 (i) *If $A, B$ are separable algebras over a field $k$, then so are $A \oplus B$ and $A \otimes B$,*

(ii) *given a $k$-algebra $A$ and a field extension $F$ of $k$, the algebra $A_F = A \otimes_k F$ is separable if and only if $A$ is,*

(iii) *for any field $k$ and any $n \geqslant 1$, the full matrix ring $k_n$ is separable.*

*Proof.* (i) The separability may be described by the existence of a section $\lambda$ for the multiplication $\mu$. If $\lambda_A, \lambda_B$ are sections for the multiplication in $A$ and $B$ respectively, then $\lambda_A + \lambda_B : A \oplus B \to (A \otimes A) \oplus (A \otimes B) \oplus (B \otimes A) \oplus (B \otimes B)$ is a section for the multiplication in $A \oplus B$, while $\lambda_A \otimes \lambda_B : A \otimes B \to A \otimes B \otimes A \otimes B \cong A \otimes A \otimes B \otimes B$ is a section for $A \otimes B$.

(ii) Let $e = \sum p_i \otimes q_i$ be a separability idempotent for $A$; then $e$ is still a separability idempotent for $A_F$, for $e\mu = 1$ and $ea = ae$ continue to hold for $a \in A_F$. Conversely, if $A_F$ is separable, choose a basis $\{u_i\}$ for $F$ over $k$ such that $u_0 = 1$ and write the idempotent for $A_F$ as

$$e = \sum p_{ij} \otimes q_{ij} \otimes u_i, \qquad \text{where } p_{ij}, q_{ij} \in A.$$

Then $\sum p_{ij} q_{ij} u_i = 1$, hence equating coefficients of $u_0$ we find that $\sum p_{0j} q_{0j} = 1$. Further, for any $a \in A$ we have $ae = ea$, i.e. $\sum a p_{ij} \otimes q_{ij} \otimes u_i = \sum p_{ij} \otimes q_{ij} a \otimes u_i$ and on equating coefficients of $u_0$ we find that $\sum a p_{0j} \otimes q_{0j} = \sum p_{0j} \otimes q_{0j} a$, hence $e_0 = \sum p_{0j} \otimes q_{0j}$ is a separability idempotent for $A$, showing $A$ to be separable.

(iii) If $A = k_n$, then $A^e = k_n \otimes k_n = k_{n^2}$, hence any $A^e$-module is semisimple, and so $A$ is separable. ■

We can now describe separable algebras over a field:

THEOREM 7.7 *Let $A$ be an algebra over a field $k$. Then $A$ is separable if and only if $A_F$ is semisimple for all extension fields $F$ of $k$. Moreover, when this holds, then $A$ is finite-dimensional over $k$.*

*Proof.* $A$ is finite-dimensional whenever it is separable or semisimple, by Prop. 7.5 and A.2, Th. 5.3.4; so we may take $A$ to be finite-dimensional in what follows. Assume $A$ separable; then by Th. 7.3 all module extensions split, i.e. every $A$-module is semisimple, so $A$ is semisimple. By Prop. 7.6, $A_F$ is also separable, so $A_F$ is semisimple, for every extension field $F$ of $k$.

Conversely, if $A_F$ is semisimple for all $F \supseteq k$, take $F$ to be an algebraic closure

of $k$. Then $A_F$ is a direct product of full matrix algebras $F_n$. By Prop. 7.6, each $F_n$ is separable and $A_F$ is separable, hence so is $A$. ■

We note that in this theorem it is enough to require $A_F$ to be semisimple for all finite field extensions of $k$. The theorem also shows that for finite-dimensional commutative algebras over a field the notion of separability introduced here reduces to that of **5.5**. For a commutative $k$-algebra $A$ is separable in the sense of this section iff $A_F$ is semisimple for all field extensions $F$ of $k$, and by the form of the radical in Artinian rings (A.2, Th. 5.4.5) this just means that $A_F$ is reduced.

## Exercises

(1) Verify that a separability idempotent for $A$ is indeed idempotent.

(2) Show that for any commutative ring $K$ and any $n \geqslant 1$, $K_n$ is separable by verifying that $\sum e_{i1} \otimes e_{1i}$ (where $e_{ij}$ are the usual matrix units) is a separability idempotent for $K_n$.

(3) Let $A$ be a separable algebra over a field $k$. Show that any right $A$-module $M$ is a direct summand of $M \otimes_k A$ and hence is projective. Deduce that $A$ is semisimple.

(4) Show that an algebra over a field $k$ is separable iff it is semisimple and its centre is a direct product of separable field extensions of $k$.

(5) Show that a $K$-algebra $A$ is separable iff the functor $M \mapsto M^A$ for any $A$-bimodule $M$, is exact.

(6) Let $k$ be a field of characteristic $p \neq 0$ and $F = k(\alpha)$ a $p$-radical extension of degree $p$, where $\alpha^p = a \in k$. Let $A$ be the $k$-algebra generated by $u$ with the defining relation $(u^p - a)^2 = 0$. Show that $J = \mathbf{J}(A)$ is spanned by $v = u^p - a$ and that $\bar{A} = A/J$ is semisimple, but $\bar{A}_F$ is not. Verify that $A$ contains no subalgebra $\cong \bar{A}$.

(7) Show that if $E/k$ is separable and $A$ is a commutative $K$-algebra, then $A_E$ is separable over $A$.

(8) Let $K$ be a commutative ring and $G$ a finite group whose order $n$ is a unit in $K$. Show that the group algebra $KG$ is separable, by verifying that $(1/n)\sum g^{-1} \otimes g$ is a separability idempotent.

(9) Let $E$ be a commutative separable $K$-algebra. Show that for any $K$-algebra $A$, $\mathrm{gl\cdot dim}(A \otimes_K E) = \mathrm{gl\cdot dim}(A)$. (Hint: use the separability idempotent of $E$ over $K$ to define an averaging operator as in Ex. (8).)

(10) Use Ex. (9) to show that a $K$-algebra $A$ is $K$-projective, i.e. an exact sequence of $A$-modules which is $K$-split is $A$-split.

## Further exercises on Chapter 6

(1) Show that if $M_1 \oplus M_2 \cong N_1 \oplus N_2$ and $M_1 \cong N_1$ then $M_2 \cong N_2$. (Hint: take the isomorphism in the form $\alpha = (\alpha_{ij})$, where $\alpha_{ij}: M_i \to N_j$ and use the fact that $\alpha$ and $\alpha_{11}$ are invertible.)

(2) If $R$ is a Dedekind domain and $\mathfrak{a}, \mathfrak{b}$ are any non-zero ideals in $R$, then $\mathfrak{a} \oplus \mathfrak{b} \cong R \oplus \mathfrak{ab}$ (A.2, Th. 9.6.11). Use this result to show that the Krull–Schmidt theorem fails over any Dedekind domain which is not a principal ideal domain.

(3) Show that for any projective left $R$-module $P$, the intersection of all maximal submodules is equal to $JP$. Give an example of a module for which this fails; can this module be finitely generated?

(4) Let $R$ be a ring which can be written as a finite direct sum of indecomposable left ideals with simple tops. Show that $R$ is semiperfect. Deduce that a ring is semiperfect iff every finitely generated module has a projective cover.

(5) Let $R$ be a semiperfect ring. Show that any homomorphic image of an indecomposable projective $R$-module is indecomposable. Does this remain true for more general indecomposable modules?

(6) Show that the endomorphism ring of any indecomposable injective module is a local ring and deduce a Krull–Schmidt theorem for injective modules.

(7) Show that finitely generated projective modules are isomorphic iff they have isomorphic tops.

(8) Let $R$ be any ring and $R_\omega$ the ring of all row-finite matrices (with countably many rows and columns) over $R$. Show that $R \prec R_\omega$ but that $R, R_\omega$ are not Morita equivalent. What is the relation between (i) the ring $R_\omega$ of all row-finite matrices, (ii) the ring ${}_\omega R$ of all column-finite matrices and (iii) $R_\omega \cap {}_\omega R$?

(9) A ring $R$ is said to be *basic* if $R/\mathbf{J}(R)$ is a direct product of a finite number of skew fields. Show that the ring of upper triangular $n \times n$ matrices over a local ring is basic.

(10) Show that for every semiperfect ring $A$ there exists a basic ring $B$ such that $B \sim A$ and that $B$ is unique up to isomorphism.

(11) A functor $T$ is called *faithfully exact* if it is faithful, exact and preserves coproducts. An object $P$ in an abelian category $\mathscr{A}$ such that $h^P = \mathscr{A}(P, -)$ is faithfully exact is called *faithfully projective*. Show that $P$ is faithfully projective iff $P$ is a projective generator such that $\mathscr{A}(P, \coprod X_i) \cong \coprod \mathscr{A}(P, X_i)$, for any family $(X_i)$ of $\mathscr{A}$-objects.

(12) Let $P$ be a faithfully projective object in an abelian category $\mathscr{A}$. Write $A = \mathscr{A}(P, P)$, $hX = \mathscr{A}(P, X)$ for $X \in Ob\, \mathscr{A}$; verify that $A$ is a ring and $hX$ a left $A$-module, under composition of maps. Verify also that $h$ is a functor from $\mathscr{A}$ to ${}_A\mathscr{M}$ and use Th. 4.1 to show that it is an equivalence of functors.

Deduce that any abelian category with coproducts and a faithfully projective object is equivalent to a category of modules over a ring (Mitchell–Freyd theorem).

(13) A short exact sequence is said to be *pure* if it stays exact under tensoring. Show that a module $F$ is flat iff every short exact sequence with third term $F$ is pure.

(14) Show that $M'$ is pure in $M$ (i.e. the short exact sequence with first two terms $M', M$ is pure) iff $uA = p$, where $u \in M'^m, p \in M'^n, A \in {}^mR^n$ implies that $u'A = p$ for some $u' \in M'^m$.

(15) Show that for any module $U$, the correspondence $U \mapsto \hat{U} = \mathrm{Hom}_{\mathbf{Z}}(U, \mathbf{K})$ is a faithful covariant functor. Show also that there is natural transformation $U \to \hat{\hat{U}}$ which is an embedding.

(16) Show that a module $M$ is flat iff for every homomorphism $\alpha: P \to M$ where $P$ is finitely generated projective and every $x \in \ker \alpha$ there is a factorization $\alpha = \beta\gamma$, where $\beta: P \to Q$, $\gamma: Q \to M, Q$ finitely generated projective, such that $x\beta = 0$.

(17) Let $A$ be a $K$-algebra, and $M$ an $A$-bimodule. Show that the sequence

$$0 \to M^A \to M \to \mathrm{Der}_K(A, M) \to H^1(A, M) \to 0$$

is exact, where $\mathrm{Der}_K(A, M) = \mathrm{Hom}_{A^e}(\Omega, M)$ is the module of derivations of $A$ into $M$.

(18) (A. I. Malcev) Let $B$ be an algebra over a field with a nilpotent ideal $I$ such that $B/I$ is separable. If $B = A_1 \oplus I = A_2 \oplus I$ are two splittings, show that $A_1, A_2$ are conjugate by an automorphism of the form $x \mapsto (1-u)^{-1}x(1-u)$, where $u \in I$. (Hint: if $a \mapsto a_i$ is an isomorphism $B/I \to A_i$, examine the cochain $f(a) = a_1 - a_2 \in I$.)

# 7
# Central simple algebras

Skew fields are more complicated than fields and much less is known about them. But in the special case of division algebras (the case of finite dimension over the centre) the situation is rather better. It is convenient to include full matrix rings over division algebras, thus our topic is essentially the class of simple Artinian rings. Although some of our results are proved in this generality we shall soon specialize to the finite-dimensional case.

There is no space to enter into such interesting questions as the discussion of division algebras over number fields, but we shall describe an important field invariant, the Brauer group, and its significance for division algebras. Then, after a look at quaternion algebras, we introduce crossed products and illustrate them on cyclic algebras.

## 7.1 Simple Artinian rings

It is clear from Wedderburn's theorem that every simple Artinian ring is an algebra over a field; it also contains a skew field and is finite-dimensional over the latter, but it need not be finite-dimensional as an algebra. We begin by not imposing any finiteness restriction, in fact we shall not even assume our rings to be Artinian, although not very much can be said in that generality.

Let $A$ be any ring and denote its centre by $C$. We may regard $A$ as $A$-bimodule or equivalently as right $A^e$-module, where $A^e = A^0 \otimes A$ as in **6.7**. The centralizer of $A^e$ acting on $A$, i.e. of all left multiplications $\lambda_a$ and all right multiplications $\rho_b$ of $A$ is the intersection of the centralizers of all the $\lambda_a$ and $\rho_b$; this is the set of all left multiplications commuting with all left multiplications, i.e. multiplications by elements of $C$. Thus the centralizer of the action of $A^e$ on $A$ is just $C$. If further, $A$ is a simple ring, it is simple as $A^e$-module and by Schur's lemma, $C$ is then a field.

Let $k$ be a field. Given a $k$-algebra $R$ and a right $R$-module $M$, the action of $R$ on $M$ is said to be *dense* if, given $x_1, \ldots, x_r \in M$ and $\theta \in \mathrm{End}_k(M)$, there exists $a \in R$ such that $x_i a = x_i \theta \, (i = 1, \ldots, r)$. In this definition we may clearly omit any $x_i$ linearly dependent on the rest. Hence an equivalent definition is obtained by requiring that for any $x_1, \ldots, x_r$, $y_1, \ldots, y_r \in M$, where the $x_i$ are linearly independent over $k$, there exists $a \in R$ such that $x_i a = y_i$. In particular, when $M$

is finite-dimensional over $k$, this means that every $k$-linear mapping of $M$ can be accomplished by acting with some element of $R$.

THEOREM 1.1 (Density theorem) *The centre of any simple ring is a field. If $A$ is a simple ring with centre $k$, then $A^e = A^0 \otimes A$ is dense in the ring of all $k$-linear transformations of $A$. In particular, when $[A:k] = n$ is finite, then*

$$A^e = A^0 \otimes A \cong \mathfrak{M}_n(k). \tag{1}$$

*Proof.* (Artin–Whaples 1943) Since $A$ is simple, $k$ is a field. Let $x_1, \ldots, x_r \in A$ be linearly independent over $k$ and $y_1, \ldots, y_r \in A$; we have to find $\varphi \in A^e$ such that $x_i\varphi = y_i$. For $r = 1$ this follows by the simplicity of $A$, so let $r > 1$. By induction there exists $\varphi_i' \in A^e$ such that $x_j\varphi_i' = \delta_{ij}$ for $i, j = 2, \ldots, r$. If $x_1\varphi_i' \notin k$ for some $i$, then for suitable $b \in A$, $\psi = \varphi_i'(b \otimes 1 - 1 \otimes b)$ satisfies $x_1\psi \neq 0$, $x_j\psi = 0$ for $j > 1$, and for some $\theta \in A^e$, $x_1\psi\theta = 1$, hence $\varphi_1 = \psi\theta$ maps $x_1$ to 1 and the $x_j (j > 1)$ to 0. Now $\varphi_i = \varphi_i' - \varphi_1(1 \otimes x_1\varphi_i')$ maps $x_i$ to 1 and the other $x$'s to 0, hence $\varphi = \Sigma\varphi_i(1 \otimes y_i)$ is the required map.

There remains the case where $\lambda_i = x_1\varphi_i' \in k$ for $i = 2, \ldots, r$. Put $\psi = \sum\varphi_i'(1 \otimes x_i) - 1$; then $x_j\psi = 0$ for $j = 2, \ldots, r$, $x_1\psi = \sum\lambda_i x_i - x_1$. By the linear independence of the $x$'s this is not zero and for a suitable $\theta \in A^e$, $x_1\psi\theta = 1$; now the proof can be completed as before.

In the finite case we have a surjective homomorphism $A^e \to \mathrm{End}_k(A) \cong k_n$, and now the isomorphism (1) follows by counting dimensions, which are $n^2$ on both sides. ■

That (1) is an isomorphism also follows from the fact, soon to be proved (in Cor. 1.3) that $A^e$ is simple, for any simple ring $A$ with centre $k$.

Let $A$ be a $k$-algebra, where $k$ is a field. Then for any $\alpha \in k$, $a, b \in A$ we have

$$\alpha(ab) = (\alpha a)b = a(\alpha b).$$

Taking $b = 1$, we see that $\alpha a = a\alpha$, thus the mapping $\alpha \mapsto \alpha.1$ is a homomorphism of $k$ into the centre of $A$; since $k$ is a field, this is actually an embedding whenever $A \neq 0$. Conversely, any ring whose centre contains $k$ as a subfield, may be considered as a $k$-algebra in this way. Thus a non-trivial $k$-algebra is essentially a ring whose centre contains $k$ as a subfield. If the centre of $A$ is precisely $k$, $A$ is said to be a *central* $k$-algebra. Throughout this section $k$ will be a field, fixed but arbitrary. Our convention will be that an algebra or $k$-algebra need not be finite-dimensional, but a *division algebra* is finite-dimensional over its centre. When the dimension is infinite, we shall speak of a *skew field.*

Consider the functor $A \otimes_k -$; we shall show that it preserves the ideal structure when $A$ is a central simple $k$-algebra.

THEOREM 1.2 (Azumaya–Nakayama 1947) *Let $A$ be a central simple*

*k-algebra. Then for any k-algebra B there is a lattice-isomorphism between the ideals of B and those of $A \otimes B$. In particular, the ideal lattice of $\mathfrak{M}_n(B)$ is isomorphic to that of B, for any $n \geqslant 1$.*

*Proof.* Consider the mappings

$$\mathfrak{b} \mapsto A \otimes \mathfrak{b} \qquad (\mathfrak{b} \text{ an ideal of } B) \tag{2}$$

$$\mathfrak{C} \mapsto \mathfrak{C} \cap B \qquad (\mathfrak{C} \text{ an ideal of } A \otimes B), \tag{3}$$

where $B$ is embedded in $A \otimes B$ by the natural mapping $x \mapsto 1 \otimes x$. Since $k$ is a field, all tensor products are exact, so we do have such an embedding. We claim that the mappings (2), (3) are mutually inverse; this will prove the result, for they establish a bijection which is clearly inclusion-preserving and hence a lattice-isomorphism. The last part then follows because $k_n$ is central simple and $B_n = B \otimes k_n$; of course it is also a consequence of Prop. 6.4.4.

We recall the intersection formula in a tensor product (A.2, (6) of **5.5**): If $U, V$ are $k$-spaces and $U = U' \oplus U''$, $V = V' \oplus V''$, then

$$U' \otimes V \cap U \otimes V' = U' \otimes V'.$$

Since $k$ is a field, every subspace is a direct summand. Now if $\mathfrak{b}$ is an ideal in $B$, then $A \otimes \mathfrak{b}$ is an ideal in $A \otimes B$ and so we have

$$(A \otimes \mathfrak{b}) \cap B = \mathfrak{b}. \tag{4}$$

Next let $\mathfrak{C}$ be an ideal in $A \otimes B$ and put $\mathfrak{b} = \mathfrak{C} \cap B$; then $A \otimes \mathfrak{b} \subseteq \mathfrak{C}$ and we have to establish equality here. Now any $c \in \mathfrak{C}$ can be written in terms of a basis $\{u_i\}$ of $A$ as $c = \sum u_i \otimes z_i$, where $z_i \in B$. Only finitely many of the $z_i$ are non-zero, say $z_i \neq 0$ for $i = 1, \ldots, r$. By Th. 1.1, $A^e$ acts densely on $A$, hence there exist $x_j, y_j \in A$ $(j = 1, \ldots, s)$ such that $\sum_j x_j u_i y_j = \delta_{i1}$ $(i = 1, \ldots, r)$. It follows that $\sum x_j c y_j = \sum_{ij} x_j u_i y_j \otimes z_i = 1 \otimes z_1 \in \mathfrak{C}$; so $z_1 \in \mathfrak{b}$ and similarly $z_i \in \mathfrak{b}$ for $i = 2, \ldots, r$. Therefore $\mathfrak{C} = A \otimes \mathfrak{b}$ and this is the desired equality. Thus (2), (3) are mutually inverse and the lattice isomorphism follows. ■

We observe that no assumptions about finiteness are needed here. The theorem has a number of important consequences.

COROLLARY 1.3 *If A is a central simple k-algebra and B is any k-algebra, then $A \otimes B$ is simple if and only if B is simple; further the centre of $A \otimes B$ is isomorphic to that of B. In particular, the tensor product of central simple k-algebras is again central simple.*

The assertion about the centre follows because the centre of a tensor product is the tensor product of the centres (A.2, Prop. 5.5.3). The rest is a consequence of Th. 1.2. ■

We shall sometimes want to know when a given division algebra over $k$ can be embedded in a matrix ring over a skew field $D$ containing $k$ in its centre. There is a simple answer when the centre of $D$ is a regular extension of $k$; it is given by the following result, taken from Schofield [1985].

PROPOSITION 1.4 *Let $D$ be a skew field whose centre $F$ is a regular extension of $k$, and $A$ a division algebra over $k$. Then $A^0 \otimes_k D$ is a simple Artinian ring; thus $A^0 \otimes D$ has a unique simple module $S$ which is finite-dimensional over $D$, say $[S:D] = s$. Now $A$ can be embedded in $\mathfrak{M}_n(D)$ if and only if $s|n$.*

*Proof.* Let $C$ be the centre of $A$. We have

$$A^0 \otimes_k D \cong A^0 \otimes_C (C \otimes_k D)$$

and $A$ is central simple over $C$; hence by Th. 1.2, the ideals of $A^0 \otimes D$ correspond to those of $C \otimes_k D$. Next, we have

$$C \otimes_k D \cong (C \otimes_k F) \otimes_F D,$$

so the ideals correspond to those of $C \otimes F$. By hypothesis this is an integral domain; since $[C \otimes_k F:F] = [C:k]$, which is finite, $C \otimes F$ is a field, and it follows that $A^0 \otimes D$ is simple. It is Artinian because its dimension over $D$ is finite. This also shows that the unique simple $(A^0 \otimes D)$-module has finite dimension over $D$.

Suppose now that $A$ is embedded in $D_n$; we may regard $D_n$ as the endomorphism ring of $D^n$, qua right $D$-module. In this way $D^n$ becomes an $(A, D)$-bimodule, i.e. a right $(A^0 \otimes D)$-module. If it is isomorphic to $S^r$ then a comparison of dimensions over $D$ shows that $n = rs$. Conversely, if $n = rs$, then $D^n \cong S^r$ and this is an $(A, D)$-bimodule, hence $A$ is embedded in $D_n$. ■

We note that in this result the regularity assumption can be omitted when $A$ is a central $k$-algebra.

We shall want to know when a given $k$-algebra can be written as a tensor product. First we shall look at the general case; in the finite-dimensional case we can then easily obtain a complete answer.

PROPOSITION 1.5 *Let $C$ be a $k$-algebra with a central simple subalgebra $A$, whose centralizer in $C$ is denoted by $A'$. Then the subalgebra of $C$ generated by $A$ and $A'$ is their tensor product over $k$.*

*If further $[A:k]$ is finite, then $C = A \otimes A'$ and the centre of $A'$ is the centre of $C$.*

*Proof.* By hypothesis, $A$ and $A'$ commute elementwise, so the mapping $(x, y) \mapsto xy$ gives rise to a homomorphism

$$A \otimes A' \to C. \tag{5}$$

Its kernel is an ideal in $A \otimes A'$, which by Th. 1.2 is of the form $A \otimes \mathfrak{a}$, where $\mathfrak{a}$

is the kernel of the restriction of (5) to $A'$. But this is the inclusion mapping, which is injective, so $\alpha = 0$ and (5) is injective. Clearly its image is the subalgebra generated by $A$ and $A'$.

Suppose now that $[A:k] = n$ is finite. We can regard $C$ as $A^e$-module, i.e. by Th. 1.1 as $k_n$-module. This module is semisimple, hence $C = \oplus C_\lambda$, where $C_\lambda$ is simple, isomorphic to $A$. Let $u_\lambda \in C_\lambda$ correspond to 1 in this isomorphism; then $u_\lambda a = a u_\lambda$ for all $a \in A$, hence $u_\lambda \in A'$ and so $C = AA'$. Therefore (5) is surjective, hence an isomorphism in this case. Now the assertion about the centre follows by Cor. 1.3. ■

In particular we find the identity

$$k_{rs} \cong k_r \otimes k_s, \tag{6}$$

already encountered in A.2, **5.5**. We remark that in general $A \otimes A'$ will be a proper subalgebra of $C$; for example, if $C = k\langle x, y, z\rangle$, the free algebra on $x, y, z$ and $A$ is the subalgebra generated by $x, y$, then $C$ and $A$ both have centre $k$ and $A' = k$, so $AA' \neq C$.

We next come to a basic result in the theory of central simple algebras, the Skolem–Noether theorem, which asserts that every automorphism of a finite-dimensional central simple algebra is inner. It is useful to have a slightly more general form of this result:

THEOREM 1.6 *Let $A$ be a simple Artinian ring with centre $k$ and $B$ any finite-dimensional simple $k$-algebra. Given any homomorphisms $f_1, f_2$ from $B$ into $A$, there exists a unit $u$ of $A$ such that*

$$bf_2 = u^{-1}(bf_1)u \qquad \text{for all } b \in B. \tag{7}$$

*Proof.* We regard $A$ as $(A^0 \otimes_k B)$-module, i.e. as $(A, B)$-bimodule in two ways:

$$(a, x, b)_i = ax(bf_i) \qquad \text{where } a, x \in A,\ b \in B,\ i = 1, 2. \tag{8}$$

By Cor. 1.3, $A^0 \otimes B$ is simple, and it is Artinian, for if $A = D_t$, where $D$ is a skew field and $t \geqslant 1$, then $A^0 \otimes B \cong (D^0 \otimes B)_t$ and this has finite dimension over $D^0$. If $V$ is the unique simple $(A^0 \otimes B)$-module, then every finitely generated $(A^0 \otimes B)$-module has the form $V^r$ for some $r \geqslant 0$, hence the $(A^0 \otimes B)$-module structures defined on $A$ by (8) are isomorphic to $V^r$, $V^s$ respectively. By comparing dimensions over $D^0$ (i.e. regarding $A$ as left $D$-module) we see that $r = s$, hence these structures are isomorphic. This means that there is a bijective linear transformation $\gamma: A \to A$ such that

$$[ax(bf_1)]\gamma = a(x\gamma)(bf_2) \qquad \text{where } a, x \in A,\ b \in B. \tag{9}$$

Putting $b = x = 1$ and writing $1\gamma = u$, we find that $a\gamma = au$ for all $a \in A$. Since $\gamma$ is surjective, there exists $v \in A$ such that $vu = 1$; but $\gamma$ is also injective and

$(uv-1)\gamma = uvu - u = 0$, hence $uv = 1$ and so $v = u^{-1}$. If we now put $a = x = 1$ in (9), we obtain $(bf_1)u = u(bf_2)$ and (7) follows. ■

Two subalgebras or elements are said to be *conjugate* if there is an inner automorphism mapping one to the other. From Th. 1.6 we immediately find

COROLLARY 1.7 *In any simple Artinian ring with centre k, isomorphic finite-dimensional simple k-subalgebras are conjugate and hence have conjugate centralizers.* ■

In particular this shows that every automorphism of a central simple finite-dimensional algebra is inner (Skolem–Noether theorem). More generally we have

COROLLARY 1.8 *Every automorphism of a finite-dimensional semisimple k-algebra which leaves the centre elementwise fixed, is inner.*

*Proof.* Let $\theta$ be the automorphism, write $A = A_1 \oplus \cdots \oplus A_r$ and denote the unit element of $A_i$ by $e_i$. Then $e_i$ is in the centre of $A$ and so $e_i\theta = e_i$, by hypothesis. It follows that $\theta$ maps each $A_i$ into itself: if $a \in A_i$, then $a\theta = (ae_i)\theta = a\theta \cdot e_i$, therefore $a\theta \in A_i$. Thus $\theta$ induces an endomorphism of $A_i$, likewise $\theta^{-1}$, so $\theta$ in fact defines an automorphism of $A_i$, $\theta_i$ say. Now the centre of $A_i$, $C_i$ say, is left fixed by $\theta_i$ so we may regard $\theta_i$ as an automorphism of the central simple $C_i$-algebra $A_i$. By Cor. 1.7 there exists a unit $u_i$ in $A_i$ such that $x\theta_i = u_i^{-1} x u_i$ $(x \in A_i)$. Now it is easily checked that $u = \sum u_i$ is a unit of $A$ inducing $\theta$. ■

We remark that Cor. 1.7 as it stands does not extend to the case of semisimple subalgebras (cf. Ex. (5)).

We next look at the relation between the dimension of a simple subalgebra and that of its centralizer.

THEOREM 1.9 (R. Brauer 1932) *Let $A$ be a simple Artinian ring with centre $k$ and $B$ a finite-dimensional simple subalgebra with centre $F$. Then the centralizer $B'$ of $B$ in $A$ is again simple with centre $F$ and the centralizer $B''$ of $B'$ equals $B$, while the centralizer $F'$ of $F$ is given by*

$$F' = B \otimes_F B'. \tag{10}$$

*Moreover,*

$$[A:B'] = [B:k], \tag{11}$$

*and if $[B:k] = r$, then*

$$A \otimes_k B^0 \cong B' \otimes_k k_r \cong B'_r. \tag{12}$$

*Proof.* We may regard $k_r$ as acting on $B \cong k^r$ by $k$-linear transformations. As such it contains the subalgebras $\rho_B$ of right multiplications and $\lambda_B$ of left multiplications. Clearly $\rho_B \cong B$, $\lambda_B \cong B^0$ and $A \otimes k_r$ is central simple, by Cor. 1.3. We have the isomorphic simple subalgebras $B \otimes k$, $k \otimes \rho_B$; they are conjugate by Cor. 1.7 and so have isomorphic centralizers:

$$B' \otimes k_r \cong A \otimes \lambda_B \cong A \otimes B^0.$$

Since $B' \otimes k_r \cong B'_r$, this proves (12), and comparing dimensions over $B'$, we obtain $r^2 = [A:B'][B:k]$, from which (11) follows on dividing by $r$.

Now $A \otimes B^0$ is simple, hence by (12), so is $B'$ (using Cor. 1.3 twice). Clearly $B'' \supseteq B$, $B''' = B'$, and replacing $B$ by $B''$ in (11), we find that $[B'':k] = r$, hence $B'' = B$.

Finally, if the centre of $B'$ is $E$, then $E \supseteq F$ and since $B'' = B$, we also have $F \supseteq E$, hence $E = F$. Thus $B, B'$ are central simple $F$-algebras, both subalgebras of $F'$, and now (10) follows from Prop. 1.5. ■

We observe that $B \otimes B'$ is in general distinct from $A$, for the two sides have centres $F$ and $k$ respectively. Only when $F = k$ do we have $F' = A$ and the above result then reduces to part of Prop. 1.5.

COROLLARY 1.10 *Let $A$ be a simple Artinian ring with centre $k$ and let $F$ be subfield of $A$ such that $F \supseteq k$ and $[F:k] = r$ is finite. Then*

$$A \otimes_k F \cong F' \otimes_k k_r.$$

*If moreover, $A$ is finite-dimensional, then $[F:k]^2$ divides $[A:k]$ and writing $B = F'$, we have $A \otimes_k B^0 \cong F_r$.*

This is just the case $B = F$ of Th. 1.9; here $[F':F] = n/r^2$, where $n = [A:k]$. ■

COROLLARY 1.11 *Let $A$ be a finite-dimensional central simple $k$-algebra and $F$ a subfield of $A$ containing $k$. Then the following are equivalent:*

(a) $F' = F$;

(b) *$F$ is a maximal commutative subring of $A$;*

(c) $[A:k] = [F:k]^2$;

(d) $[A:k] = [A:F]^2$.

*In particular, every maximal commutative subfield $F$ of a central division $k$-algebra $D$ satisfies $[D:k] = [F:k]^2 = [D:F]^2$.*

*Proof.* Clearly $F$ is a maximal commutative subring of $A$ iff $F' = F$, and $F$ is a subfield by hypothesis. Now for any subfield $F$, $F' \supseteq F$ and $[A:k] = [A:F][F:k] = [F:k][F':k]$, hence $[A:F]^2 \geqslant [A:k] \geqslant [F:k]^2$, with equality in either place iff $F' = F$. ■

We remark that a central simple algebra $A$ may have no subfield $F$ satisfying the conditions of Cor. 1.11, e.g. $A = k_n$, where $k$ is algebraically closed and $n > 1$. Nevertheless, the dimension of a central simple algebra is always a perfect square, for by Wedderburn's theorem, $A \cong D_n$ where $D$ is a skew field, again with centre $k$. If $F$ is a maximal subfield of $D$, then by Cor. 1.11 applied to $D$ we find that $[D:k] = [F:k]^2$, hence $[A:k] = n^2[F:k]^2$. In the next section we shall meet another proof of this important fact.

As another application of Th. 1.9 we have Wedderburn's theorem on finite fields (A.2, Th. 3.8.6). We shall need the following remark about finite groups:

LEMMA 1.12 *Let $G$ be a finite group and $H$ a proper subgroup. Then $G$ cannot be written as the union of all the conjugates of $H$.*

*Proof.* Let $|H| = h, (G:H) = n$, so that $|G| = hn$. If $a_1, \ldots, a_n$ is a right transversal for $H$ in $G$, then each conjugate of $H$ has the form $a_i^{-1}Ha_i$. There are $n$ such conjugates and each contains $h$ elements, but the unit element is common to all of them, so their union contains at most $(h-1)n + 1$ elements in all. Since $n > 1$, this is less than $hn = |G|$, so not every element of $G$ is included. ■

THEOREM 1.13 (Wedderburn's theorem on finite fields) *Any finite skew field is commutative.*

*Proof.* Let $D$ be a finite skew field, with centre $k$ and let $F$ be a maximal subfield. Then $F$ is a finite field; all maximal subfields of $D$ have the same degree $r$, say, over $k$ (Cor. 1.11) and hence are isomorphic, as minimal splitting fields of $x^{q^r} - x$, where $q = |k|$. By Cor. 1.7 they are conjugate to $F$. Now each element of $D$ lies in some maximal commutative subfield of $D$, hence $D$ is the union of conjugates of $F$. It follows that the multiplicative group $D^\times$ is a finite group, equal to the union of the conjugates of $F^\times$. But this is impossible by Lemma 1.12, hence $D$ must be commutative. ■

## Exercises

(1) Show that every finite-dimensional central simple algebra over a finite field $F$ has the form $\mathfrak{M}_n(F)$.

(2) Let $A$ be a finite-dimensional $k$-algebra. Show that if $A_E \cong E_n$ for some extension field $E/k$ and some $n \geqslant 1$, then $A$ is central simple over $k$.

(3) Let $D$ be a skew field with centre $k$ and let $E$ be a finite-dimensional subalgebra (necessarily a skew field). Show that if $E'$ is the centralizer of $E$, then the centralizer of $E'$ is $E$ and $[D:E'] = [E:k]$.

(4) Let $C$ be a finite-dimensional $k$-algebra and $A$ a central simple subalgebra. Show that $[A:k]|[C:k]$.

(5) Show that Cor. 1.7 no longer holds for semisimple subalgebras. (Hint: take appropriate diagonal subalgebras isomorphic to $k^2$ of $k_3$.)

(6) Let $R, S$ be $k$-algebras, where $k$ is a field, and let ${}_RU$, $V_S$ be modules as indicated, where $R$ acts densely on $U$ with centralizer $k$. Show that for any $0 \neq u_0 \in U$ the mapping $v \mapsto u_0 \otimes v$ embeds $V$ in $U \otimes V$. Construct a lattice-isomorphism between the $S$-submodules of $V$ and the $(R, S)$-bimodules of $U \otimes V$.

(7) Let $D$ be a skew field with centre $k$ and let $F$ be a maximal subfield of $D$. Show that $D_F$ is a dense ring of linear transformations on $D$ as $F$-space. Show that if either $[D:F]$ or $[F:k]$ is finite, then so is the other and $D_F$ is then a full matrix ring over $F$. (Hint: use the regular representation of $D$ to get a homomorphism $D \otimes_k F \to F_n$, where $n$ is the dimension of $D$ as right $F$-space.)

(8) Let $\delta$ be a derivation of a central simple algebra $A$. By representing $\delta$ as an isomorphism of subalgebras of $\mathfrak{M}_2(A)$ (cf. **2.2**) show that $\delta$ is an inner derivation.

(9) (Wedderburn 1921) Let $D$ be a skew field with centre $k$. Show that any two elements of $D$ with the same minimal equation over $k$ are conjugate.

(10) Let $D$ be a skew field with centre $C$ and $K$ a subfield with centre $H$. Show that $[K:H] \leqslant [D:C]$ and when equality holds and both are finite, then $H \subseteq C$ and $D \cong K \otimes_H C$ (A. Kupferroth).

(11) Show that if $A$ is a central simple $k$-algebra and $B$ is any $k$-algebra, then $A \otimes B$ is semisimple iff $B$ is.

(12) Let $D$ be a skew field with centre $k$ and $E$ a skew subfield with centre $C$. Show that $E$ and the subfield generated by $C$ and $k$ are linearly disjoint over $C$.

## 7.2 The Brauer group

In this section all algebras will be finite-dimensional over the ground field $k$.

Let $A$ be a central simple $k$-algebra. We know that $A$ is a full matrix ring over a skew field:

$$A \cong D_m \cong D \otimes k_m.$$

Here $D$ is a central division algebra over $k$, unique up to $k$-isomorphism, while $m$ is unique. We shall call $D$ the *skew field component* of $A$. Two central simple $k$-algebras $A, B$ are said to be *similar*, $A \sim B$, if their skew field components are isomorphic. Clearly this is an equivalence relation; we shall denote the class of $A$ by $(A)$. We now show that tensor multiplication induces a multiplication of these equivalence classes.

If $A, B$ are central simple $k$-algebras, then so is $A \otimes B$, by Cor. 1.3. Moreover, if $A \sim A'$, $B \sim B'$, say $A \cong C \otimes k_m$, $B \cong D \otimes k_n$, then $A \otimes B \cong (C \otimes k_m) \otimes (D \otimes k_n) \cong C \otimes D \otimes k_{mn}$, hence $A \otimes B \sim C \otimes D$, and similarly $A' \otimes B' \sim C \otimes D$, whence $A \otimes B \sim A' \otimes B'$. The multiplication of similarity classes is associative and commutative, by the corresponding laws for tensor products. Moreover, for any $A$, we have $A \otimes k \cong A$ and $A \otimes A^0 \cong k_n$ for some $n \geqslant 1$, hence

$$(A)(k) = (A) \qquad (A)(A^0) = (k).$$

This shows that the class $(k)$ is neutral for multiplication and $(A^0)$ is the inverse of $(A)$, and it proves

THEOREM 2.1 *For any field $k$, the similarity classes of finite-dimensional central simple $k$-algebras form an abelian group with respect to the multiplication induced by the tensor product.*

We still need to check that the collection of classes is actually a set; this is easily seen if we observe that the central simple algebras are finite-dimensional over $k$. ■

The group so obtained is called the *Brauer group* of $k$ and is written $\mathbf{B}_k$; its elements are the *Brauer classes* of central simple $k$-algebras. The Brauer group is an invariant of the field $k$ which provides information about the central division algebras over $k$. Later, in **7.5**, we shall meet another description of $\mathbf{B}_k$, as a cohomology group.

As an example take an algebraically closed field $F$. If $D$ is a division algebra over $F$, then for any $a \in D$, $F(a)$ is a finite extension field, hence $F(a) = F$ because $F$ is algebraically closed, and so $a \in F$, i.e. $D = F$. Thus there are no division algebras over $F$ apart from $F$ itself, and we conclude that $\mathbf{B}_F = 0$, i.e. the Brauer group of an algebraically closed field is trivial. Of course once we drop the restriction on the dimension, we can find skew fields with centre $F$, cf. **9.3** and **11.1**.

For a closer study of $\mathbf{B}_k$ we need to examine the behaviour of algebras under ground field extension. Let $A$ be a (finite-dimensional) $k$-algebra and $F$ an extension field of $k$ (not necessarily finite-dimensional over $k$). Then the $F$-algebra defined by

$$A_F = A \otimes_k F$$

is again finite-dimensional; in fact any $k$-basis of $A$ will be an $F$-basis for $A_F$, so that we have

$$[A_F : F] = [A : k]. \tag{1}$$

We note that $x \mapsto x \otimes 1$ defines an embedding of $A$ in $A_F$. If $A$ has centre $k$, then $A_F$ has centre $F$, by Cor. 1.3. From this fact we can deduce

PROPOSITION 2.2 *If $A$ is a finite-dimensional central simple $k$-algebra, then $[A:k] = r^2$ for some integer $r$, and there is a finite extension $F$ of $k$ such that $A_F \cong F_r$; hence $A$ can be embedded in $F_r$.*

*Proof.* If $E$ is an algebraic closure of $k$, then $A_E$ is a full matrix algebra over $E$, say $A_E \cong E_r$. Comparing dimensions and remembering (1), we find that $[A:k] = [A_E:E] = [E_r:E] = r^2$. Now there is an embedding

$$A \to A_E = E_r, \tag{2}$$

by the above remark. It remains to show that we can replace $E$ by a finite extension. Let $u_1, \ldots, u_n$ be a basis of $A$ and $U_1, \ldots, U_n$ the matrices over $E$ which correspond to the $u$'s under the mapping (2). Then we can express the matrix units in $E_r$ as $e_{ij} = \sum \alpha_{ijv} U_v$ for some $\alpha_{ijv} \in E$. Denote the finite set of entries of the $U_v$ and the $\alpha_{ijv}$ by $X$. Since $E$ is algebraic over $k$, $X$ generates a finite extension $F$ of $k$ and it is clear that $A_F \cong F_r$. Since $A$ is embedded in $A_F$, it is also embedded in $F_r$. ■

We note that this result could not be deduced from Prop. 1.4, because here $F/k$ is never regular. Prop. 2.2 provides no explicit bound on $[F:k]$, but we shall soon meet such a bound, in Cor. 2.7.

For any central simple $k$-algebra $A$ the integer $\sqrt{[A:k]}$ is called the *degree* of $A$, written Deg $A$. If $A \cong D \otimes k_m$, then $\operatorname{Deg} A = m(\operatorname{Deg} D)$; clearly the degree of $D$ is also an invariant of $A$, called the *index* of $A$. It is a measure of how far $A$ deviates from being a full matrix algebra over $k$. We note that the index is an invariant of the Brauer class of $A$, while the degree determines $A$ up to isomorphism within its Brauer class.

If $F$ is an extension field of $k$, then the mapping $A \mapsto A_F$ induces a mapping of Brauer classes, for if $A \cong D \otimes k_m$, then $A_F \cong D \otimes k_m \otimes F \cong D \otimes F_m$, so that $(A_F) = (D_F)$. The mapping is a homomorphism, for $(A \otimes_k B)_F \cong A \otimes_k B \otimes_k F \cong A_F \otimes_F B_F$, so we have a group homomorphism

$$\mathbf{B}_k \to \mathbf{B}_F.$$

Its kernel $\mathbf{B}(F/k)$, the *relative Brauer group*, consists of those classes $(A)$ over $k$ for which $A_F \cong F_m$ for some $m$. Such a class is said to be *split* by $F$ and $F$ is called a *splitting field* for this class, or for the algebra $A$. If $A_F = A \otimes F \cong F_m$, then on taking anti-isomorphisms, we find that $A^0 \otimes F \cong F_m$, hence $F$ splits $A$ iff it splits $A^0$. Any central simple $k$-algebra has a splitting field, which may be taken of finite degree over $k$, by Prop. 2.2.

Next we examine more closely which fields split a given Brauer class, but first we establish a relation between the indices of the extensions:

PROPOSITION 2.3 (Index reduction lemma, A. A. Albert) *Let $F/k$ be a finite field extension, say $[F:k] = r$, and $A$ a central simple $k$-algebra. If $A$, $A_F$ have indices $m, \mu$, then $\mu | m$ and $m/\mu | r$.*

*Proof.* We may take $A = D$ to be a division algebra, without loss of generality. If $C$ is the skew field component of $D_F$, then

$$D \otimes F = D_F \cong C \otimes_F F_q; \tag{3}$$

comparing dimensions, we find $m^2 r = \mu^2 q^2 r$, hence $m = \mu q$ and so $\mu | m$.

Secondly, the regular representation of $F$ (by right multiplication) defines an embedding of $F$ in $k_r$, because $F \cong k^r$ as $k$-space. Thus $F$ is embedded in $k_r$ and hence $D \otimes F$ is embedded in $D \otimes k_r$. By (3) we obtain an embedding of $F_q$ in $D \otimes k_r$. But $F \supseteq k$, so $D \otimes k_r$ contains $k_q$ as central simple subalgebra. If $B$ denotes the centralizer of $k_q$ in $D \otimes k_r$, then by Prop. 1.5,

$$D \otimes k_r \cong k_q \otimes B \cong k_q \otimes G \otimes k_s,$$

where $G$ is the skew field component of $B$ (in fact, $G \cong D$ by uniqueness). Comparing dimensions, we find that $r = qs$, thus $m/\mu = q | r$. ∎

The factor $q$ in $m = \mu q$ is called the *index reduction factor*; we note that $q = r$ whenever $F$ is isomorphic to a subfield of $D$.

The next two corollaries are immediate consequences of Prop. 2.3.

COROLLARY 2.4 *Let $A$ be a central simple $k$-algebra and $F/k$ a field extension. If $[F:k]$ is prime to the index of $A$, then $A$ and $A_F$ have the same index. In particular, if $A$ is a division algebra of degree prime to $[F:k]$, then $A_F$ is again a division algebra.* ∎

COROLLARY 2.5 *The degree of a splitting field (over $k$) of a central simple $k$-algebra $A$ is divisible by the index of $A$.* ∎

We now obtain a criterion for a given finite extension field of $k$ to split a given Brauer class over $k$:

THEOREM 2.6 *Let $w \in \mathbf{B}_k$ and let $F$ be an extension field of $k$, of degree $r$. Then $F$ splits $w$ if and only if some algebra in $w$ contains $F$ as a maximal subfield; this algebra necessarily has degree $r$.*

*Proof.* Let $D$ be the division algebra in the class $w$ and let $n$ be the least integer such that $F$ can be embedded in $D_n$. Then the centralizer $F'$ of $F$ in $D_n$ is a skew field, by the minimality of $n$, and $F$ is the centre of $F'$. By Cor. 1.10,

$$D \otimes_k F \sim D_n \otimes_k F \cong F' \otimes_k k_r \sim F'. \tag{4}$$

Since $F'$ has centre $F$, this shows that $F$ splits $D$ iff $F' = F$, i.e. iff $F$ is a maximal subfield of $D_n$. ∎

COROLLARY 2.7 *Let $D$ be a central division algebra over $k$. Then any maximal subfield of $D$ is a splitting field for $D$.*

For by Cor. 1.11 any maximal subfield $F$ satisfies $[D:k] = [F:k]^2$, so the theorem may be applied. ■

The next step is to show that the splitting field of a central simple algebra can always be taken to be a separable extension of the ground field. We prove more than this, namely that we can actually find a separable splitting field in the skew field component. More precisely, the proof below (due to Herstein) shows that every algebraic skew field extension contains a separable element.

THEOREM 2.8 (Köthe 1932) *Every central division algebra $D$ over $k$ contains a maximal commutative subfield (hence splitting $D$) which is separable over $k$.*

*Proof.* Clearly we may assume that $\operatorname{char} k = p \neq 0$. Our first task will be to find a separable extension $F$ of $k$ in $D$. Each element of $D$ is algebraic over $k$; if some $a \notin k$ is separable over $k$, then $k(a)$ is the required extension. Otherwise there are no separable extensions of $k$, so each element of $D$ is $p$-radical over $k$, $x^{p^r} \in k$ for some $r = r(x)$. Hence we can find $a \notin k$ such that $a^p \in k$. Denote by $\delta$ the inner derivation induced by $a$, $\delta: x \mapsto xa - ax$. Then $x\delta^p = xa^p - a^p x = 0$, but of course $\delta \neq 0$, because $a \notin k$. Choose $b \in D$ such that $c = b\delta \neq 0$, $b\delta^2 = 0$. Then $c\delta = 0 = a\delta$, so if $u = bc^{-1}a$, then $u\delta = cc^{-1}a = a$. Writing this out, we have $ua - au = a$, i.e. $u = 1 + aua^{-1}$, but $u^q \in k$ for some $q = p^s$, hence $u^q = 1 + (aua^{-1})^q = 1 + u^q$ (because $u^q \in k$). We obtain $1 = 0$, a contradiction. It follows that $D$ contains a proper separable extensions of $k$.

Taking a separable extension of maximal degree in $D$, we obtain a maximal separable extension $F$. By Th. 1.9, its centralizer $F'$ is simple with centre $F$, but as centralizer in a division algebra $F'$ is itself a division algebra, so if $F' \neq F$, then $F$ has a proper separable extension $E$, by the first part. But then $E$ is separable over $k$, which contradicts the maximality of $F$. Hence $F' = F$, i.e. $F$ is a maximal subfield and by Cor. 2.7, a splitting field of $D$, separable by construction. ■

We remark that the result holds for any skew fields that are algebraic over $k$ but not necessarily finite-dimensional. For we can use Zorn's lemma instead of a dimension argument to obtain a maximal separable extension.

COROLLARY 2.9 *Every Brauer class of $k$ has a splitting field which is a finite Galois extension of $k$.*

For let $D$ be a division algebra in $w \in \mathbf{B}_k$ and let $F$ be a maximal separable extension in $D$. Then $F$ is contained in a finite Galois extension of $k$, and this will also split $D$. ■

Of course the Galois splitting field need not be contained as a subfield in $D$.

Later in 7.5 we shall find that splitting subfields that are Galois lead to the crossed product construction.

A basic question in the theory of central simple algebras is this: when is the tensor product of two division algebras again a division algebra? In essence this is a question about the index of a tensor product, and little explicit information is available. We first take a simple case, where there is a complete answer.

PROPOSITION 2.10 *If $C$, $D$ are two central division $k$-algebras of coprime degrees, then $C \otimes D$ is again a division algebra.*

*Proof.* By Cor. 1.3, $A = C \otimes D$ is central simple, say $A \cong K_n$ for a skew field $K$. If the simple $A$-module is denoted by $V$, then $A \cong V^n$; here $V$ is isomorphic to a minimal right ideal, hence a $C$-space, and so $n$ divides $[A:C] = [D:k]$. Similarly $n|[C:k]$, hence $n = 1$. ■

Our next result provides a description of the index of a tensor product, even when only one of the factors has finite degree. We remark that if $R$ is a simple Artinian ring and $R_r \cong D_n$, where $D$ is a skew field, then $n = rs$ and $R \cong D_s$. For by Wedderburn's theorem, $R$ has the form $R \cong K_s$ where $K$ is a skew field. It follows that $K_{rs} \cong D_n$ and by uniqueness, $n = rs$ and $K \cong D$, and so $R \cong D_s$.

THEOREM 2.11 *Let $F/k$ be a field extension of degree $d$, and let $C$, $D$ be skew fields with centres $k$, $F$ respectively. If either $C$ or $D$ is of finite degree, then*

$$C \otimes_k D \cong G_m, \tag{5}$$

*where $G$ is a skew field with centre $F$. Moreover,*

(i) *if* $\operatorname{Deg} C = r$, *then*

$$D_q \cong C^0 \otimes_k G, \qquad \text{where } mq = r^2, \tag{6}$$

*and $q$ is the least integer such that $C^0$ can be embedded in $D_q$;*

(ii) *if* $\operatorname{Deg} D = s$, *then $D^0$ can be embedded in $C_p$ for $p = s^2d/m$ but no smaller $p$, the centralizer of $D^0$ in $C_p$ is isomorphic to $G_d$ and if $F'$ denotes the centralizer of $F$ in $C_p$, then*

$$F' \cong D^0 \otimes_F G_d, \qquad \text{and } mp = s^2d. \tag{7}$$

*In particular, in case* (i) $q|r^2$ *and $C \otimes D$ is a skew field if and only if $q = r^2$, while in case* (ii), $p|s^2d$ *and $C \otimes D$ is a skew field if and only if $p = s^2d$.*

(iii) *When both $C, D$ have finite degrees $r, s$ respectively, and $p, q$ are as in* (i), (ii) *then $p, q$ are related by the equation*

$$pr^2 = qs^2d, \tag{8}$$

*and* $\operatorname{Deg} G = t$, *where*

$$t = sq/r = rp/sd, \qquad m = r^2/q = s^2 d/p. \tag{9}$$

From (i), (ii) it is clear that $r, s, d$ are independent, while $p, q$ are related as in (8), and $m, t$ are determined by (9) in terms of $r, s, d, p, q$.

*Proof.* The algebra $C \otimes_k D$ is simple with centre $F$, by Th. 1.2. If $C$ has finite degree, then $C \otimes D$ is finite-dimensional over $D$; if $D$ has finite degree, $C \otimes D$ is finite-dimensional over $C$. In either case it is Artinian and by Wedderburn's theorem it has the form $G_m$ for some $m$, where $G$ is a skew field with centre $F$.

(i) Suppose now that $\operatorname{Deg} C = r$; then $C$ and hence $C^0$ can be embedded in $k_{r^2}$ and hence in $D_{r^2}$. Let $q$ be the least integer such that $C^0$ can be embedded in $D_q$ as $k$-algebra. Then by Prop. 1.5, $D_q \cong C^0 \otimes_k E$, where $E$ is a simple algebra with centre $F$, by Th. 1.2. Moreover, $E$ is Artinian, for if $\mathfrak{a}$ is a left ideal, then $C^0 \otimes \mathfrak{a}$ is a left $D$-space, hence the length of chains of left ideals is bounded by $q$. Thus $E$ is a matrix ring over a skew field. Taking Brauer classes we have $(E) = (C)(D) = (G)$, therefore $E \cong G_n$, but if $n > 1$, then we can replace $q$ by $q/n$, which contradicts the minimality of $q$. Hence $n = 1$ and $D_q \cong C^0 \otimes G$. It follows that $D_{qm} \cong C^0 \otimes G_m \cong C^0 \otimes C \otimes D \cong k_{r^2} \otimes D \cong D_{r^2}$ and so $qm = r^2$; thus (6) is established.

(ii) Next assume that $\operatorname{Deg} D = s$; then $D$ and with it $D^0$ can be embedded in $F_{s^2}$ and hence in $k_{s^2 d}$. Let $p$ be the least integer for which $D^0$ can be embedded in $C_p$ and denote by $H$ the centralizer of $D^0$ in $C_p$. Then by Th. 1.9, $F' \cong D^0 \otimes_F H$, where $F'$ is the centralizer of $F$ and $H$ is simple with centre $F$, while $C_p \otimes_k D \cong H_{s^2}$, by (12) of **7.1**. Comparing this relation with (5), we see that $H_{s^2} \cong G_{mp}$. By Prop. 1.4, $p = [V:C]$, where $V$ is a simple $(D \otimes C)$-module. Since $C \otimes D \cong G_m$, we have $mp = [C \otimes D:C] = [D:k] = s^2 d$. Hence $mp = s^2 d$ and $H \cong G_d$.

Finally, when both $D$ and $C$ have finite degrees, then by combining (6), (7) we see that $m = r^2/q = s^2 d/p$, therefore $pr^2 = qs^2 d$, and if $\operatorname{Deg} G = t$, then a comparison of degrees in (5), (6) yields $t = sq/r = rp/sd$. ■

Consider the special case where $D = F$. Then $s = 1$ and $mp = d$, so we obtain a result essentially contained in Prop. 2.3:

COROLLARY 2.12 *Let $C$ be a skew field with centre $k$ and $F$ a finite extension of $k$. Then $C_F \cong G_m$ for some skew field $G$, where $m|[F:k]$, with equality if and only if $F$ can be embedded in $C$.* ■

We remark that here $C$ need not be finite-dimensional over $k$.

*Examples of Brauer groups.* 1. We have already seen that the Brauer group of an algebraically closed field is trivial, thus e.g. $\mathbf{B}_{\mathbf{C}} = 0$. More generally this holds for any field which is separably closed, by Th. 2.8.

2. The Brauer group of any finite field is trivial. For any division algebra over a finite field $F$ is itself a finite skew field, hence commutative by Th. 1.13, and so it reduces to $F$. This case will be generalized in the next section.

3. The Brauer group of the real numbers has order 2. For the only algebraic extensions of $\mathbf{R}$ are $\mathbf{R}, \mathbf{C}$, so any division algebra has degree 1 or 2, and as we shall see in **7.4**, the only algebra of degree 2 is the algebra of quaternions.

4. If $F$ is a complete field for a discrete valuation with finite residue class field (e.g. the $p$-adic field $\mathbf{Q}_p$), then $\mathbf{B}_F = \mathbf{Q}/\mathbf{Z}$.

5. If $F$ is an algebraic number or function field, then $\mathbf{B}_F$ is a subgroup of the direct sum of the Brauer groups of the corresponding local fields, described in 3, 4 above (cf. Weil [1967] or Reiner [1976]). More precisely we have an exact sequence (Hasse reciprocity):

$$0 \to \mathbf{B}_F \to \oplus \mathbf{B}_{F_p} \to \mathbf{Q}/\mathbf{Z} \to 0,$$

where $F_p$ are the completions of $F$.

### Exercises

(1) Define Brauer classes for any central simple $k$-algebra, not necessarily finite-dimensional, and show that these classes form a monoid whose group of units is $\mathbf{B}_k$.

(2) In Prop. 2.3 show that $q$ is the degree of the largest subfield common to $F$ and $D$.

(3) Show that any skew field $p$-radical over its centre is commutative.

(4) Prove Th. 2.8 in detail for skew fields algebraic over $k$.

(5) Let $D$ be a central division $k$-algebra. For any automorphism $\alpha$ of $D$ as skew field define the *inner order* as the least $r$ such that $\alpha^r$ is inner. Show that the inner order of any such $\alpha$ divides the order of the restriction $\alpha|k$.

## 7.3 The reduced norm and trace

We have, in **5.6** of A.2, met the notions of norm and trace of an element in a finite-dimensional $k$-algebra. We recall that for an $m$-dimensional algebra $A$ with basis $u_1, \ldots, u_m$ the right multiplication is represented by a matrix $\rho(a) = (\rho_{ij}(a))$, where

$$u_i a = \sum \rho_{ij}(a) u_j \qquad \text{for all } a \in A. \tag{1}$$

Here $\rho: A \to \mathfrak{M}_m(k)$ is the regular representation and the trace and norm are defined in terms of it by

$$\mathrm{Tr}(a) = \sum \rho_{ii}(a), \qquad Nm(a) = \det(\rho(a)). \tag{2}$$

When $A$ is central simple over $k$, we have $m = n^2$ and for a splitting field $F$

of $A$ we have

$$A_F \cong F_n, \qquad \text{where } n^2 = [A:k]. \tag{3}$$

Let $e_{ij}$ be the standard basis of matrix units for $F_n$ and write $a \in A_F$ as $a = \sum a_{ij} e_{ij}$. Then the equation (1) takes the form

$$e_{rs} a = \sum_j a_{sj} e_{rj}:$$

hence the matrix $\rho(a)$ has as $(rs, uv)$-entry $(a_{sv}\delta_{ur})$ and the equations (2) become in this case

$$\mathrm{Tr}(a) = \sum_{rs} a_{ss} = n \cdot \sum a_{ss}, \qquad Nm(a) = \det(a_{sv}\delta_{ur}) = (\det(a_{rs}))^n. \tag{4}$$

This is most easily seen by writing $A_F$ as $V^n$, where $V$ is a minimal right ideal, corresponding to a single row of $F_n$. The right action is right multiplication by $a$, so that the matrix $\rho(a)$ is the diagonal sum of $n$ terms $a$, and we obtain (4). In particular, (4) shows that $\mathrm{Tr} = 0$ whenever $n$ is a multiple of $\mathrm{char}\, k$, and it suggests that we can get more information by taking the trace and determinant of the matrix for $a$ itself as invariants.

This is accomplished as follows, for any central simple $k$-algebra. Let $F/k$ be a Galois extension which splits $A$:

$$A \xrightarrow{\lambda} A \otimes_k F \xrightarrow{\mu} F_n,$$

where $\lambda$ is the natural embedding and $\mu$ is an isomorphism. Then $\lambda\mu$ embeds $A$ in $F_n$ and there we have the usual norm and trace. We define the *reduced trace* $T(a)$ and the *reduced norm* $N(a)$ of any $a \in A$ as

$$T_{A/k}(a) = T(a) = \mathrm{Tr}(a\lambda\mu), \qquad N_{A/k}(a) = N(a) = \det(a\lambda\mu) \qquad \text{for } a \in A. \tag{5}$$

We note that whereas $\lambda$ is the canonical mapping $a \mapsto a \otimes 1$, $\mu$ is not uniquely determined, but the definition (5) is independent of the choice of $\mu$, for two isomorphisms of $A_F$ with $F_n$ differ by an automorphism of $F_n$ which must be inner, by the Skolem–Noether theorem, and so leave $T, N$ unaffected. From the definition $T(a)$, $N(a)$ lie in $F$, but if $\sigma$ is a $k$-automorphism of $F$, then $\sigma$ induces a $k$-automorphism of $F_n$, which gives another representation $a \mapsto (a\lambda\mu)^\sigma$. Since $\sigma$ leaves $a \in A_F$ fixed, we have $T(a)^\sigma = T(a)$, $N(a)^\sigma = N(a)$. This holds for all $\sigma \in \mathrm{Gal}(F/k)$, hence $T(a)$, $N(a) \in k$. Further, if $F'$ is another splitting field of $A$, we can find a Galois extension $E$ to contain both $F$ and $F'$, and it follows that $F$ and $F'$ give rise to the same $T$ and $N$. The following familiar properties of trace and norm are easily verified; here $[A:k] = n^2$:

**R.1** $T(a+b) = T(a) + T(b)$, $T(\alpha a) = \alpha T(a)$, $T(1) = n$, *where* $\alpha \in k$,
**R.2** $T(ab) = T(ba)$,

**R.**3 $N(ab) = N(a)N(b)$, $N(\alpha a) = \alpha^n N(a)$, $N(1) = 1$,
**R.**4 $\mathrm{Tr}(a) = n \cdot T(a)$, $Nm(a) = N(a)^n$.

We also have a product formula for the reduced norm and trace. For a field extension $F/k$ we shall write $N_{F/k}$ and $T_{F/k}$ for the usual norm and trace.

PROPOSITION 3.1 *Let $A$ be a central simple $K$-algebra and $B$ a simple subalgebra of $A$, with centre $F$. Suppose that* $\mathrm{Deg}\, A = n$, $\mathrm{Deg}\, B = r$, $[F:k] = t$; *then* $rt \mid n$, *say* $n = rst$, *and for any* $b \in B$,

$$N_{A/k}(b) = N_{F/k}(N_{B/F}(b))^s, \qquad T_{A/k}(b) = s \cdot T_{F/k}(T_{B/F}(b)). \tag{6}$$

*Proof.* Let $E$ be a Galois splitting field of $A$ containing $F$. Then $A_E \cong E_n$ and $E$ also splits $B$, hence under the mapping $A \to A_E$, $B$ becomes

$$B \otimes_k E = (B \otimes_k F) \otimes_F E \cong (B \otimes_F E) \otimes_k F \cong (E \otimes_k F)_r. \tag{7}$$

We thus have an embedding of $(E \otimes_k F)_r$ in $E_n$, so $E_n$ is an $r \times r$ matrix ring, $E_n \cong D_r$, where $D$ is simple Artinian, hence $D \cong E_{n/r}$ by uniqueness.

Since $A_E \cong E_n$, there is a unique simple right $A_E$-module $V \cong E^n$. By (7), $B \otimes E$ is faithfully represented by endomorphisms of $U = (E \otimes F)^r$. Now $[U:E] = r \cdot [E \otimes F : E] = rt$ and $B \otimes E$ acts on $V$, hence $V \cong U^s$ for some $s$, and a comparison of dimensions shows that $n = rst$. For any $b \in B$ we have $b \otimes 1 \in B \otimes_F E \cong E_r$, so $b \otimes 1$ is represented by an $r \times r$ matrix and $T_{B/F}(b) = \mathrm{Tr}(b \otimes 1)$, $N_{B/F}(b) = \det(b \otimes 1)$. If we now tensor with $F$ and consider $b \otimes 1 \otimes 1$ in $(B \otimes_F E) \otimes_k F = (E \otimes_k F)_r$, we have

$$\det(b \otimes 1 \otimes 1) = N_{F/k}(N_{B/F}(b)), \qquad \mathrm{Tr}(b \otimes 1 \otimes 1) = T_{F/k}(T_{B/F}(b)),$$

and since $V \cong U^s$, we obtain (6). ∎

We note the special case $B = F$:

COROLLARY 3.2 *Let $A$ be a central simple $k$-algebra of degree $n$ and $F$ a subfield of $A$, of degree $t$ over $k$. Then $t \mid n$, say $n = st$, and for any $a \in F$,*

$$T_{A/k}(a) = s \cdot T_{F/k}(a), \qquad N_{A/k}(a) = N_{F/k}(a)^s. \quad ∎$$

In particular, when $A$ is a division algebra and $F$ a maximal subfield, then $s = 1$ and the norm and trace in $F$ coincide with the reduced norm and trace in $A$. Further, if in Cor. 3.2 we take $F = k(a)$, then $N_{F/k}(a)$ is (up to sign) the constant term of the minimal polynomial of $a$ over $k$, and $a$ is invertible precisely when this constant term is non-zero. We deduce

PROPOSITION 3.3 *Let $A$ be a central simple algebra. Then an element $a$ of $A$ is invertible if and only if $N(a) \neq 0$.* ∎

Let us denote the group of units of $A$ by $\mathbf{U}(A)$. The reduced norm defines a mapping $\mathbf{U}(A) \to k^\times$ which by **R**.3 above is a homomorphism. Since $k$ is commutative, the commutator subgroup $\mathbf{U}(A)'$ is mapped to 1 in this homomorphism. Let us define the *Whitehead group* of $A$ as

$$\mathbf{K}_1(A) = \mathbf{U}(A)^{ab} = \mathbf{U}(A)/\mathbf{U}(A)'.$$

By what has been said, the reduced norm induces a homomorphism $v: \mathbf{K}_1(A) \to k^\times$. The kernel of this homomorphism is called the *reduced Whitehead group* and denoted by $\mathbf{SK}_1(A)$. We thus have the exact sequence

$$1 \to \mathbf{SK}_1(A) \to \mathbf{K}_1(A) \xrightarrow{v} k^\times \to \operatorname{coker} v \to 1. \tag{8}$$

For a matrix ring over a field, $\mathbf{K}_1$ is easily determined:

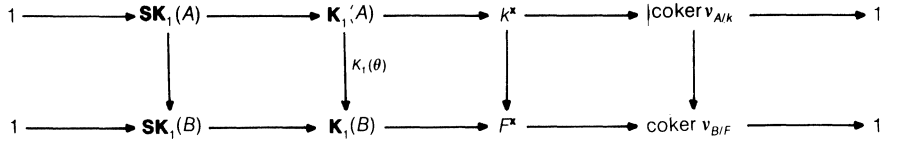
LEMMA 3.4 *For any field $k$ and any $n \geqslant 2$, except $n = 2$ and $k = \mathbf{F}_2, \mathbf{F}_3$,*

$$\mathbf{SK}_1(k_n) = 1 \qquad \text{and} \qquad \mathbf{K}_1(k_n) \cong k^\times. \tag{9}$$

*Proof.* We have a surjective homomorphism $\mathbf{GL}_n(k) \to k^\times$, given by the determinant, with kernel $\mathbf{SL}_n(k)$; hence $\mathbf{SK}_1(k_n) \cong \mathbf{SL}_n(k)/\mathbf{GL}_n(k)'$. Now by Prop. 4.7.2 we have $\mathbf{SL}_n(k) = \mathbf{GL}_n(k)'$ except when $n = 2$ and $k = \mathbf{F}_2, \mathbf{F}_3$, so (9) follows with these exceptions. ■

The exceptions in Lemma 3.4 are treated in Ex. (1), (2). We next describe the diagram resulting from an algebra homomorphism.

THEOREM 3.5 *Let $F/k$ be a field extension and let $A$, $B$ be simple algebras with centres $k$, $F$ respectively. If there is a $k$-algebra homomorphism $\theta: A \to B$, then* $\operatorname{Deg} B = d \cdot \operatorname{Deg} A$ *for some $d \geqslant 1$, and there are homomorphisms such that the diagram*




*commutes, where the mapping $k^\times \to F^\times$ is the power mapping $x \mapsto x^d$.*

*Proof.* It is clear that $\theta$ maps $\mathbf{U}(A)$ to $\mathbf{U}(B)$ and so induces a homomorphism $\mathbf{K}_1(\theta): \mathbf{K}_1(A) \to \mathbf{K}_1(B)$. The $F$-subalgebra of $B$ generated by $A\theta$ is a homomorphic image of $A_F$ and hence, by the simplicity of the latter, isomorphic to $A_F$. If $a \in A$, then $N_{A/k}(a) = N_{A_F/F}(a) = N_{F(a)/F}(a)^r$, where $r \cdot [F(a):F] = \operatorname{Deg} A_F$, and $N_{B/F}(a\theta) = N_{F(a)/F}(a)^s$, where $s \cdot [F(a):F] = \operatorname{Deg} B$. Now $\operatorname{Deg} A_F = \operatorname{Deg} A$ is a divisor of $\operatorname{Deg} B$, because $A_F$ is embedded in $B$. Hence $\operatorname{Deg} B = d \cdot \operatorname{Deg} A$, and so $s \cdot [F(a):F] =$

$d \cdot \mathrm{Deg}\, A = dr \cdot [F(a):F]$, therefore $s = rd$ and $N_{B/F}(a) = N_{A/k}(a)^d$. This shows that the central square in (10) commutes and this determines the outer vertical arrows. ∎

We note the special case when $A$ is any central simple $k$-algebra and $B = A_F$. Then we obtain an exact commutative diagram of the form (10) with $B = A_F$, where the mapping $k^\times \to F^\times$ is the inclusion mapping (because now $d = 1$). In particular, taking $F$ to be a splitting field $A$, we have an isomorphism $\mathbf{K}_1(A_F) \cong F^\times$, by Lemma 3.4 (with the exceptions listed), hence $\mathbf{SK}_1(A_F)$ and coker $v_{A_F/F}$ are then trivial. We remark that any $a \in K$ satisfies $N(a) = a^n$; it follows that coker $v$ is a group of exponent dividing $n$, the index of $A$. It can be shown that $\mathbf{SK}_1(A)$ also has finite exponent, dividing $\prod p_i^{\alpha_i - 1}$, where $\prod p_i^{\alpha_i}$ is the index of $A$ (cf. Draxl [1983]). In fact for many ground fields, e.g. all algebraic number fields, it can be shown that $\mathbf{SK}_1(A) = 1$ and it was an open problem for many years whether (apart from the trivial exceptions of Ex. (1), (2)) algebras with non-trivial reduced Whitehead group exist (Tannaka–Artin problem). In 1975 Platonov gave examples of algebras with non-trivial reduced Whitehead group; we shall meet some simple examples due to Draxl later, in **9.3**.

The reduced norm can be used to show that $\mathbf{B}_k$ is trivial for certain fields $k$. In any central division algebra $A$ of degree $r$ we have $N(x) \neq 0$ for $x \neq 0$, and taking a basis $u_1, \ldots, u_n$ $(n = r^2)$ of $A$, we can write the general element of $A$ as $x = \sum \xi_i u_i$. Now $N(x)$ becomes a *form*, i.e. a homogeneous polynomial of degree $r$ in the $r^2$ variables $\xi_i$. A field $k$ is said to be *quasi-algebraically closed* or a *$C_1$-field* if every form of degree $d$ in $n > d$ variables has a non-trivial zero. With this definition we have

THEOREM 3.6 *Every $C_1$-field has a trivial Brauer group.*

*Proof.* Let $k$ be a $C_1$-field; we have to show that there are no central division algebras other than $k$. Let $D$ be a central division algebra of degree $r$ over $k$. The reduced norm $N(x)$ is a form of degree $r$ in $r^2$ variables, and $N(x) = 0$ has no non-trivial solutions, hence $r^2 \leqslant r$, so $r = 1$ and $D = k$, as claimed. ∎

An obvious example of $C_1$-fields are the algebraically closed fields; we shall soon meet other examples. For the moment we note a reduction that is sometimes useful:

PROPOSITION 3.7 *Any finite extension of a $C_1$-field is again $C_1$.*

*Proof.* Let $F/k$ be an extension of degree $r$ and take a basis $v_1, \ldots, v_r$ of $F$ over $k$. If $f(x_1, \ldots, x_n)$ is a form of degree $d < n$ with coefficients in $F$, let us write $x_\lambda = \sum \xi_{\lambda i} v_i$ and consider

$$g(\xi_{11}, \ldots, \xi_{nr}) = N_{F/k}(f(\textstyle\sum \xi_{1i} v_i, \ldots, \sum \xi_{ni} v_i)).$$

We claim that $g$ is homogeneous of degree $dr$ in the $\xi$s:

$$g(\lambda\xi_{11},\dots,\lambda\xi_{nr}) = N_{F/k}(f(\textstyle\sum\lambda\xi_{1i}v_i,\dots,\sum\lambda\xi_{ni}v_i))$$
$$= N_{F/k}(\lambda^d f(\textstyle\sum\xi_{1i}v_i,\dots,\sum\xi_{ni}v_i)) = \lambda^{dr}N_{F/k}(f).$$

Clearly $g$ has coefficients in $k$, and it is of degree $dr$ in the $nr$ variables $\xi_{\lambda i}$. Since $dr < nr$ and $k$ is $C_1$, it follows that $g(\xi') = 0$ for some $\xi'_{\lambda_i} \in k$, not all 0. This gives $x'_\lambda = \sum \xi'_{\lambda i} v_i \in F$ not all 0, such that $N_{F/k}(f(x')) = 0$, hence $f(x') = 0$, so $f$ has a non-trivial zero in $F$, as claimed. ■

Let us show that every finite field is $C_1$. By Prop. 3.7 we can limit ourselves to $\mathbf{F}_p$, but this is no easier. We shall need a formula for power sums in $\mathbf{F}_q$:

LEMMA 3.8 *Let $k = \mathbf{F}_q$, the field of $q = p^r$ elements. Then*

$$S_m = \sum_{x\in k} x^m = \begin{cases} -1 & \text{if } q-1 \mid m, \\ 0 & \text{otherwise.} \end{cases}$$

*Proof.* We have $x^{q-1} = 1$ for $x \neq 0$ and $= 0$ if $x = 0$, hence $S_{q-1} = \sum x^{q-1} = q - 1 = -1$; similarly $S_{d(q-1)} = -1$. If $q - 1 \nmid m$, then $a^m \neq 1$ for some $a \in k$, hence $S_m = \sum x^m = \sum (ax)^m = a^m S_m$, so $(1 - a^m)S_m = 0$; since $a^m \neq 1$, we conclude that $S_m = 0$. ■

With the help of this formula we can show that the number of points on the hypersurface over $\mathbf{F}_q$ defined by a polynomial in more variables than its degree is divisible by $p$:

THEOREM 3.9 (Chevalley–Warning 1934) *Let $f$ be a polynomial in $n$ variables over $k = \mathbf{F}_q$, where $q = p^r$. Write $V(f)$ for the set of zeros of $f$. If $n$ is greater than the degree of $f$, then*

$$|V(f)| \equiv 0 \pmod{p}. \tag{11}$$

*In particular, if $f$ has zero constant term, then it has a non-trivial zero.*

*Proof.* For each $x \in k^n$ we have

$$1 - f(x)^{q-1} \equiv \begin{cases} 1 & \text{if } x \in V(f), \\ 0 & \text{if } x \notin V(f). \end{cases}$$

Thus $1 - f(x)^{q-1}$ is the characteristic function of $V(f)$, and summing over all points $x$ of $V(f)$, we find

$$|V(f)| \equiv \sum (1 - f(x)^{q-1}) = -\sum f(x)^{q-1},$$

because the total number of points in $k^n$ is $q^n \equiv 0$. Now $f(x)$ is a linear

combination of terms $x_1^{\nu_1} \cdots x_n^{\nu_n}$. We have

$$\sum_x x_1^{\nu_1} \cdots x_n^{\nu_n} = \prod_{i=1}^{n} \sum_{x_i \in k} x_i^{\nu_i} = \prod_{i=1}^{n} S_{\nu_i}. \tag{12}$$

Now if $\nu_i = 0$ for some $i$, then $S_{\nu_i} \equiv 0 \pmod{q}$ and we get zero, so we may assume that $\nu_i > 0$ for $i = 1, \ldots, n$. But by Lemma 3.8, $S_m = 0$ unless $q-1 \mid m$, and since $\sum \nu_i \leqslant d(q-1) < n(q-1)$, where $d$ is the degree of $f$, it follows that some $\nu_i$ is not divisible by $q-1$. So in any case the sum in (12) is zero and (11) follows. ■

If $f(0) = 0$, then the number of non-zero roots of $f = 0$ is $\equiv -1 \pmod{p}$, hence it is non-zero. In particular, this is so if $f$ is homogeneous of positive degree, hence we obtain

COROLLARY 3.10 *Every finite field is* $C_1$. ■

This then shows that $\mathbf{B}_{\mathbf{F}_q} = 0$; in Ex. (4) we shall meet another proof of this fact.

As a third example of $C_1$-fields we consider function fields of transcendence degree 1:

THEOREM 3.11 (Tsen's theorem) *Let $k$ be an algebraically closed field and $F$ a field of functions in one variable over $k$. Then* $\mathbf{B}_F = 0$.

*Proof.* $F$ is a finite algebraic extension of the rational function field $k(t)$. By Prop. 3.7 it will be enough to show that $k(t)$ is a $C_1$-field. Let $f(x_1, \ldots, x_n)$ be a polynomial over $k(t)$, homogeneous of degree $d < n$. We shall show that $f(x) = 0$ has a solution when the $x_i$ are quadratic polynomials in $t$. Write

$$x_i = \xi_{i0} + \xi_{i1} t + \xi_{i2} t^2;$$

then

$$f(x_1, \ldots, x_n) = p_0 + p_1 t + \cdots + p_{2d} t^{2d},$$

where $p_\lambda$ is a polynomial in the $\xi$'s. We thus have $2d+1$ homogeneous polynomials $p_\lambda$ in the $3n$ variables $\xi_{ij}$. We can eliminate one of the $\xi$'s by singling out a polynomial in which it occurs and forming resultants with the remaining polynomials. This diminishes the number of polynomials and variables by one. By continuing this process we eventually obtain a polynomial in at least two variables, provided that the number of variables is greater than the number of polynomials, i.e. $3n > 2d+1$, and this holds because $d < n$. Since $k$ is algebraically closed, our polynomial has a non-trivial zero, and this means that all the $p_\lambda$ have a non-trivial common zero, as we had to show. ■

## Exercises

(1) Show that for $A = \mathfrak{M}_2(\mathbf{F}_2)$, $\mathbf{SK}_1(A) = \mathbf{K}_1(A) = \mathbf{C}_2$, the cyclic group of order 2.

(2) Show that for $A = \mathfrak{M}_2(\mathbf{F}_3)$, $SK_1(A) = \mathbf{C}_3$, $K_1(A) = \mathbf{C}_6$. (Hint: verify that the mapping $\begin{pmatrix} x & 1 \\ -1 & 0 \end{pmatrix} \mapsto x$ extends to a homomorphism $\mathbf{SL}_2(\mathbf{F}_3) \to \mathbf{F}_3^+$; cf. Cohn (1966).)

(3) Let $A$ be a central simple $k$-algebra and let $p$ be a prime dividing the index of $A$. Show that there is a finite extension $F$ of $k$ such that $A_F$ has index $p$.

(4) Use Th. 1.13 to show that every central simple algebra over a finite field $F$ splits, and deduce that $F$ has trivial Brauer group.

## 7.4 Quaternion algebras

The first skew field was discovered by W. R. Hamilton in 1843. His aim had been to find a generalization of the complex numbers, to represent three-dimensional vectors; it took some 12 years to realize that four dimensions rather than three were needed and that the commutative law had to be given up. The algebra he found, which he called the *quaternions*, was a four-dimensional $\mathbf{R}$-algebra $\mathbf{H}$ with basis $1, i, j, k$ and multiplication table:

$$i^2 = j^2 = k^2 = -1, \qquad ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j. \tag{1}$$

The group of order 8 generated by $i, j, k$ is called the *quaternion group*. Hamilton and his followers developed an elaborate geometrical calculus on the basis of the quaternions, but this will not concern us here. For us the quaternions form the simplest division algebra and an important tool in the general theory.

Let $k$ be any field of characteristic not 2 and let $a, b \in k^\times$. We define the *quaternion algebra* $(a, b; k)$ as the $k$-algebra with basis $1, u, v, uv$ and multiplication rules

$$u^2 = a, \qquad v^2 = b, \qquad vu = -uv. \tag{2}$$

In this notation Hamilton's algebra becomes $(-1, -1; \mathbf{R})$. When $\operatorname{char} k = 2$, we define the *quaternion algebra* $(a, b; k]$ as the algebra with basis $1, u, v, uv$ and multiplication rules

$$u^2 = a, \qquad v^2 + v = b, \qquad vu = uv + u. \tag{3}$$

It is easily checked that in either case the quaternion algebra is central simple; hence by Wedderburn's theorem, it is either a division algebra or it is split, i.e. a $2 \times 2$ matrix ring over $k$.

Let $A$ be any quaternion algebra. Then any element $\alpha$ not in $k$ is quadratic over $k$, and its equation may be written

$$\alpha^2 - t(\alpha)\alpha + n(\alpha) = 0, \tag{4}$$

where $t(\alpha)$ and $n(\alpha)$ are the trace and norm respectively. Explicitly, if

$\alpha = t + xu + yv + zuv$, then for $\operatorname{char} k \neq 2$,

$$t(\alpha) = 2t, \qquad n(\alpha) = t^2 - x^2a - y^2b - z^2ab,$$

while for $\operatorname{char} k = 2$ we have

$$t(\alpha) = y, \qquad n(\alpha) = t^2 + x^2a + y^2b + z^2ab + ty + xza.$$

This is most easily seen by observing that $A$ has an *involution*, i.e. an antiautomorphism whose square is 1, $\alpha \mapsto \bar{\alpha}$, such that $t(\alpha) = \alpha + \bar{\alpha}$, $n(\alpha) = \alpha\bar{\alpha}$.

Our first result shows that quaternion algebras effectively include all four-dimensional division algebras:

THEOREM 4.1 *Let $k$ be any field. Then any central simple $k$-algebra $A$ possessing a two-dimensional splitting field is either split or a quaternion algebra.*

*Proof.* Since $A$ has a two-dimensional splitting field, it is four-dimensional by Th. 2.6, and so is either $k_2$ or a division algebra. Leaving the first case aside, we see by Th. 2.8 that the splitting field may be taken to be a subfield of $A$ and separable over $k$. Suppose first that $\operatorname{char} k \neq 2$; then we may take the splitting field to be $F = k(u)$, where $u^2 = a \in k$. Clearly $u \mapsto -u$ defines an automorphism of $F$ which by Cor. 1.7 is induced by an inner automorphism of $A$, say $x \mapsto v^{-1}xv$. Hence $v^{-1}uv = -u$, and $v^2$ centralizes $A$ and so lies in $k$, say $v^2 = b \in k$. Now it is easily verified that $1, u, v, uv$ are linearly independent and clearly span $A$, and they satisfy (2) by construction.

When $\operatorname{char} k = 2$, we can take as our splitting field $F = k(v)$, where $v^2 + v = b \in k$. Now $v \mapsto v + 1$ is an automorphism, hence there exists $u$ such that $u^{-1}vu = v + 1$. Again $u^2$ centralizes $A$, so $u^2 = a \in k$, and as before $1, u, v, uv$ form a basis for $A$, and (3) holds. ■

As a consequence we have

COROLLARY 4.2 (Frobenius' theorem) *The only division algebras over the real field are* **R**, **C** *and* **H**, *the Hamilton quaternions.*

*Proof.* Let $D$ be a division algebra over **R**. Since **C** is the only proper algebraic field extension of **R**, if $D \neq \mathbf{R}, \mathbf{C}$, it must be non-commutative. Let $F$ be a maximal subfield of $D$; $F$ is a proper extension of **R**, hence $F \cong \mathbf{C}$ and by Cor. 2.7, $F$ is a splitting field of $D$. Thus $D$ is a quaternion algebra, $(a, b; \mathbf{R})$ say, by Th. 4.1. If $a$ or $b$ is positive, $D$ is easily seen to split, hence $a, b < 0$ and on dividing the basis elements $u, v$ by $\sqrt{-a}$, $\sqrt{-b}$ respectively we reach the form $(-1, -1; \mathbf{R})$. ■

In general $(a, b; k)$ may be split; conditions for this to happen are given by

PROPOSITION 4.3 *Let $H = (a, b; k)$ be a quaternion algebra over a field $k$ of characteristic not 2. Then the following conditions are equivalent:*

(a) $H = (1, -1; k)$,
(b) *$H$ splits, i.e.* $H \cong k_2$,
(c) *$H$ is not a skew field,*
(d) $n(x) = 0$ *for some* $x \neq 0$ *in* $H$,
(e) $a = N(\alpha)$ *for some* $\alpha \in k(\sqrt{b})$, *where $N$ is the norm from* $k(\sqrt{b})$ *to* $k$,
(f) $ax^2 + by^2 = z^2$ *has a non-zero solution in* $k$.

*Proof.* (a)$\Rightarrow$(b). Consider the map $H \to k_2$ defined by

$$u \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad v \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad uv \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

where $u^2 = 1$, $v^2 = -1$. It is easily checked that this map preserves the defining relations of $H$ and so defines a homomorphism from $H$ to $k_2$. This is clearly surjective and so is an isomorphism, by a comparison of dimensions.

(b)$\Rightarrow$(c) is clear, as is (c)$\Rightarrow$(d), for if $n(x) \neq 0$, then $x$ has an inverse, as we see from (4). (d)$\Rightarrow$(e). If $b$ is a square in $k$, then $k(\sqrt{b}) = k$ and the conclusion follows. Otherwise take $q \neq 0$ with $n(q) = 0$; on writing $q = t + xu + yv + zuv$, we have $0 = n(q) = t^2 - ax^2 - by^2 - abz^2$, hence

$$a = (t^2 - by^2)/(x^2 + bz^2),$$

and this shows $a$ to be a norm in $k(\sqrt{b})$.

(e)$\Rightarrow$(*f*). If $a = N(\alpha) = x^2 - by^2$, then $a + by^2 = x^2$, so (f) holds for $(1, y, x)$. Finally (f)$\Rightarrow$(a): Let $ax^2 + by^2 = z^2$, where $x, y$ are not both zero, say $x \neq 0$. Then

$$\frac{1}{a} + b\left(\frac{y}{ax}\right)^2 = \left(\frac{z}{ax}\right)^2;$$

changing variables, we have $z^2 - by^2 = a^{-1}$. Taking the basis of $H$ to be $1, i, j, k$, put $u = zi + yk$; then $u^2 = az^2 - aby^2 = a(z^2 - by^2) = 1$. Thus $u^2 = 1$; further, we have $ju = -uj$, so if $v = [(1 - b) + (1 + b)u]j/2b$, then $uv = -vu$ and $v^2 = -1$. This shows that $H = (1, -1; k)$. ■

## Exercises

(1) Verify directly that $(1, 1; \mathbf{R})$ is split.

(2) Show that in the Hamilton quaternions the equation $x^2 + 1 = 0$ has infinitely many solutions, which are all conjugate.

(3) Show that two elements of a quaternion algebra satisfying the same irreducible equation either commute or are conjugate. Deduce that every quaternion of norm 1 is a commutator, i.e. $\mathbf{SK}_1(H) = 1$.

(4) Show that in characteristic 2, $(a, b; k|$splits iff $a = N(\alpha)$ for some $\alpha \in k(\wp^{-1}(b))$.

(5) Show that $(a, b; k)$ is multiplicative in each factor, i.e. $(a, b; k) \otimes_k (a', b; k) \cong (aa', b; k)$ and similarly for the other factor. Likewise for $(a, b; k|$ in characteristic 2.

(6) Show that if $H \cong (a, b; k)$ is not split but is split by $k(\sqrt{a'})$, then $H \cong (a', b'; k)$ for suitable $b' \in k$.

(7) (Albert, Draxl) Show that if $(a', b'; k) \otimes_k (a'', b''; k)$ is similar to a quaternion algebra $(a, b; k)$, then there exist $c', c'', d$ such that $(a', b'; k) \cong (c', d; k)$ and $(a'', b''; k) \cong (c'', d; k)$. Deduce that a tensor product of two quaternion algebras $H, K$ is split iff $H, K$ have a common splitting field which is separable quadratic over $k$.

## 7.5 Crossed products

For a closer study of the Brauer group there is a concrete representation of central simple algebras which is often useful, namely as a crossed product. Such a representation does not exist for each algebra, but there is one in every Brauer class.

DEFINITION A central simple $k$-algebra is called a *crossed product* if it contains a maximal subfield $F$ such that $F/k$ is a Galois extension. As maximal subfield $F$ will then be a splitting field (Th. 2.6).

It is easy to see that every Brauer class contains a crossed product: if $D$ is a central division $k$-algebra, then $D$ has a separable splitting field $F$, by Th. 2.8, and the normal closure $E$ of $F/k$ is a Galois extension of $k$. Let $[F:k] = r$, $[E:F] = n$; then $E \subseteq D_n$ and $[E:k] = nr$, $[D_n:k] = n^2r^2$, hence $E$ is a maximal subfield of $D_n$, by Cor. 1.11. Thus $D_n$ is a crossed product, though $D$ itself need not be (because its maximal subfield may not be Galois over $k$). The situation was first studied in the 1930's by Hasse, Albert and others, and it was found that every central division algebra over $\mathbf{Q}$ is a crossed product, but it was only much later that Amitsur (1972) gave examples of central division algebras that are not crossed products.

Crossed products have an explicit description which is of importance (and which accounts for the name). Let $A$ be a crossed product, with Galois splitting field $F$ over $k$ as subfield. Denote by $U$ the group of units of $A$ and by $N$ the normalizer of $F^\times$ in $U$:

$$N = \{u \in U | u^{-1} F u \subseteq F\}.$$

Then $F^\times$ is a normal subgroup of $N$ and we have the exact sequence

$$1 \to F^\times \to N \to \Gamma \to 1,$$

where $\Gamma = N/F^\times$. Thus $A$ determines a group extension $N$ of $F^\times$ by $\Gamma$. We shall

show that (i) $\Gamma \cong \mathrm{Gal}(F/k)$, (ii) every extension of $F^\times$ by $\mathrm{Gal}(F/k)$ determines a crossed product (up to isomorphism). This will also provide an explicit form for $A$.

Any $u \in N$ defines an automorphism $\alpha(u)$ of $F^\times$ by the rule $x^{\alpha(u)} = u^{-1}xu$; this automorphism leaves $k$ elementwise fixed, so we have a mapping

$$\alpha: N \to \mathrm{Gal}(F/k), \tag{1}$$

which is clearly a homomorphism. Its kernel is the centralizer of $F^\times$ in $N$, which is $F^\times$, because $F$ is a maximal subfield. To show that $\alpha$ is surjective, let $\sigma \in \mathrm{Gal}(F/k)$; then by Skolem–Noether (Cor. 1.7) $\sigma$ is induced by an inner automorphism of $A$, say $\sigma = \alpha(u)$, where $u \in U$. By definition, $u \in N$, so (1) is surjective and this shows that $\Gamma \cong \mathrm{Gal}(F/k)$.

Returning to our crossed product $A$, let us take a transversal $\{u_\sigma\}$ of $\Gamma = \mathrm{Gal}(F/k)$ in $N$, so that the elements of $N$ have the form $u_\sigma a (a \in F^\times, \sigma \in \Gamma)$. We shall indicate the action of $\Gamma$ on $F$ by putting exponents, thus

$$au_\sigma = u_\sigma a^\sigma \qquad (a \in F, \sigma \in \Gamma). \tag{2}$$

Further, we have

$$u_\sigma u_\tau = u_{\sigma\tau} c_{\sigma,\tau} \qquad \text{for some } c_{\sigma,\tau} \in F^\times, \tag{3}$$

where the $c_{\sigma,\tau}$ satisfy the factor set condition (by the associativity of $N$):

$$c_{\rho,\sigma\tau} c_{\sigma,\tau} = c_{\rho\sigma,\tau} c_{\rho,\sigma}^\tau. \tag{4}$$

We assert that $A$ is determined completely as right $F$-space with basis $u_\sigma (\sigma \in \Gamma)$ and multiplication rule (2), (3). We know that $[A:k] = n^2 = [A:F][F:k]$ and $[F:k] = n$, hence $[A:F] = n$, so the dimension is correct, since there are $n = |\Gamma|$ basis elements. It only remains to show that the $u_\sigma$ are right linearly independent over $F$. If there is a non-trivial relation

$$\sum u_\sigma a_\sigma = 0, \qquad \text{where } a_\sigma \in F, \tag{5}$$

let us take such a relation with the fewest non-zero coefficients. Pick $\sigma_0 \in \Gamma$ such that $a_{\sigma_0} \neq 0$ and multiply on the left by $u_{\sigma_0}^{-1}$ so as to obtain a relation (5) with $a_1 \neq 0$. The left-hand side of (5) cannot consist of a single term, hence $a_\tau \neq 0$ for some $\tau \neq 1$. Let $b \in F$ be such that $b^\tau \neq b$ and take the commutator of (5) with $b$:

$$0 = \sum u_\sigma b^\sigma a_\sigma - \sum u_\sigma a_\sigma b = \sum u_\sigma a_\sigma (b^\sigma - b).$$

The coefficient of $u_\tau$ is $a_\tau(b^\tau - b) \neq 0$, so this relation is non-trivial, but it has fewer terms than (5), because the coefficient of $u_1$ is $a_1(b - b) = 0$. This contradicts the minimality of (5) and it shows that the $u_\sigma$ are right $F$-linearly independent. We note that this is essentially the argument of Dedekind's lemma (Lemma 3.5.1 of A.2).

Suppose now that we are given a (finite) Galois extension $F/k$ with group $\Gamma$, and a group extension $N$ of $F^\times$ by $\Gamma$, where $\Gamma$ acts on $F^\times$ by automorphisms.

Let us take a transversal $\{u_\sigma\}$ of $F^\times$ in $N$; this determines a factor set for which (3) holds. We define an algebra $A$ by taking the right $F$-space on the $u_\sigma$ as basis, with multiplication defined by (2), (3). Then we claim that $A$ is a crossed product.

In the first place, $A$ is simple, for if $\bar{A}$ is a non-zero quotient, it is spanned by the $\bar{u}_\sigma (\sigma \in \Gamma)$ over $F$ and $\bar{u}_\sigma \neq 0$ because $u_\sigma$ is a unit in $A$ and so cannot map to 0. Now the same argument as before shows that the $\bar{u}_\sigma$ are linearly independent over $F$, hence the mapping $\sum u_\sigma a_\sigma \mapsto \sum \bar{u}_\sigma a_\sigma$ is injective and $A$ is simple.

Next we note that $A$ has centre $k$. For if $x = \sum u_\sigma a_\sigma$ lies in the centre, then $bx = xb$ for all $b \in F$, so $\sum u_\sigma a_\sigma (b^\sigma - b) = 0$. Hence $a_\sigma (b^\sigma - b) = 0$ for all $b \in F$ and $\sigma \in \Gamma$, therefore $a_\sigma = 0$ for $\sigma \neq 1$ and $x = u_1 a \in F$. Now $u_\sigma x = x u_\sigma = u_\sigma x^\sigma$, hence $x^\sigma = x$ for all $\sigma \in \Gamma$, and so $x \in k$. Thus $k$ is the centre of $A$. Finally $[A:F] = [F:k]$ by construction, so $F$ is a splitting field of $A$. This proves

THEOREM 5.1 *Any crossed product $A$ over $k$ with Galois splitting field $F$ contained in $A$ is defined up to isomorphism by an extension $N$ of $F^\times$ by* $\mathrm{Gal}(F/k)$ *and conversely, any such extension $N$ defines a crossed product.* ■

We now examine when two factor sets define isomorphic crossed products. Identifying our two isomorphic algebras, we have to compare two transversals $\{u_\sigma\}$, $\{u'_\sigma\}$ in our crossed product $A$; two such transversals are related by equations

$$u'_\sigma = u_\sigma a_\sigma \qquad \text{for some } a_\sigma \in F^\times,$$

and if the factor set for $\{u'_\sigma\}$ is $\{c'\} = \{c'_{\sigma,\tau}\}$, then

$$c'_{\sigma,\tau} = c_{\sigma,\tau} a_{\sigma\tau}^{-1} a_\sigma^\tau a_\tau;$$

hence the factor sets $c$, $c'$ are associated (cf. (13) of **4.1**). The factor sets form a group $C$ under multiplication, the group of 2-cocycles, in which the bounding cocycles form a subgroup $B$. These are the cocycles associated to 1:

$$c_{\sigma,\tau} = a_{\sigma\tau}^{-1} a_\sigma^\tau a_\tau.$$

The quotient $C/B$ is just $H^2(\Gamma, F^\times)$, the second cohomology group of $\Gamma$ with coefficients in $F^\times$, and by Th. 5.1 we have a mapping

$$H^2(\Gamma, F^\times) \mapsto \mathbf{B}_k. \tag{6}$$

The above remarks show this mapping to be injective and its image is $\mathbf{B}(F/k)$, the subgroup of Brauer classes split by $F$, already encountered in **7.2**. Since each central simple $k$-algebra has a separable splitting field, contained in a Galois extension of $k$, it follows that $\mathbf{B}_k$ is the union of the $\mathbf{B}(F/k)$, as $F$ ranges over the finite Galois extensions of $k$.

It remains to show that (6) is a homomorphism. To establish this fact we need to show that the tensor product of algebras corresponds to the Baer product of the extensions. Take $w \in \mathbf{B}_k$, with Galois splitting field $F$ and let $B \in w$.

Put $[F:k] = n, [B:k] = r^2$ and let $V$ be an $F$-space of dimension $r$; then

$$B^0 \otimes F \cong B_F^0 \cong F_r \cong \mathrm{End}_F(V).$$

Let $A$ be the centralizer of $B^0$ in $\mathrm{End}_k(V)$; then $A \sim B$ and $[A:k] = n^2$; since $A$ contains $F$, we have $[A:F] = n$, so $F$ is a maximal subfield and $A$ is a crossed product. We can realize this situation by taking $V$ to be a simple left ideal in $B_F$; then $V$ is a $(B, F)$-bimodule, i.e. a right $(B^0 \otimes F)$-module, and $[V:F] = r$. Moreover, $A$ has a right $F$-basis $\{u_\sigma\}$ and each $u_\sigma$ defines an $F$-semilinear transformation with automorphism $\sigma$:

$$(\alpha x + \beta y)u_\sigma = \alpha^\sigma x u_\sigma + \beta^\sigma y u_\sigma \qquad \text{for any } x, y \in V,\ \alpha, \beta \in F.$$

Thus $A$ is spanned over $F$ by semilinear transformations.

Given two Brauer classes $w, w'$, let us take $B \in w$, $B' \in w'$ and let $V$, $V'$ be simple left ideals in $B_F$, $B'_F$ respectively, where $F$ is a Galois splitting field for $w$ and $w'$. Then $B_F^0 = \mathrm{End}_F(V)$, $B_F'^0 = \mathrm{End}_F(V')$, hence

$$\mathrm{End}_F(V \otimes_F V') \cong B_F^0 \otimes_F B_F'^0 \cong (B^0 \otimes_k B'^0)_F \cong (B \otimes_k B')_F^0.$$

Denote the respective centralizers of $B^0$, $B'^0$, $(B \otimes B')^0$ by $A, A', A''$; then $A \sim B$, $A' \sim B'$, $A'' \sim B \otimes B'$, hence $(A)(A') = (A'')$. Let $N, N', N''$ be the normalizers of $F^\times$ in $A, A', A''$ respectively; to show that (6) is a homomorphism we must prove that $N''$ is just the Baer product of $N$ and $N'$. To find this Baer product, let $N \circ N'$ be the pullback of the mappings $N \to \Gamma$, $N' \to \Gamma$, i.e. the subgroup of $N \times N'$ of elements of the form $(u_\sigma \alpha, u'_\sigma \beta)$, where $\alpha, \beta \in F^\times$, $u_\sigma, u'_\sigma$ are transversals of $\Gamma$ in $N, N'$ respectively and $u_1 = u'_1 = 1$ for simplicity. The set $L = \{(\lambda, \lambda^{-1}) | \lambda \in F^\times\}$ is a normal subgroup of $N \circ N'$ and the elements of $N \circ N'/L$ can uniquely be written as $(u_\sigma, u'_\sigma \alpha)$ and it is easily verified that this is an extension of $F^\times$ by $\Gamma$ with factor set equal to the product of those of $N$ and $N'$, thus it is the Baer product.

Now each element of $N \circ N'$ defines a semilinear transformation on $V \otimes V'$:

$$(u_\sigma \alpha, u'_\sigma \alpha') : v \otimes v' \mapsto v(u_\sigma \alpha) \otimes v'(u'_\sigma \alpha').$$

Hence we have a homomorphism $f : N \circ N' \to N''$. Clearly it is surjective and the kernel consists of all $(u_\sigma \alpha, u'_\sigma \alpha')$ inducing the identity, i.e. $\sigma = 1$, $\alpha\alpha' = 1$. Hence $\ker f = L$ and so $N''$ is isomorphic to $N \circ N'/L$, the Baer product; this shows (6) to be a homomorphism. We sum up the result as

THEOREM 5.2 *Let $F/k$ be a finite Galois extension with group $\Gamma$. Then*

$$\mathbf{B}(F/k) \cong H^2(\Gamma, F^\times). \quad \blacksquare \tag{7}$$

Once we have the homomorphism property, the injectivity also follows from the explicit criterion for splitting:

PROPOSITION 5.3 *Let $A$ be a central simple $k$-algebra of degree $n$ which is*

*a crossed product with maximal subfield $F$. Then $A \cong k_n$ if and only if the extension of $F^\times$ by $\mathrm{Gal}(F/k)$ in $A$ splits.*

*Proof.* Write $\Gamma = \mathrm{Gal}(F/k)$. If the extension of $F^\times$ by $\Gamma$ splits, then we can realize it with 1 as a factor set. Put $A = \mathrm{End}_k(F) \cong k_n$. Each $\sigma \in \Gamma$ acts as $k$-linear transformation of $F \cong k^n$, while $F$ itself acts by right multiplication. By Dedekind's lemma the $\sigma$ are linearly independent over $F$, so we obtain an $n$-dimensional $F$-space, i.e. an $n^2$-dimensional $k$-space, which must be all of $A$, by comparing dimensions. Thus $A$ is realized as a crossed product. Conversely, any factor set for the algebra $k_n$ is associated to the trivial factor set (which we saw defines $k_n$) and hence itself corresponds to a split extension. ■

Let us examine the Brauer group in a little more detail.

THEOREM 5.4 *The Brauer group $\mathbf{B}_k$ is a torsion group. More precisely, if $w \in \mathbf{B}_k$ has index $r$, then $w^r = 1$. Hence for an extension $F/k$ of degree $n$, $n \cdot \mathbf{B}(F/k) = 0$.*

*Proof.* Let $w \in \mathbf{B}_k$ have index $r$ and take a Galois splitting field $F$ of $w$, where $[F:k] = n$, say. By Th. 5.2, $w$ corresponds to an element $c$ of $H^2(\Gamma, F^\times)$ and it follows from Prop. 4.1.6 that the order of $c$ and hence of $w$ divides $n$, but we want to get the sharper bound $r$.

Let $A \in w$ be a crossed product with $F$ as maximal subfield and let $V$ be a minimal right ideal of $A$; then $[V:F] = r$ and we can represent $A$ by $F$-linear transformations of $V$. Let $v_1, \ldots, v_r$ be a right $F$-basis of $V$; then for any $a \in A$ we have $v_j a = \sum v_i \alpha_{ij}$, or in matrix form,

$$(v)a = (v)\alpha,$$

where $(v) = (v_1, \ldots, v_r)$ and $\alpha = (\alpha_{ij})$. In particular, if $u_\sigma \mapsto U_\sigma$, we have

$$(v)u_\sigma u_\tau = (v)U_\sigma u_\tau = (v)u_\tau U_\sigma^\tau = (v)U_\tau U_\sigma^\tau,$$
$$(v)u_\sigma u_\tau = (v)u_{\sigma\tau} c_{\sigma,\tau} = (v)U_{\sigma\tau} c_{\sigma,\tau},$$

hence

$$U_{\sigma\tau} c_{\sigma,\tau} = U_\tau U_\sigma^\tau, \tag{8}$$

where the $U_\sigma$ are $r \times r$ matrices over $F$. Now write $d_\sigma = \det U_\sigma$ and take determinants in (8):

$$d_{\sigma\tau} c_{\sigma,\tau}^r = d_\tau d_\sigma^\tau,$$

hence $\{c_{\sigma,\tau}^r\}$ is a splitting factor set, therefore $(A)^r = 1$. ■

If $k$ is a perfect field of characteristic $p$, then it can be shown that the $p$-component of $\mathbf{B}_k$ is trivial (cf. Ex. (2)).

The order of $w$ as an element of $\mathbf{B}_k$ is called its *exponent*. For any central simple algebra $A$, its *exponent* is defined as the exponent of its Brauer class, and Th. 5.4 may be expressed by saying that for any Brauer class, the exponent divides the index. The question naturally arises whether the exponent is always equal to the index. For rational division algebras this is true, but not in general. However, we do have the following connexion:

PROPOSITION 5.5 *For any $w \in \mathbf{B}_k$ the index and exponent have the same prime factors.*

*Proof.* Let $w$ have index $m$ and exponent $t$, so that $t|m$, by Th. 5.4. We have to show that any prime factor $p$ *of* $m$ also divides $t$. Take a Galois splitting field $F$ of $w$, with group $\Gamma$ and let $S$ be a Sylow $p$-subgroup of $\Gamma$ with fixed field $E$. Then $[E:k] = (\Gamma:S) = v$ is prime to $p$, while $|S| = p^\alpha$. Let $A \in w$; the index reduction factor from $A$ to $A_E$ divides $v$ (Prop. 2.3) and so is prime to $p$, hence the index of $A_E$ is still divisible by $p$ and it is enough to show that $p$ also divides the exponent. Now $A_E$ is a central simple $E$-algebra which does not split but which is split by $F$. Since $[F:E] = p^\alpha$, its exponent is a positive power of $p$, as we had to show. ■

This result leads to a remarkable decomposition formula:

THEOREM 5.6 *Any central division algebra $D$ of degree $m = q_1 \cdots q_r$, where the $q_i$ are powers of distinct primes, has the decomposition*

$$D \cong D^{(1)} \otimes \cdots \otimes D^{(r)}, \tag{9}$$

*where $D^{(i)}$ is a central division algebra of degree $q_i$.*

We note that the assertion is that (9) is an isomorphism. The corresponding assertion, with isomorphism replaced by similarity is a trivial consequence of the basis theorem for abelian groups, applied to the cyclic subgroup generated by the Brauer class of $D$.

*Proof.* The class $(D)$ has exponent $n = q_1' \cdots q_r'$, where $q_i' | q_i$ and by Th. 5.5, $q_i' > 1$. By the basis theorem for abelian groups, $(D)$ can be written as a product of classes which are powers of $(D)$ with prime power exponent. Let $D^{(i)}$ be a division algebra similar to a power of $D$ with exponent $q_i'$; then

$$D^{(1)} \otimes \cdots \otimes D^{(r)} \sim D.$$

By Th. 5.5 the $D^{(i)}$ have coprime degrees and by Prop. 2.10, we have a division algebra on the left, hence the two sides are isomorphic. ■

A central simple $k$-algebra is called *primary* if it is not equal to $k$ and it contains no proper central simple $k$-subalgebra. Thus if $A$ is not primary, it is either $k$ or it has a central simple subalgebra $B \neq k, A$. By Prop. 1.5 we have $A = B \otimes B'$, where $B'$ is the centralizer of $B$ in $A$. Bearing in mind Th. 5.6 and the relation $k_{rs} \cong k_r \otimes k_s$, we see that any primary algebra is a division algebra of prime power degree or of the form $k_p$. Thus we obtain

PROPOSITION 5.7 *Every central simple algebra is a tensor product of primary algebras. The primary k-algebras are $k_p$ where p is a prime, and certain division algebras of prime power degree.* ■

A division algebra of prime power degree is not necessarily primary, though this does hold over an algebraic number field.

### Exercises

(1) Let $F/k$ be a finite Galois extension with group $\Gamma$ of order $n$. Show that $F \otimes_k k\Gamma \cong k_n$. (Hint: use a normal basis of $F/k$.)

(2) Let $k$ be a perfect field of characteristic $p \neq 0$ and $D$ a central division $k$-algebra. Show that $\operatorname{Deg} D$ is prime to $p$. (Hint: Use Th. 2.8 to show that $D$ contains no proper extension of degree $p$ of $k$.) Deduce that $\mathbf{B}_k$ has trivial $p$-component.

(3) Show that every division $k$-algebra has a splitting field which is a tensor product of extensions of $k$ of prime power degrees.

(4) Let $G$ be a group whose centre $Z$ is free abelian and of finite index $n$ in $G$. Show, by constructing a suitable crossed product with group $G/Z$, that $G$ can be embedded in a division algebra of degree $n$.

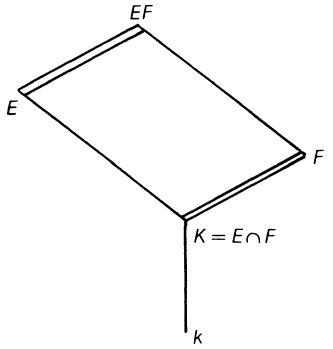
(5) Let $A, B$ be central division $k$-algebras that are crossed products with groups $G, H$. Show that if $A \otimes B$ is a division algebra, then it is a crossed product with group $G \times H$.

## 7.6 Change of base field

Let us consider the effect of changes in the base field on a crossed product. We begin by recalling a result from Galois theory.

PROPOSITION 6.1 *Let $F/k$ be a Galois extension and $E$ any field extension of $k$, where $E$, $F$ are both contained in the same field. Then $EF/E$ is Galois, with group isomorphic to the subgroup of* $\mathrm{Gal}(F/k)$ *corresponding to $E \cap F$.* ■

This is essentially a translation of the parallelogram rule (cf. A.2, Prop. 3.10.2).




The isomorphism is obtained by taking an automorphism of $EF/E$ and restricting it to $F$; this provides an isomorphism with $\mathrm{Gal}(F/E\cap F)$.

In the above situation let us write $G = \mathrm{Gal}(F/k)$ and denote by $H$ the subgroup leaving $E\cap F$ fixed, so that $H \cong \mathrm{Gal}(EF/E)$. Any factor set $\{c\}: G\times G \to F^\times$ gives by restriction to $H$ a factor set $\{c'\}$: $H\times H \to (EF)^\times$. It is clear that a split factor set has a split restriction; hence the inclusion $H\subseteq G$ gives rise to a homomorphism, the *restriction*

$$\mathrm{res}: H^2(G, F^\times) \to H^2(H, (EF)^\times). \tag{1}$$

In what follows we shall write $(F/k, c)$ for the crossed product over $F/k$ with factor set $\{c\}$.

THEOREM 6.2 (Restriction theorem) *Let $F/k$ be a Galois extension with group $G = \mathrm{Gal}(F/k)$, let $E/k$ be any extension (within a field containing $F$) and let $H$ be the subgroup of $G$ corresponding to $E\cap F$, so that $H \cong \mathrm{Gal}(EF/E)$. Then*

$$(F/k, c)_E \sim (EF/E, c'), \tag{2}$$

*where $\{c'\}$ is the factor set $\{c\}$ restricted to $H$:*

$$\begin{array}{ccc} H^2(G,F) & \xrightarrow{\ \mathrm{res}\ } & H^2(H,F) \\ \downarrow & & \downarrow \\ \mathbf{B}(F/k) & \xrightarrow{\ \otimes K\ } & \mathbf{B}(F/K) \end{array}$$

*Proof.* Put $K = E\cap F$; we shall show

$$(F/k, c)\otimes_k K \cong (F/K, c')\otimes k_r, \qquad \text{where } r = [K:k], \tag{3}$$

$$(F/K, c')\otimes_K E \cong (EF/E, c'). \tag{4}$$

It is clear that (2) is a consequence of (3) and (4). Let us write $A = (F/k, c)$; this algebra contains $F$ and hence $K$ as a subfield. If $K'$ denotes the centralizer of

$K$ in $A$, then by Brauer's theorem (Th. 1.9)

$$K' \otimes k_r \cong A \otimes_k K,$$

and this will establish (3) if we can show that $K' \cong (F/K, c')$. In $A$ take $x = \sum u_\sigma a_\sigma (a_\sigma \in F)$; we have $x \in K'$ iff $xy = yx$ for all $y \in K$, i.e. $\sum u_\sigma a_\sigma y = \sum y u_\sigma a_\sigma = \sum u_\sigma y^\sigma a_\sigma$. Thus we must have $a_\sigma = 0$ whenever $y^\sigma \neq y$ for some $y$, i.e. when $\sigma \notin \mathrm{Gal}(F/K) = H$. This shows that $K' = \{\sum u_\sigma a_\sigma | a_\sigma \in F, \sigma \in H\}$ and (3) follows.

To establish (4), we write $B = (F/K, c')$. The $K$-algebra homomorphism

$$E \otimes_K F \to EF$$

is surjective and both sides have dimension $[EF:E] = |H| = [F:K]$ over $E$, so it is an isomorphism (thus the composite $EF$ is independent of the choice of embedding; this depends on $F$ being normal over $K$, cf. Th. 5.5.5). Hence

$$B \otimes_K E = \sum_{\sigma \in H} (u_\sigma \otimes 1) EF \cong (EF/E, c'),$$

and (4) follows. ■

There is a second operation, called *inflation*, corresponding to the natural homomorphism $G \to G/N$, for $N \lhd G$. Given a factor set $\{c\}$ on $\bar{G} = G/N$, we define a factor set $\{\bar{c}\}$ on $G$ by the inflation rule (where $\sigma \mapsto \bar{\sigma}$ is the natural homomorphism from $G$ to $G/N$):

$$\bar{c}_{\sigma,\tau} = c_{\bar{\sigma},\bar{\tau}} \tag{5}$$

THEOREM 6.3 (Inflation theorem) *Let $k \subseteq K \subseteq F$, where $F/k$, $K/k$ are Galois extensions, $G = \mathrm{Gal}(F/k)$ and $N$ is the (normal) subgroup corresponding to $K$. If $\{c\}$ is any factor set on $G/N$ and $\{\bar{c}\}$ is the corresponding factor set on $G$ derived by the inflation rule* (5), *then*

$$(F/k, \bar{c}) \cong (K/k, c) \otimes k_r, \qquad \text{where } r = [F:K]. \tag{6}$$

*Proof.* Let $[F:K] = r$, $[K:k] = s$, $\bar{G} = G/N$, and define

$$B = (K/k, c) \otimes k_r.$$

The field $F$ can be embedded in $K_r$ and hence in $B$; now $B$ is a central simple $k$-algebra split by $K$, hence also by $F$ and $\mathrm{Deg}\, B = rs = [F:k]$, so $F$ is a maximal subfield of $B$, therefore $B$ is a crossed product. We shall prove that $B \cong (F/k, \bar{c})$ by constructing an explicit embedding of $F$ in $B$; this will establish (6).

We take a $K$-basis $e_1, \ldots, e_r$ for $F$ and define $T(x) = (t_{ij}(x))$ by

$$e_i x = \sum t_{ij}(x) e_j \qquad \text{for } x \in F.$$

On writing $e = (e_1, \ldots, e_r)^T$, we can express this equation in matrix form as

$$ex = T(x)e \qquad \text{where } T(x) \in K_r. \tag{7}$$

Since $e_i^\sigma \in F$ for all $\sigma \in G$, we have

$$e^\sigma = P_\sigma e, \qquad \text{where } P_\sigma \in K_r. \tag{8}$$

It follows that $P_{\sigma\tau}e = e^{\sigma\tau} = (P_\sigma e)^\tau = P_\sigma^\tau e^\tau = P_\sigma^\tau P_\tau e$, hence

$$P_{\sigma\tau} = P_\sigma^{\bar\tau} P_\tau, \tag{9}$$

where we have replaced $\tau$ by $\bar\tau$ in the action on $P_\sigma$ because the latter has entries in $K$. Applying $\sigma$ to (7), we have $e^\sigma x^\sigma = T(x)^\sigma e^\sigma$, i.e. $P_\sigma T(x^\sigma) = T(x)^\sigma P_\sigma$ and again $T(x)$ has entries in $K$, so that

$$P_\sigma T(x^\sigma) = T(x)^{\bar\sigma} P_\sigma. \tag{10}$$

We claim that for any right $K$-basis $u_{\bar\sigma}$ of $(K/k, c)$,

$$v_\sigma = u_{\bar\sigma} P_\sigma \tag{11}$$

is a right $F$-basis for $B \cong (K/k, c) \otimes k_r$, with the isomorphism $\sum v_\sigma a_\sigma \leftrightarrow \sum u_{\bar\sigma} P_\sigma T(a_\sigma)$. For the proof we need only verify the conditions on $v_\sigma$: using (10), we have

$$T(x)v_\sigma = T(x)u_{\bar\sigma}P_\sigma = u_{\bar\sigma}T(x)^{\bar\sigma}P_\sigma = u_{\bar\sigma}P_\sigma T(x^\sigma) = v_\sigma T(x^\sigma),$$

and

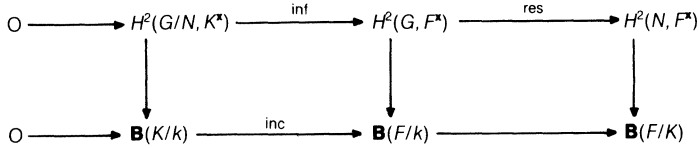
$$v_\sigma v_\tau = u_{\bar\sigma}P_\sigma u_{\bar\tau}P_\tau = u_{\bar\sigma}u_{\bar\tau}P_\sigma^{\bar\tau}P_\tau = u_{\bar\sigma\bar\tau}c_{\bar\sigma,\bar\tau}P_{\sigma\tau} = v_{\sigma\tau}c_{\bar\sigma,\bar\tau}.$$

This shows that $B \cong (F/k, \bar c)$, as claimed, and it proves (6). ■

Since $r > 1$ except in the trivial case $F = K$, we obtain from Th. 6.3,

COROLLARY 6.4 *A central simple algebra obtained by inflation is never a division algebra.* ■

We note that the natural homomorphism $G \to G/N$ induces the inflation homomorphism $H^2(G/N, K^\times) \to H^2(G, F^\times)$ and Th. 6.3 can again be expressed as a commutative square, which together with the previous one gives the commutative diagram:




The bottom row is easily seen to be exact: a central simple $k$-algebra split by $F$ will split as $K$-algebra iff it is split by $K$. Hence the top sequence is also exact.

As a third operation we have the *corestriction* (or *transfer*), which for $k \subseteq F$ provides a homomorphism $\mathbf{B}_F \to \mathbf{B}_k$.

Let $B$ be an $F$-algebra with a finite group $G$ of automorphisms such that each element not equal to 1 of $G$ restricts to a non-trivial automorphism of $F$. As

usual we write $B^G$ and $F^G$ for the subset fixed by $G$. If $A$ is any $k$-algebra and $B = A_F$, where $k = F^G$, then $B^G = A$, as is easily verified.

We begin by showing that $B^G$ can be expressed in terms of the trace, where for $b \in B$ we define $\operatorname{tr} b = \sum b^\sigma$ (summed over all $\sigma \in G$) and $\operatorname{tr} B = \{\operatorname{tr} b \mid b \in B\}$.

LEMMA 6.5 *Let $B$ be an $F$-algebra with a finite group of automorphisms which induce distinct automorphisms on $F$. Then $\operatorname{tr} B$ coincides with the fixed algebra $B^G$ and if $F^G = k$, then*

$$B \cong B^G \otimes_k F. \tag{12}$$

*Proof.* Clearly $\operatorname{tr} B \subseteq B^G$ and both $\operatorname{tr} B, B^G$ are vector spaces over $k$. Let $C$ be the $F$-space spanned by $\operatorname{tr} B$; if $C \subset B$, then there is an $F$-linear functional $\varphi: B \to F$ such that $\varphi(C) = 0$ but $\varphi \neq 0$, say $\varphi(u) \neq 0$. For any $\alpha \in F$ we have

$$0 = \varphi(\operatorname{tr}(\alpha u)) = \varphi(\sum \alpha^\sigma u^\sigma) = \sum \alpha^\sigma \varphi(u^\sigma).$$

By hypothesis the automorphisms $\sigma$ of $F$ are distinct and hence by Dedekind's lemma they are linearly independent, contradicting the assumption that $\varphi(u) \neq 0$. Hence $\varphi$ must vanish and $C = B$.

We thus have a canonical map $B^G \otimes F \to B$, which we assert is injective. Any element of $B^G \otimes F$ can be written as $x = \sum b_i \otimes \alpha_i$, where $b_i \in B^G$, $\alpha_i \in F$ and we may take the $b_i$ to be linearly independent over $k$. Suppose that $x \neq 0$, but that $\sum b_i \alpha_i = 0$; then for any $\beta \in F$,

$$0 = \operatorname{tr}(\beta \sum b_i \alpha_i) = \operatorname{tr}(\sum \beta \alpha_i b_i) = \sum \operatorname{tr}(\beta \alpha_i) b_i.$$

By the linear independence of the $b_i$ we have $\operatorname{tr}(\beta \alpha_i) = 0$ for all $\beta \in F$, hence

$$\sum \beta^\sigma \alpha_i^\sigma = 0 \qquad \text{for all } \beta \in F,$$

and this again contradicts Dedekind's lemma, unless $\alpha_i = 0$ for all $i$, so $x = \sum b_i \otimes \alpha_i = 0$ and our mapping is injective. By comparing dimensions in (12), we see that this is an isomorphism. ■

If $B$ is simple with centre $F$, then (12) shows that $B^G$ is a central simple $k$-algebra such that $[B:F] = [B^G:k]$.

Let $E/k$ be a Galois extension with group $G$ and let $B$ be an $E$-algebra. We define $B^\sigma$, for $\sigma \in G$, as an $E$-algebra on $B$ with the same ring structure as $B$ but with scalar multiplication

$$\alpha \cdot x = \alpha^\sigma x \qquad \text{for } \alpha \in E,\ x \in B.$$

Given any separable extension $F/k$, let $E/k$ be a Galois extension containing $F/k$, with group $G$, and denote by $H$ the subgroup corresponding to $F$. Let $\sigma_1, \ldots, \sigma_n$, where $n = [F:k]$, be a left transversal of $H$ in $G$: $G = \bigcup H\sigma_i$. Then for

any $F$-algebra $B$ the corestriction from $F$ to $k$ is defined as follows. Put

$$B^{(G:H)} = B^{\sigma_1} \otimes_E \cdots \otimes_E B^{\sigma_n}, \tag{13}$$

and define a $G$-action on $B^{(G:H)}$ by writing, for any $\sigma \in G$, $\sigma_i \sigma = \sigma' \sigma_{i'}$, where $\sigma' \in H$ and $i \mapsto i'$ is a permutation of $1, \ldots, n$ determined by $\sigma$. Now put

$$(b_1 \otimes \alpha_1)^\sigma \otimes \cdots \otimes (b_n \otimes \alpha_n)^\sigma = (b_{1'} \otimes \alpha_{1'}^{\sigma'}) \otimes \cdots \otimes (b_{n'} \otimes \alpha_{n'}^{\sigma'}). \tag{14}$$

To check that this is indeed a $G$-action, let $\sigma_i \sigma = \sigma' \sigma_{i'}, \sigma_{i'} \tau = \tau' \sigma_{i*}$. Then $\sigma_i \sigma = \sigma' \tau' \sigma_{i*}$ and

$$\left( \bigotimes_{i=1}^n (b_i \otimes \alpha_i)^\sigma \right)^\tau = \bigotimes_{i=1}^n (b_{i'} \otimes \alpha_{i'}^{\sigma'})^\tau = \bigotimes_{i=1}^n (b_{i*} \otimes \alpha_{i*}^{\sigma'\tau'}) = \bigotimes_{i=1}^n (b_i \otimes \alpha_i)^{\sigma\tau}$$

Now the *corestriction* of $B$ is the subalgebra of $B^{(G:H)}$ fixed by $G$:

$$\mathrm{cor}_{F/k} B = (B^{(G:H)})^G. \tag{15}$$

For example, taking $B = F$, we have $F^{(G:H)} = \otimes F^{\sigma_i}$, and the fixed ring is $k^{\otimes n} = k$, hence we have

$$\mathrm{cor}_{F/k}(F) = k. \tag{16}$$

PROPOSITION 6.6 *Let $F/k$ be a separable extension of degree $n$ and $B$ a central simple $F$-algebra. Then $\mathrm{cor}_{F/k} B$ is a central simple $k$-algebra which depends only on $F/k$, not on the Galois extension $E$ or the choice of transversal. Moreover the correspondence $\mathrm{cor}_{F/k}$ is a homomorphism from $\mathbf{B}_F$ to $\mathbf{B}_k$.*

*Proof.* If $B$ is any central simple $F$-algebra, then $C = B^{(G:H)}$ is simple with centre $E$, by Cor. 1.3. Hence $C^G$ has centre $k$; now any ideal of $C^G$ gives rise to an ideal of $C$, and the simplicity of the latter shows $C^G$ to be simple.

Now let $N$ be a normal subgroup of $G$ contained in $H$ and let $L$ be the corresponding subfield; thus $F \subseteq L \subseteq E$ and $L/k$ is Galois, with group $G/N$. The transversal $\sigma_1, \ldots, \sigma_n$ of $H$ in $G$ is still a transversal of $\bar{H} = H/N$ in $\bar{G} = G/N$. Suppose that $A = \mathrm{cor}_{F/k}(B)$ is formed as above, going via $E$; going via $L$ we obtain $C = B^{(\bar{G}:\bar{H})}$ and it is clear that $C = (B^{(G:H)})^N$. Therefore $C^{\bar{G}} = ((B^{(G:H)})^N)^{\bar{G}} = (B^{(G:H)})^G$, so we reach the same algebra going via $L$. Given any two Galois extensions $L_1$, $L_2$ of $F$, we can find a Galois extension $E/k$ containing both, and the above argument shows that using $L_1$ or $L_2$ we obtain the same algebra $\mathrm{cor}_{F/k}(B)$ as if we had used $E$, hence all three cases give the same result. Further, suppose that $B = B' \otimes B''$, write $C = B^{(G:H)}$ and define $C'$, $C''$ similarly in terms of $B'$ and $B''$. Then by the associativity and commutativity of the tensor product, $C = C' \otimes C''$, hence $C^G = C'^G \otimes C''^G$ and this shows the corestriction to be a homomorphism. ■

For our next result we note that if $K$ is any commutative ring and $E$ is a

commutative $K$-algebra, then for any $K$-modules $U, V$, writing again $U_E = U \otimes_K E$ etc., we have the $K$-module isomorphism

$$U_E \otimes_E V_E \cong (U \otimes_K V)_E, \tag{17}$$

given by the mapping $(u \otimes \alpha) \otimes (v \otimes \beta) \mapsto (u \otimes v) \otimes \alpha\beta$.

PROPOSITION 6.7 *Let $F/k$ be a separable extension of degree $n$ and let $A$ be a central simple $k$-algebra. Then*

$$\mathrm{cor}_{F/k}(A_F) \cong A^{\otimes n}. \tag{18}$$

*Proof.* Take a Galois extension $E/k$ containing $F$, with group $G$ and subgroup $H$ corresponding to $F$. We have $A_F \otimes_F E \cong A \otimes_k E = A_E$ and $A_E^\sigma = A \otimes E^\sigma$, hence

$$(A_F)^{(G:H)} = \otimes_E (A_E)^{\sigma_i} \cong \otimes_E (A \otimes_k E^{\sigma_i}) \cong A^{\otimes n} \otimes (\otimes E^{\sigma_i});$$

taking fixed subrings and using (16), we obtain (18). ■

We remark that in the isomorphism with cohomology groups, (18) corresponds to the formula $\mathrm{cor} \circ \mathrm{res} = n$.

### Exercises

(1) Let $F/k$ be a finite Galois extension. Show that for any $k$-algebra $A$, $\mathrm{tr}\, A_F = A$.

(2) Given a field $F$ of characteristic $p \neq 0$, an $F$-algebra $B$ and a group $G$ of automorphisms of $B$ inducing distinct automorphisms of $F$, show that $p \nmid |G|$.

(3) Let $A$ be a central simple $k$-algebra of degree $p^r$, where $p$ is prime. Show that $A$ is similar to a crossed product of degree $p^s$, for some $s$. (Hint: take a Galois splitting field and a subfield corresponding to a Sylow $p$-subgroup.)

## 7.7 Cyclic algebras

The simplest crossed products are those with cyclic groups; they are called *cyclic* algebras and can be brought to the following simple form. Let $F/k$ be a cyclic Galois extension of degree $n$, with group $\Gamma$ generated by $\sigma$. We shall choose an $F$-basis for our algebra $A$ as follows: $u_0 = 1$, $u_i = u^i (i = 1, \ldots, n-1)$ where $u$ is an element of $A$ inducing $\sigma$. Since $u^i$ induces $\sigma^i$, we have indeed an $F$-basis; moreover, $u^n = \alpha \in F$ and since $u\alpha^\sigma = \alpha u = u^{n+1} = u\alpha$, we have $\alpha^\sigma = \alpha$, so $\alpha \in k$. The multiplication table for $A$ is

$$u^i u^j = \begin{cases} u^{i+j} & \text{if } i+j < n, \\ \alpha u^{i+j-n} & \text{if } i+j \geqslant n. \end{cases} \tag{1}$$

The element $u$ is called the *canonical generator* of $A$. Thus $A$ is determined up

to isomorphism by $F/k$, $\alpha$ and $\sigma$, and one also writes $A = (\alpha, F/k, \sigma)$. We note that by Prop. 5.3, $(1, F/k, \sigma) \cong k_n$; this algebra can be realized as endomorphism ring of $F = k^n$, e.g. by taking a normal basis in $F$. Then $\sigma$ acts by cyclic permutation of the coordinates.

Our first concern is to find when two presentations give isomorphic algebras:

PROPOSITION 6.1 *Let $F/k$ be a cyclic Galois extension of degree $n$. Two cyclic algebras $(\alpha, F/k, \sigma)$ and $(\beta, F/k, \sigma)$ are isomorphic if and only if $\beta/\alpha = N_{F/k}(c)$ for some $c \in F$. In particular $(\alpha, F/k, \sigma)$ splits if and only if $\alpha = N_{F/k}(c)$ for some $c \in F$.*

This condition is more briefly expressed by saying that $\alpha$(resp. $\beta/\alpha$) is a norm from $F$ to $k$.

*Proof.* Assume that $(\alpha, F/k, \sigma) \cong (\beta, F/k, \sigma)$; then the canonical generators $u, v$ are related by the equation $v = uc$, where $c \in F$. Hence $v^i = (uc)^i = u^i c c^\sigma \cdots c^{\sigma i - 1}$; for $i = n$ we find $v^n = (uc)^n = u^n N(c)$, i.e. $N(c) = \beta/\alpha$. Conversely, if $N(c) = \beta/\alpha$, then the same calculation shows that $(uc)^n = \beta$, so that $uc$ is a canonical generator for the isomorphic algebra $(\beta, F/k, \sigma)$. ■

Suppose that $k$ contains a primitive $n$th root of 1, say $\omega$. Then any cyclic algebra $A$ of degree $n$ over $k$ can be written as follows. The splitting field $F$ is of the form $F = k(v)$, where $v^n = \beta \in k$. Taking $u \in A$ such that $u^{-1} v u = \omega v$, we have $u^n = \alpha \in k$, and so $A$ has the $k$-basis $u^i v^j (i, j = 0, 1, \ldots, n-1)$, with the defining relations

$$u^n = \alpha, \qquad v^n = \beta, \qquad vu = \omega uv. \tag{2}$$

By a *symbol* $(\alpha, \beta; k)_n$ or $(\alpha, \beta)_n$ one understands a cyclic algebra over $k$ with the presentation (2). We note the following consequence of (2):

$$(u + v)^n = \alpha + \beta. \tag{3}$$

For on expansion the left-hand side of (3) is a sum of products of degree $n$. Consider a product $P$ involving $i$ factors $u$ and $n - i$ factors $v$; by moving the last factor to the first place we obtain $\omega^i P$, whether this factor was $u$ or $v$. Now $P$ occurs with all its cyclic conjugates and by what has been said, their sum is $(1 + \omega^i + \cdots + \omega^{i(n-1)})P$ which is 0 except when $i = 0$ or $n$; so the sum reduces to $u^n + v^n$, which is $\alpha + \beta$, as claimed.

By applying Prop. 7.1 to a symbol, we obtain

COROLLARY 7.2 *Let $k$ be a field containing a primitive $n$th root of 1. Then a symbol $(\alpha, \beta; k)_n$ splits if and only if $\alpha$ is a norm from $k(\beta^{1/n})$ to $k$.* ■

Consider next a cyclic algebra $A$ of prime degree $p$ over a field $k$ of characteristic $p$. If $F$ is a Galois splitting field of degree $p$ over $k$, then $F$ contains $v$ such that $\wp v = v^p - v = \beta \in k$ and $u$ such that $u^{-1} v u = v + 1$. Hence $u^p = \alpha \in k$ and so

$A$ has the basis $u^i v^j$ $(i,j = 0,1,\ldots,p-1)$ with the defining relations

$$u^p = \alpha, \qquad v^p - v = \beta, \qquad vu = u(v+1). \tag{4}$$

This algebra is also denoted by $(\alpha, \beta; k|_p$.

We can obtain a more symmetric form by putting $t = u^{-1}v$. Then $tu - ut = u^{-1}vu - v = v + 1 - v = 1$. To find $t^p$ we note that in $k$,

$$\prod(x - i) = x^p - x;$$

further, $vu^{-1} = u^{-1}(v-1)$, therefore $t^p = u^{-1}v \cdots u^{-1}v = u^{-p}(v-(p-1)) \cdots (v-1)v = u^{-p}(v^p - v) = \alpha^{-1}\beta$. Thus $A$ may be defined by $t, u$ with the defining relations

$$t^p = \gamma, \qquad u^p = \alpha, \qquad tu - ut = 1,$$

where we have put $\gamma = \beta/\alpha$.

In 7.2 we saw that every central simple algebra is similar to a crossed product. It can actually be shown that when $k$ has a primitive $m$th root of 1, then every central division $k$-algebra of exponent $m$ is similar to a tensor product of cyclic algebras of degree $m$. This is the content of the Merkuryev–Suslin (1982) theorem, whose proof is beyond the level of this book (cf. Rowen [1988] for an illuminating discussion and a proof of various special cases).

## Exercises

(1) Show that $(\alpha, \beta; k)_n$ splits if it contains an element $x$ such that $1, x, \ldots, x^{n-1}$ are linearly independent and $x^n$ is an $n$th power in $k$.

(2) Show that if $(\alpha, F/k, \sigma)$ is cyclic of degree $n$, then $x^n - \alpha$ is irreducible over $k$ and $D$ contains a maximal subfield generated by a root of $x^n = \alpha$.

(3) Show that $(\alpha, F/k, \sigma) \otimes (\beta, F/k, \sigma) \sim (\alpha\beta, F/k, \sigma)$.

(4) Show that if $(\alpha, F/k, \sigma)$ has degree $n$, then for any $r$ prime to $n$, $(\alpha^r, F/k, \sigma^r) \cong (\alpha, F/k, \sigma)$.

(5) (Wedderburn) Show that a cyclic algebra $(\alpha, F/k, \sigma)$ of degree $n$ is a division algebra if $\alpha^n$ is the least power of $\alpha$ which is a norm.

(6) Let $k \subseteq F \subseteq E$, where $E/k$ is cyclic of degree $n$ and $[E:F] = d$. If $\mathrm{Gal}(E/k)$ is generated by $\sigma$ and $\sigma|F = \bar{\sigma}$, then $(\alpha, F/k, \bar{\sigma}) \sim (\alpha^d, E/k, \sigma)$.

(7) With the notation of Th. 6.2, if $E/k$ is cyclic, then $(\alpha, E/k, \sigma)_F \sim (\alpha, EF/F, \sigma^r)$, where $r = [E \cap F:k]$.

(8) Let $k$ be a field with a primitive $n$th root of 1. Show that if $A = (\alpha, F/k, \sigma)$ is a cyclic division algebra of degree $n^2$, then $A$ can also be represented as a crossed product with group $\mathbf{C}_n \times \mathbf{C}_n$.

## Further exercises on Chapter 7

(1) Let $D$ be a skew field, $\alpha$ an endomorphism and $\delta$ a $(1, \alpha)$-derivation of $D$. Show that if $D$ is finite-dimensional over its centre, then either $\alpha$ or $\delta$ must be inner.

(2) Let $D$ be a skew field with centre $k$ but not algebraic over $k$. Show that $D \otimes k(t)$ is a simple Noetherian ring, not a skew field (and not a full matrix ring over a skew field).

(3) (Albert) Let $D$ be a skew field which is totally ordered (cf. **8.3** below), and suppose that $D$ is algebraic over its centre $k$. Show that the conjugates of any positive element are again positive. Deduce that the sum of the conjugates of any non-zero element cannot be zero, and hence prove that $D$ must be commutative.

(4) (Kharchenko) Let $A$ be a central simple $k$-algebra, of finite degree. By regarding $A$ as right $A^e$-module show that every $k$-linear mapping of $A$ to itself has the form $f: x \mapsto \sum a_i x b_i$ $(a_i, b_i \in A)$. Deduce the existence of a non-constant central polynomial, i.e. a polynomial with values in $k$ (cf. **9.7**).

(5) Let $F/k$ be a field extension. Show that there is an exact sequence

$$0 \to \mathbf{B}(F/k) \to \mathbf{B}_k \xrightarrow{f} \mathbf{B}_F.$$

Identify coker $f$ in case $F/k$ is a Galois extension.

(6) Suppose that in a central division $k$-algebra $D$ of degree $n$ every extension is $p$-radical. Show that $x^n \in k$ for all $x \in D$. By extension to a splitting field obtain a contradiction, and hence give another proof of Th. 2.8 (J.-P. Serre).

(7) Let $A, B$ be crossed products with factor sets $(a)$, $(b)$ respectively. Given a Galois splitting field $F$ for both $A$ and $B$, write $F = k(\theta)$, put $P = A \otimes B$ and $e = \prod (\theta \otimes 1 - 1 \otimes \theta^\sigma)/(\theta - \theta^\sigma) \otimes 1$, where the product is taken over all $\sigma \neq 1$ in $\mathrm{Gal}(F/k)$. Verify that for the minimum polynomial $f$ of $\theta$ over $k$, $f(\theta \otimes 1) = f(\theta) \otimes 1 = 0$ and $(c \otimes 1)e = (1 \otimes c)e$ for all $c \in F$. Hence show that $e$ is idempotent and that $ePe$ is a crossed product with factor set $(a)$ $(b)$.

(8) Let $F = k(\sqrt{c})$ (char $k \neq 2$). Show that there is a central division $k$-algebra with $F$ as maximal subfield iff the form $x^2 + cy^2$ is not universal over $k$. Similarly in characteristic 2, if $F$ is generated over $k$ by a root of $x^2 + x + c = 0$, the same holds iff $x^2 + xy + cy^2$ is not universal.

(9) Show that $\mathbf{SL}_2(\mathbf{F}_3)$ has order 24 but is not $S_4$. (Hint: show that its derived group is the quaternion group. It is known as the *binary tetrahedral* group.)

(10) The Hilbert norm residue symbol $(a, b)_p$ is defined to be 1 or $-1$ according as $(a, b; \mathbf{Q}_p)$ does or does not split, where $\mathbf{Q}_p$ is the $p$-adic field when $p$ is prime and $\mathbf{R}$ when $p = \infty$. Using Prop. 4.3 (e), show that for fixed $b$ and $p$ the $a$ with $(a, b)_p = 1$ form a group under multiplication. (Note that the law of quadratic reciprocity, Ex. (23), Ch. 3

of A.2, may be expressed as $\prod(a,b)_p = 1$, where the product is taken over all primes and $p = \infty$.)

(11) Let $k$ be a field with a primitive $n$th root of 1 and $F \supseteq k$. Show that for any $\alpha \in k$, $\beta \in F$, $\mathrm{cor}_{F/k}(\alpha, \beta; F)_n \sim (\alpha, N_{F/k}\beta; k)_n$.

(12) Let $A$ be a central simple $k$-algebra of index $p^\alpha m$, where $p$ is a prime not dividing $m$ and $\alpha \geqslant 1$. Show that there is a separable extension $F$ of degree prime to $p$ over $k$ such that $A_F$ has index $p^\alpha$.

(13) Show that the Brauer classes split by a cyclic extension $F/k$ of degree $n$ form a group $H^2(\mathbf{C}_n, F^\times) \cong k^\times / N_{F/k}(F^\times)$.

(14) Let $D$ be a central division $k$-algebra of prime degree $p$. If $D$ has a maximal subfield $E$ not its own normalizer in $D^\times$, show that $E/k$ is Galois and deduce that $D$ is a crossed product.

(15) (L. E. Dickson) Let $D$ be a central division $k$-algebra of degree 3. Let $x_1 \in D \setminus k$ have minimal polynomial $(t - x_1)(t - x_2)(t - x_3)$. Find $y \in D^\times$ such that $x_i y = y x_{i+1}$ ($i \bmod 3$) and show that either $k(y)$ or $k(x_1 y)$ is not its own normalizer. Deduce that $D$ is a crossed product. (Hint: try a quadratic polynomial in the $x$'s for $y$.)

(16) Show that every central division algebra of degree 6 is cyclic.

# 8
# Quadratic forms and ordered fields

Both volumes 1 and 2 had a chapter on quadratic forms, but some quite basic matters are still left over. There is the Clifford algebra, a useful means of studying quadratic spaces and the orthogonal groups defined on them, while the Witt ring (the Witt group of A.2, **6.4**, equipped with a ring structure) allows us to consider the collection of all quadratic spaces over a given field at once. The discussion of ordered fields is rounded off by a presentation of the Artin–Schreier theory of formally real fields. Both orthogonal and symplectic groups give rise to families of simple groups which are described, and a final section takes a brief glance at the situation in characteristic 2, which was left aside so far.

## 8.1 The Clifford algebra of a quadratic space

Let $k$ be any field and $V$ a vector space over $k$. We recall that a *quadratic form* on $V$ is a mapping $q: V \to k$ such that

**Q.**1 $q(\alpha x) = \alpha^2 q(x)$ *for all* $x \in V$ *and* $\alpha \in k$,
**Q.**2 $q(x+y) - q(x) - q(y) = B(x,y)$ *is a bilinear form on* $V$.

The pair $(V,q)$ is also called a *quadratic space*. The function $B$ in **Q.**2 is a symmetric bilinear form. Suppose that $\operatorname{char} k \neq 2$, the case that will mainly concern us here. Then we define the bilinear form associated with the quadratic form $q$ by

$$b(x,y) = \tfrac{1}{2}B(x,y). \tag{1}$$

With the help of the symmetric bilinear form $b$ we can reach $q$ by the equation

$$b(x,x) = q(x). \tag{2}$$

Thus we can pass from $q$ to $b$ and back, so in characteristic not 2 quadratic forms and symmetric bilinear forms are equivalent.

Given a quadratic space $(V,q)$, suppose that we have a linear mapping $\lambda: V \to A$ into a $k$-algebra $A$. The mapping $\lambda$ is said to be *admissible* and $A$ is *compatible*

with $\lambda$, if

$$(x\lambda)^2 = q(x) \qquad \text{for all } x \in V. \tag{3}$$

On the left we have a product in $A$ and on the right a scalar multiple of the unit element in $A$. From (3) we find by linearization

$$x\lambda \cdot y\lambda + y\lambda \cdot x\lambda = B(x, y), \tag{4}$$

where $B$ is as in **Q**.2. In particular, if $x$ and $y$ are orthogonal vectors in $V$, then $x\lambda$ and $y\lambda$ anticommute, while by (3) $x\lambda$ is invertible in $A$ if $x$ is anisotropic, i.e. $q(x) \neq 0$. Conversely, if $x\lambda$ has an inverse $c$ in $A$, then $x\lambda \cdot c = 1$, hence $q(x)c^2 = (x\lambda)^2 c^2 = 1$ and so $q(x) \neq 0$. Thus $x\lambda$ is invertible iff $q(x) \neq 0$.

DEFINITION Let $(V, q)$ be a quadratic space over a field $k$. By a *Clifford algebra* for $(V, q)$ one understands a $k$-algebra $C$ with a linear mapping $\mu: V \to C$ which is admissible and such that $C, \mu$ is universal for this property. Thus for any compatible algebra $A$ with map $\lambda: V \to A$ there exists a unique homomorphism $\lambda': C \to A$ such that $\lambda = \mu\lambda'$.

To prove the existence of $C$ we form the tensor algebra on $V$, as a mod 2 graded algebra: $\mathbf{T}(V) = \mathbf{T}_0(V) \oplus \mathbf{T}_1(V)$, where $\mathbf{T}_0, \mathbf{T}_1$ are the components of even and odd degree in $\mathbf{T}(V)$, respectively. Let $I$ be the ideal of $\mathbf{T}(V)$ generated by the elements $x^2 - q(x)(x \in V)$. Since these elements are homogeneous of degree 0 (mod 2), $I$ is a graded ideal and the quotient $C = T/I$ is again mod 2 graded, and it has the required universal property by construction: Given an admissible mapping $\lambda: V \to A$, we can extend $\lambda$ to a homomorphism from $\mathbf{T}(V)$ to $A$ because the former is free; this homomorphism maps $I$ to 0, hence it can be factored via $T/I = C$, and it is unique because it is determined on a generating set of $C$, namely the image in $C$ of $V$. Further, $C = C(V)$ is unique up to isomorphism, as universal object. Let us sum up our findings.

PROPOSITION 1.1 *Every quadratic space $(V, q)$ has a Clifford algebra $C(V)$, defined as the universal algebra compatible with a mapping $\mu_V: V \to C(V)$, and isometric spaces have isomorphic Clifford algebras.*

Only the last part still needs proof. If $f: V \to W$ is an isometry, then $f\mu_W: V \to W \to C(W)$ is admissible, hence there is a unique homomorphism $\varphi: C(V) \to C(W)$ such that $f\mu_W = \mu_V\varphi$. Since $f$ is invertible, so is $\varphi$ and it is the required isomorphism. ■

As we have seen, the Clifford algebra $C$ of $V$ is mod 2 graded; we write $C = C_0 \oplus C_1$ and call the subalgebra $C_0$ the *even* Clifford algebra; the elements of $C_0, C_1$ are called *even* and *odd* respectively.

We shall soon see that the Clifford algebra of an $n$-dimensional space has

dimension $2^n$; for the moment we observe that when $\operatorname{char} k \neq 2$,

$$\dim C \leqslant 2^n, \qquad \text{where } \dim V = n. \tag{5}$$

For we can take an orthogonal basis $e_1, \ldots, e_n$ in $V$ (A.2, Th. 6.2.2). If the image of $e_i$ in $C$ is written $u_i$, then $C$ is generated by $u_1, \ldots, u_n$ and by (3) and (4), $u_i u_j = -u_j u_i (i \neq j), u_i^2 = q(e_i)$. It follows that any product of $u$'s can be arranged in ascending order of suffixes by inserting a scalar coefficient. Thus $C$ is spanned by the elements

$$u_{i_1} \cdots u_{i_r}, \qquad i_1 < \cdots < i_r, \qquad r = 0, 1, \ldots, n. \tag{6}$$

The number of these products is $2^n$ and (5) follows.

We pause to give some examples. We recall that a space with the quadratic form $a_1 x_1^2 + \cdots + a_n x_n^2$ (relative to an orthogonal basis) is denoted by $\langle a_1, \ldots, a_n \rangle$.

1. If $V$ is 1-dimensional, say $V = \langle a \rangle$, then $C(V) = k[x]/(x^2 - a)$. Hence $C_0, C_1$ are both one-dimensional and $C$ has basis $1, u$ with multiplication $u^2 = a$. If $a \neq 0, C$ is either the direct sum of two copies of $k$ or a quadratic extension field, namely $k(\sqrt{a})$, according as $a$ is or is not a square in $k$.

2. Let $V = \langle a, b \rangle$, where $a, b \in k^\times$. Then $C(V)$ is generated by two elements $u, v$ such that $u^2 = a, v^2 = b, uv = -vu$; this is the quaternion algebra $(a, b; k)$, cf. **7.4**.

3. If $V$ has the basis $u, v$ with $q(u) = q(v) = 0$, $b(u, v) = 1$, then $C(V)$ has the basis $1, u, v, uv$ with defining relation $uv + vu = 1$. This is just the full matrix ring $k_2$. We note that this applies even in characteristic 2.

4. If the quadratic form on $V$ is identically zero, then $C(V)$ reduces to the exterior algebra on $V$, cf. **2.4**.

To study Clifford algebras we shall need the notion of a graded product. Given two graded algebras $A = \oplus A_i$, $B = \oplus B_i$, we define their *graded tensor product* $A \hat{\otimes} B$ as the algebra whose underlying vector space is $A \otimes B$, again graded by total degree, with multiplication

$$(a_i \hat{\otimes} b_j)(c_r \hat{\otimes} d_s) = (-1)^{jr} a_i c_r \hat{\otimes} b_j d_s, \tag{7}$$

where the suffixes indicate the degree. In particular, this definition applies to mod 2 graded algebras. It is easily checked that the multiplication so defined is associative and distributive.

We can now describe the structure of Clifford algebras more precisely.

THEOREM 1.2 *Let $V$ be a quadratic space over a field $k$ of characteristic not equal to 2, with an orthogonal basis $e_1, \ldots, e_n$. Then the $2^n$ products*

$$e_{i_1} \cdots e_{i_r}, \qquad i_1 < \cdots < i_r, \qquad r = 0, 1, \ldots, n, \tag{8}$$

*constitute a basis of $C(V)$; in particular, the admissible mapping $\mu: V \to C(V)$ is injective. Thus $\dim C(V) = 2^n$ and if $n > 0$, then $\dim C_0 = \dim C_1 = 2^{n-1}$.*

*If $(V,q),(V',q')$ are two quadratic spaces and $(V\perp V',q+q')$ is their orthogonal sum, then*

$$C(V\perp V')\cong C(V)\hat{\otimes}C(V'). \tag{9}$$

*Proof.* We begin with the last part. The inclusion $V\to V\perp V'$ induces a homomorphism $f:C(V)\to C(V\perp V')$; similarly we have a homomorphism $f':C(V')\to C(V\perp V')$. The images of $f$ and $f'$ anticommute, for if $x\in V,x'\in V'$, then $x\cdot x'=0$ in $V\perp V'$, hence $xf\cdot x'f'+x'f'\cdot xf=0$ in $C(V\perp V')$. Thus the generators anticommute, and by induction on the degree this holds for the subalgebras generated. We therefore have a homomorphism of graded algebras:

$$C(V)\hat{\otimes}C(V')\to C(V\perp V'). \tag{10}$$

To find the inverse, consider the mapping

$$(x,x')\mapsto x\hat{\otimes}1+1\hat{\otimes}x' \tag{11}$$

of $V\perp V'$ into $C(V)\hat{\otimes}C(V')$. We have

$$\begin{aligned}(x\hat{\otimes}1+1\hat{\otimes}x')^2&=x^2\hat{\otimes}1+(x\hat{\otimes}1)(1\hat{\otimes}x')+(1\hat{\otimes}x')(x\hat{\otimes}1)+1\hat{\otimes}x'^2\\&=q(x)+q(x')+x\hat{\otimes}x'-x\hat{\otimes}x'\\&=q(x)+q(x').\end{aligned}$$

The right-hand side represents the quadratic form on $V\perp V'$, hence the algebra $C(V)\hat{\otimes}C(V')$ is compatible with the mapping (11), so by the universal property of $C(V\perp V')$ we obtain a homomorphism inverse to (10). So (9) is proved, as isomorphism of graded algebras.

Let $(V,q)$ be any quadratic space of dimension $n$ and write $V=\langle a_1,\dots,a_n\rangle$. By what has been proved, we have an isomorphism

$$C(V)\cong C(\langle a_1\rangle)\hat{\otimes}\cdots\hat{\otimes}C(\langle a_n\rangle).$$

Now $\dim C\langle a\rangle=2$ by the special case treated earlier, hence the right-hand side has dimension $2^n$. As we saw, the $2^n$ elements (6) span $C(V)$, hence they form a basis; in particular $e_1,\dots,e_n$ are linearly independent, so the admissible mapping $\mu$ is injective. The assertions about $C_0,C_1$ now follow because $C_m$ is spanned by the products (8) with $r\equiv m\pmod 2$. ■

This also provides another proof of Th. 2.4.1 on exterior algebras, which is just the special case where the quadratic form is identically zero. In what follows we shall assume that $V$ is a regular quadratic space, i.e. $V^\perp=0$; the structure of $C(V)$ is then expressible in terms of quaternion algebras.

Let us write $(a,b)$ for the quaternion algebra $(a,b;k)$ and $(a)$ for the Clifford algebra of the space $\langle a\rangle$. Then we have the formulae

$$(a)\hat{\otimes}(b)=(a,b), \tag{12}$$

$$(a_1)\hat{\otimes}(a_2)\hat{\otimes}(a_3) = (-a_1a_2, -a_2a_3)\otimes(-a_1a_2a_3), \tag{13}$$

$$(a_1, a_2)\hat{\otimes}(a_3, a_4) = (-a_1a_2, a_2a_3)\otimes(-a_1a_2a_3, a_4), \tag{14}$$

where on the right we can omit the $\hat{}$ because in each case the first factor is even. The formula (12) is a special case of (9). To prove (13), let $(a_i)$ be spanned by $1, u_i$, where $u_i^2 = a_i$. The $u_i$ anticommute, hence $u_1u_2$ and $u_2u_3$ anticommute and $(u_1u_2)^2 = -a_1a_2$, $(u_2u_3)^2 = -a_2a_3$, therefore $u_1u_2$ and $u_2u_3$ generate a quaternion subalgebra of the left of (13), while $u_1u_2u_3$ generates a subalgebra of the centre. Hence the left of (13) contains the right-hand side, so equality holds since both sides are eight-dimensional. This proves (13); now using (12) and (13), we have

$$\begin{aligned}(a_1, a_2)\hat{\otimes}(a_3, a_4) &= (a_1)\hat{\otimes}(a_2)\hat{\otimes}(a_3)\hat{\otimes}(a_4)\\ &= (-a_1a_2, -a_2a_3)\hat{\otimes}(-a_1a_2a_3)\hat{\otimes}(a_4)\\ &= (-a_1a_2, -a_2a_3)\otimes(-a_1a_2a_3, a_4), \qquad \text{i.e. (14).}\end{aligned}$$

The formulae (12)–(14) can be used to describe the structure of $C(V)$, but we shall need to recall the definition of the determinant of a quadratic space. Given a quadratic space $(V, q)$ with associated bilinear form $b$, take any basis $u_1, \ldots, u_n$ of $V$; then $d = \det(b(u_i, u_j))$ is called the *determinant* of $V$. For example, the determinant of $\langle a_1, \ldots, a_n\rangle$ is $a_1 \cdots a_n$. When the basis of $V$ is changed, $d$ changes only by a square factor, and we shall usually regard it as being determined modulo a square, i.e. as an element of $k^\times/k^{\times 2}$. Sometimes we shall need the *discriminant* $\delta$ of $V$, defined as

$$\delta = (-1)^{n(n-1)/2}d = (-1)^m d, \qquad \text{if } \dim V = 2m \text{ or } 2m+1. \tag{15}$$

We note that the space is regular iff $d \neq 0$, and when $\dim V \equiv 0$ or $1 \pmod 4$, or when $-1$ is a square in $k$, then $\delta = d$; otherwise $\delta = -d$. For a regular quadratic form the Clifford algebra is close to being a simple algebra:

THEOREM 1.3 *Let $(V, q)$ be a regular quadratic space over a field $k$ of characteristic not 2, and denote its discriminant by $\delta$.*

(i) *If $\dim V$ is even, then $C(V)$*
*If $\dim V$ is odd, then $C(V)_0$* } *is simple with centre $k$.*

(ii.a) *Suppose that $\delta$ is not a square in $k$.*

*If $\dim V$ is odd, then $C(V)$*
*If $\dim V$ is even, then $C(V)_0$* } *is simple with centre $k(\sqrt{\delta})$.*

(ii.b) *Suppose that $\delta$ is a square in $k$.*

*If $\dim V$ is odd, then $C(V)$*
*If $\dim V$ is even, then $C(V)_0$* } *is a direct product of two simple algebras with centre $k$.*

The relation between $C, C_0$ and their centres is easily remembered by bearing in mind that the dimension of a central simple algebra is a perfect square. Thus if $\delta$ is not a square in $k$ and $k(\sqrt{\delta}) = F$, then for $\dim V = 2m$, $[C:k] = 2^{2m}$, $[C_0:F] = 2^{2(m-1)}$, and when $\dim V = 2m+1$, then $[C:F] = 2^{2m}$, $[C_0:k] = 2^{2m}$.

*Proof.* Let $V = \langle a_1, \ldots, a_n \rangle$; using (13) we can step by step replace two-dimensional factor by quaternion factors. At the first stage, $C(V) = (a_1) \hat{\otimes} \cdots \hat{\otimes} (a_n) = (-a_1a_2, -a_2a_3) \hat{\otimes} (-a_1a_2a_3) \hat{\otimes} (a_4) \hat{\otimes} \cdots \hat{\otimes} (a_n)$; here the number of two-dimensional factors has been reduced by two, while the first factor is an even quaternion algebra. Suppose that $n = 2m$; then after $m$ such steps we have a tensor product of $m$ quaternion algebras, hence $C(V)$ is then central simple over $k$. If $n = 2m+1$, then after $m$ steps we have a product of $m$ quaternion algebras by $((-1)^m a_1 \cdots a_n) = (\delta)$, but as we have seen, this is $k[x]/(x^2 - \delta)$, which is $k(\sqrt{\delta})$ in case (ii.a) and $k \times k$ in case (ii.b). So we have either a simple algebra with centre $k(\sqrt{\delta})$ or a direct product of two central simple $k$-algebras.

Now consider $C_0 = C(V)_0$. When $n = 2m+1$, $C_0$ is the tensor product of $m$ quaternion algebras, which is central simple over $k$; when $n = 2m$, we have after $m-1$ steps a product of $m-1$ even quaternion algebras by the algebra

$$((-1)^m a_1 \cdots a_{2m-1}) \hat{\otimes} (a_{2m}) \cong ((-1)^m a_1 \cdots a_{2m-1}, a_{2m}).$$

If the basis of this quaternion algebra is $1, u, v, uv$, then $(uv)^2 = (-1)^m a_1 \cdots a_{2m} = \delta$, so the even part of this algebra is again $k[x]/(x^2 - \delta)$, which is $k(\sqrt{\delta})$ in case (ii.a) and $k \times k$ in case (ii.b), so we again have either a central simple $k(\sqrt{\delta})$-algebra or a direct product of two central simple $k$-algebras. ■

Sometimes it is useful to have an explicit expression for the centre of $C(V)$:

PROPOSITION 1.4 *Let $(V, q)$ be as in Th. 1.3. If $\dim V = n$ is odd, then the centre of $C(V)$ is two-dimensional with basis $1, u_1 \cdots u_n$ and no non-zero element anticommutes with all of $C(V)$. If $\dim V$ is even, then the centre of $C(V)$ is $k$ and any anticommuting element is a scalar multiple of $u_1 \cdots u_n$.*

*Proof.* Let $u_1, \ldots, u_n$ be an orthogonal basis of $V$; we first note how an element changes under conjugation by $u_i$:

$$u_i^{-1} u_{i_1} \cdots u_{i_r} u_i = \lambda u_{i_1} \cdots u_{i_r},$$

where $\lambda = -1$ if an odd number of suffixes $i_1, \ldots, i_r$ differ from $i$ and $\lambda = 1$ otherwise. For a given product both possibilities are realized unless $r = 0$ or $n$, and $u_1 \cdots u_n$ commutes with all elements of $C$ for odd $n$ and anticommutes for even $n$. The assertion is an immediate consequence. ■

### Exercises

(1) Show that the graded tensor product of anticommutative algebras is again anticommutative.

(2) Prove Th. 1.2 for infinite-dimensional quadratic spaces.

(3) For any quadratic space $V$, find the relation between $V^\perp$ (the radical of $V$) and the Jacobson radical of $C(V)$.

(4) Show that every Clifford algebra is a graded tensor product of a Clifford algebra on a regular quadratic space and an exterior algebra.

(5) Let $V$ be a quadratic space and $u \in V$ anisotropic. Verify that right multiplication by $u$ defines an isomorphism $C(V)_0 \cong C(u^\perp)$.

(6) Verify that $C\langle 1^{k+2}\rangle \cong C\langle(-1)^k\rangle \otimes C\langle 1^2\rangle$, $C\langle(-1)^{k+2}\rangle \cong C\langle 1^k\rangle \otimes C\langle(-1)^2\rangle$.

## 8.2 The spinor norm

The object of this section is to relate the orthogonal group of a quadratic space to its Clifford algebra; here we shall assume that $\operatorname{char} k \neq 2$. We recall from A.2, **6.3** that for any anisotropic vector $u$ we have an orthogonal transformation

$$\sigma_u : x \mapsto x - \frac{2b(x,u)}{q(u)}u, \tag{1}$$

the *symmetry* with respect to $u$; it reverses vectors along $u$ and preserves vectors orthogonal to $u$, thus it is the reflexion in the hyperplane orthogonal to $u$. Further we recall that in an $n$-dimensional regular quadratic space any orthogonal transformation is a product of at most $n^2$ symmetries (A.2, Cor. 6.3.3).

Let $(V, q)$ be a regular quadratic space, $\mathbf{O}(V)$ its orthogonal group and $C(V)$ its Clifford algebra. Any $\theta \in \mathbf{O}(V)$ can be combined with the canonical inclusion $V \to C(V)$ and by the universal property extends to a unique automorphism of $C(V)$; thus we have a mapping

$$\mathbf{O}(V) \to \operatorname{Aut}(C(V)), \tag{2}$$

which is easily seen to be a group homomorphism. To analyse it further, we recall the Skolem–Noether theorem, Cor. 7.1.7, which coupled with Th. 1.3 tells us that when $\dim V$ is even, every automorphism of $C(V)$ is inner. This suggests a representation of orthogonal transformations by inner automorphisms, but there are two complications. Firstly, the element inducing an inner automorphism is determined only up to a scalar factor, and secondly we have to take the grading into account. To meet the first point we shall start by taking an invertible element of $C(V)$ and use it to define an orthogonal transformation. To allow for the grading one could develop a theory of central simple *graded* algebras

(as in Lam [1980], Ch. 4), but for our purpose the following development will be sufficient.

Let $u$ be an anisotropic vector in $V$. Then $u$ is invertible in $C(V)$ and for any $x \in V$ we have $xu + ux = 2b(x, u)$, hence

$$u^{-1}xu + x = 2b(x,u)u^{-1} = \frac{2b(x,u)}{q(u)}u.$$

Thus the symmetry (1) can be written as

$$x\sigma_u = -u^{-1}xu, \qquad \text{where } x, u \in V, q(u) \neq 0. \tag{3}$$

The anisotropic vectors in $V$ form a group under multiplication in $C(V)$; this group is denoted by $\Gamma(V)$ and is called the *Clifford group*; its elements are sometimes called *versors*. Let $z \in \Gamma(V)$; if $z = v_1 \cdots v_r$ say $(v_i \in V)$, then $\gamma_z = \sigma_{v_1} \cdots \sigma_{v_r}$ is an orthogonal transformation, which by (3) can be written $x\gamma_z = (-1)^r z^{-1}xz$. If $v$ is the automorphism of $C(V)$ extending the orthogonal transformation $x \mapsto -x$ of $V$, then this may be written as

$$x\gamma_z = z^{-1}xz^v, \qquad x \in V,\ z \in \Gamma(V). \tag{4}$$

It is clear that the mapping $z \mapsto \gamma_z$ is a homomorphism of $\Gamma(V)$ into $\mathbf{O}(V)$; since every orthogonal transformation is a product of symmetries (A.2, Cor. 6.3.3), this mapping is actually surjective. To find the kernel, assume that $\gamma_z = 1$. Then $xz^v = zx$ for all $x \in V$; writing $z = z_0 + z_1$, where $z_i \in C_i$, we have $x(z_0 - z_1) = (z_0 + z_1)x$, and on equating odd and even components we see that $xz_0 = z_0x$, $xz_1 = -z_1x$ for all $x \in V$. By Prop. 1.4, $z_1 = 0$, $z_0 \in k$, hence $\ker \gamma = k^\times$. We sum up our conclusions as

THEOREM 2.1 *Let $(V, q)$ be a regular quadratic space* (char $k \neq 2$). *Then the mapping $\gamma$ defined by* (4) *gives rise to an exact sequence*

$$1 \longrightarrow k^\times \longrightarrow \Gamma(V) \xrightarrow{\gamma} \mathbf{O}(V) \longrightarrow 1. \quad \blacksquare$$

This means that each orthogonal transformation determines a versor up to a scalar multiple, and it shows that (2) is an isomorphism, with inverse induced by $\gamma$, once the grading has been allowed for.

To make the correspondence more precise we restrict ourselves to versors of norm 1, which are obtained as follows. The mapping $x \mapsto -x$ of $V$ can also be extended to a unique antiautomorphism of $x \mapsto x^J$ of $C(V)$, called the *fundamental involution*, and the *norm* $N$ on $C(V)$ is then defined as

$$N(x) = xx^J, \qquad \text{for } x \in C(V). \tag{5}$$

It is clear that $N$ defines a group homomorphism from $\Gamma(V)$ to $k^\times$ such that $N(x^v) = N(x)$. The versors $x$ such that $N(x) = 1$ form a subgroup $\Gamma(V)$ called the *reduced Clifford group* and denoted by pin $(V)$, for reasons which will become

clear later. Its image under $\gamma$ is called the *reduced orthogonal group* or *spinor kernel* and will be denoted by $\mathbf{O}'(V)$. To identify the cokernel of the map $\mathbf{O}'(V) \to \mathbf{O}(V)$ we define a homomorphism $\theta: \mathbf{O}(V) \to k^\times/k^{\times 2}$ by the following rule: Any $\alpha \in \mathbf{O}(V)$ may be written as a product of symmetries

$$\alpha = \sigma_{u_1} \cdots \sigma_{u_r}. \tag{6}$$

When (6) holds, we define

$$\theta(\alpha) \equiv q(u_1) \cdots q(u_r) \qquad (\mathrm{mod}\, k^{\times 2}). \tag{7}$$

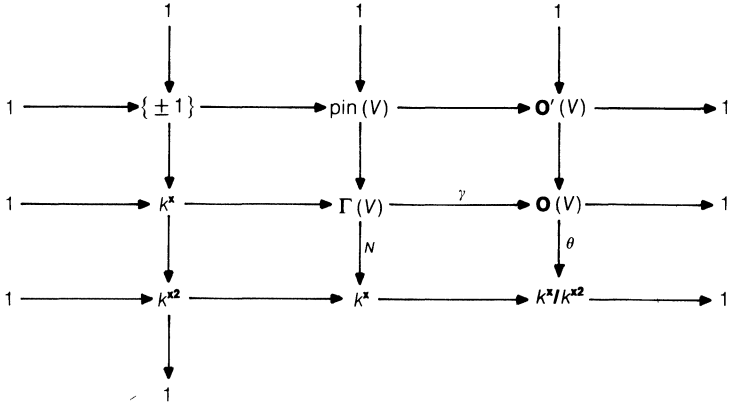
To show that this is well defined we have to verify that when $\sigma_{u_1} \cdots \sigma_{u_r} = 1$, then $q(u_1) \cdots q(u_r)$ is a square in $k$. In this case $r$ is even, because 1 is a rotation, so the element $a = u_1 \cdots u_r$ of $C(V)$ induces the identity on $V$ and so lies in the centre of $C(V)$. Hence $a \in k$ and

$$q(u_1) \cdots q(u_r) = u_1 \cdots u_r u_r \cdots u_1 = aa^J = a^2,$$

which proves the assertion. Thus $\theta(\alpha)$ is a well defined residue class mod $k^{\times 2}$, and it is clear that $\theta(\alpha\beta) = \theta(\alpha)\theta(\beta)$. The resulting homomorphism $\theta$ is called the *spinor norm.*

We remark that $\theta$ is not invariant under a scale change: if we replace $q$ by $\lambda q$, then $\theta(\sigma_u)$ changes by a factor $\lambda$. For this reason $\theta$ is usually restricted to $\mathbf{SO}(V)$; now a scale change $q \mapsto \lambda q$ replaces $\theta(\alpha)$ by $\lambda^r \theta(\alpha)$ and since $r$ is even, this is a square. We also note that $\theta(-1)$ is the determinant of the quadratic form.

The relations between the various groups may be summarized in the following commutative diagram with exact rows and columns:




(8)

We note that the spinor norm is trivial whenever every norm is a square, e.g. for the complex numbers or a positive definite form over the real numbers.

If we restrict ourselves to proper orthogonal transformations and assume that every norm is a square, we obtain the following exact sequence, corresponding to the first row of the diagram (8) and revealing the (whimsical)

origin of pin:

$$1 \to \{\pm 1\} \to \mathrm{Spin}(V) \to \mathbf{SO}(V) \to 1. \tag{9}$$

Here $\mathbf{SO}(V)$ is the special orthogonal group and $\mathrm{Spin}(V)$, called the *spin group*, is the inverse image under $\gamma$ (restricted to $\mathrm{pin}(V)$) of $\mathbf{SO}(V)$. The exactness follows by definition.

As an example consider a regular two-dimensional space $V = \langle a, b \rangle$. Its Clifford algebra is the quaternion algebra $(a, b)$. Specializing to the real quaternions we have $a = b = -1$; the algebra has the basis $1, i, j, ij = k$ and $N(x_0 + x_1 i + x_2 j + x_3 k) = x_0^2 + x_1^2 + x_2^2 + x_3^2$. Thus every norm is a square in $\mathbf{R}$; moreover, $\mathrm{Spin}(V)$ is the group of matrices $P = \begin{pmatrix} x & -\bar{y} \\ y & \bar{x} \end{pmatrix}$, $x, y \in \mathbf{C}$. We have $\det P = x\bar{x} + y\bar{y} = 1$, which is the special unitary group $\mathbf{SU}_2(\mathbf{C})$. Now (9) takes the form

$$1 \to \{\pm 1\} \to \mathbf{SU}_2(\mathbf{C}) \to \mathbf{SO}_3(R) \to 1.$$

This is the double covering of the rotation group in 3-space, also called the *spin representation.*

As a second example consider a *hyperbolic plane.* We recall from A.2, **6.4** that this is a two-dimensional quadratic space $H$ with a basis $u, v$ such that $q(u) = q(v) = 0$, $b(u, v) = 1$. On changing the basis to $e = u + v$, $f = u - v$ we find that $b(e, f) = 0$, $q(e) = -q(f) = 1$; thus $H = \langle 1, -1 \rangle$. Its Clifford algebra has basis $1, u, v, uv$, and is seen to be the matrix ring with matrix units $e_{12} = u$, $e_{21} = v$, $e_{11} = uv$, $e_{22} = vu$. More generally, a hyperbolic space is an orthogonal sum of hyperbolic planes. To find its Clifford algebra we note by (14) of **8.1** that $(-1, 1) \otimes (1, -1) = (1, 1) \otimes (1, -1)$. Thus we can replace the graded by the ungraded tensor product and so find as the Clifford algebra of a $2r$-dimensional hyperbolic space $\mathfrak{M}_{2^r}(k)$.

### Exercises

(1) In any quadratic space $(V, q)$ verify that $\theta(-1)$ is the determinant of $q$.

(2) Show that $\theta: \mathbf{O}(V) \to k^\times / k^{\times 2}$ is the unique homomorphism such that $\theta(\sigma_u) \equiv q(u) \pmod{k^{\times 2}}$ for all anisotropic vectors $u$.

(3) Show that a quaternion $u$ is pure (i.e. of trace 0) iff $u \notin k$, $u^2 \in k$. Verify that every pure quaternion of non-zero norm is invertible and satisfies $x\sigma_u = -uxu^{-1}$ for all pure quaternions $x$.

## 8.3 Formally real fields

There are close links between the theory of ordered fields and quadratic forms, and it is convenient at this point to develop that theory, both for the use it

makes of quadratic forms and because it is needed later. The algebraic theory of ordered fields goes back to a classic series of papers by Artin and Schreier in the 1920s, but there has also been further development in the last 20 years, which has helped to simplify the basic theory.

We recall from A.2, **6.5** that an *ordered ring* is a non-trivial ring $R$ with a total ordering '$>$' satisfying

**O.1** $x > x', y > y' \Rightarrow x + y > x' + y'$,

**O.2** $x > 0, y > 0 \Rightarrow xy > 0$.

As usual we write '$x \geqslant y$' to mean: '$x > y$ or $x = y$' and use $<, \leqslant$ for the opposite ordering. Further, $x$ is called *positive* if $x > 0$ and *negative* if $x < 0$.

A *cone* in a ring $R$ is a subset $P$ containing all squares but not $-1$ and containing with $x, y$ also $x + y, xy$. In an ordered ring $R$ the ordering may be described by its *positive cone*

$$P = \{x \in R \mid x \geqslant 0\}, \tag{1}$$

which has the further property, where as usual, $-P = \{-x \mid x \in P\}$,

$$R = P \cup -P, \qquad P \cap -P = \{0\}. \tag{2}$$

It is easily checked that conversely, any cone satisfying (2) is the positive cone of an ordering on $R$. We note that for a field the second condition in (2) is redundant, for if $P \cap -P$ contains a non-zero element $c$, then $c, -c \in P$, hence $-1 = c \cdot (-c) \cdot (c^{-1})^2 \in P$.

To study the possible orderings of a field it is helpful to consider the elements which are positive under every ordering. For any field $K$ let us define its *core* as the set $C$ of sums of squares:

$$C = C(K) = \{\textstyle\sum a_i^2 \mid a_i \in K\}.$$

We observe that $C$ is closed under addition, multiplication and inversion. For addition this is clear; next if $a = \sum a_i^2$, $b = \sum b_j^2$, then $ab = \sum_{ij} (a_i b_j)^2$ and $a^{-1} = a(a^{-1})^2$. By definition $C$ contains every square in $K$, so it is a cone precisely when $-1 \notin C$.

If $K$ is an ordered field, then clearly all non-zero elements of $C(K)$ are positive, in particular, $-1 \notin C(K)$. A field is said to be *formally real* if its core does not contain $-1$, and so is a cone. By what has been said, every ordered field is formally real. Further, every formally real field is of characteristic 0, for $C$ contains $1, 1+1, 1+1+1, \ldots$ and these must all be different from $-1$. The following rephrasing of the definition is often useful:

PROPOSITION 3.1 *A field is formally real if and only if $-1$ cannot be written as a sum of squares.* ■

Our first task is to show that every formally real field can be ordered. It is no harder to prove a relative version, using a relative notion of core.

Let $K$ be any field with a subfield $F$ and suppose that $F$ is ordered. Then the *$F$-core* of $K$ is the set $\Gamma_F(K)$ of all sums of squares with positive coefficients from $F$:

$$\Gamma_F(K) = \{\sum a_i\lambda_i^2 \mid \lambda_i \in K, a_i \in F, a_i > 0\}.$$

If $K$ is any field of characteristic 0, it contains $\mathbf{Q}$ as prime subfield and this is ordered; now the $\mathbf{Q}$-core of $K$ is just the core defined earlier. As in that case we can show that $\Gamma_F(K)$ is closed under addition, multiplication and inversion and contains all squares. Hence $\Gamma_F(K)$ is a cone iff it does not contain $-1$. If $K$ has an ordering extending that of $F$, then $\Gamma_F(K)$ clearly does not contain $-1$ and so is a cone. We shall show conversely that if $\Gamma_F(K)$ is a cone, then the ordering of $F$ can be extended to $K$. We single out two properties of cones:

LEMMA 3.2 *Let $K$ be any field.* (i) *If $P$ is any cone on $K$ and $a, b \in K$ are such that $ab \in P$, then either $P + aP$ or $P - bP$ is a cone.* (ii) *Every cone on $K$ is contained in a maximal cone and any maximal cone $P$ satisfies $P \cup -P = K$.*

*Proof.* (i) Clearly $P + aP$, $P - bP$ contain $P$ and they admit addition, multiplication and inversion. For example if $x = x_1 - bx_2$, $y = y_1 - by_2$, then $xy = x_1y_1 + b^2x_2y_2 - b(x_1y_2 + x_2y_1)$, and similarly for the other cases. Hence if neither $P + aP$, nor $P - bP$ is a cone, then both contain $-1$, say $x + au = -1 = y - bv$; hence $-abuv = 1 + x + y + xy$ and so $-1 = x + y + xy + abuv \in P$, a contradiction.

(ii) The existence of a maximal cone containing a given cone follows by Zorn's lemma. Now let $P$ be a maximal cone and take $a \in K^\times$. Then $a^2 \in P$, hence by (i), $P + aP$ or $P - aP$ is a cone, say the former; but $P + aP \supseteq P$, so by maximality $P = P + aP$ and thus $a \in P$. ■

Now it is an easy matter to establish a criterion for extendibility:

THEOREM 3.3 *Let $K$ be a field with an ordered subfield $F$. Then the ordering of $F$ can be extended to an ordering of $K$ if and only if the $F$-core $\Gamma_F(K)$ is a cone, i.e. $-1 \notin \Gamma_F(K)$.*

*Proof.* We have seen the necessity of this condition. Conversely, when it holds, then by Lemma 3.2 (ii), $\Gamma_F(K)$ is contained in a maximal cone $P$ and $P \cup -P = K$. It follows that $P$ defines an ordering on $K$ and this ordering extends the ordering on $F$, because $P$ contains $\Gamma_F(K)$, which by construction contains the positive cone of $F$. ■

If we apply this result to a field $K$ of characteristic 0, taking $F$ to be $\mathbf{Q}$, we see that $K$ can be ordered iff it is formally real. More precisely we have

COROLLARY 3.4 *Let $K$ be a formally real field and $P$ any cone on $K$. Then $P$ is the intersection of all maximal cones containing $P$. In particular, $K$ can be ordered.*

*Proof.* It is clear that the intersection of maximal cones containing $P$ is again a cone containing $P$, so to prove the first part we have to find, for any $a \notin P$, a maximal cone containing $P$ but not $a$. We note that $P - aP$ is a cone, for if not, then $-1 = x - ay$ for some $x, y \in P$, and so $ay = x + 1$, hence $a = (x+1)y^{-1} \in P$, which is a contradiction. By Lemma 3.2 there is a maximal cone $Q$ containing $P - aP$; hence $Q \supseteq P$, but $a \notin Q$ because $-a \in Q$. Now it follows that the intersection of all maximal cones containing $P$ contains no element not in $P$ and so equals $P$. The second part follows because a formally real field always has a cone, namely its core. ■

From Cor. 3.4 we see that in any formally real field $K$ the core is just the intersection of all maximal cones. In other words, an element of $K$ is a sum of squares iff it is positive under every ordering of $K$. Such an element is also called *totally positive.*

We next investigate which algebraic extensions of ordered fields allow an extension of the ordering. A field is said to be *real closed* if it is formally real but has no algebraic extensions that are formally real; thus a real closed field is a maximal formally real extension field in its algebraic closure. By the *real closure* of a formally real field $K$ we understand a maximal formally real extension of $K$ in its algebraic closure. For example, the field of all real algebraic numbers is real closed and it is the real closure of $\mathbf{Q}$. It is easy to see that a real closure always exists:

THEOREM 3.5 *Every formally real field has a real closure.*

*Proof.* Let $K$ be formally real; by Cor. 3.4 we may take it to be ordered. Consider the family $\mathscr{F}$ of all ordered extensions of $K$ in its algebraic closure; $\mathscr{F}$ itself is partially ordered by order-preserving inclusions, thus for $E_1, E_2 \in \mathscr{F}$ we write $E_1 \leqslant E_2$ if $E_2$ is an extension of $E_1$ and the ordering of $E_1$ is induced by that of $E_2$. Clearly $\mathscr{F}$ is inductive, and so by Zorn's lemma it has a maximal member $E$. This is the required real closure of $K$. ■

Real closed field can be described more explicitly; to do so we need to look at two particular cases of algebraic extensions:

LEMMA 3.6 *Let $K$ be an ordered field and $L$ an algebraic extension of $K$. Then the ordering of $K$ can be extended to $L$ provided that either $[L:K]$ is odd, or $L = K(\sqrt{a})$, where $a \in K$, $a > 0$.*

*Proof.* Suppose first that $L/K$ is an extension of odd degree $n$. We shall use induction on $n$. Since $\operatorname{char} K = 0$, the extension is simple, by the theorem of the primitive element (A.2, Th. 3.9.2), say $L = K(\alpha)$, where the minimal polynomial $p$ of $\alpha$ over $K$ has degree $n$. For $n = 1$ the result clearly holds, so assume that $n > 1$. By Th. 3.3 we have to show that $\Gamma_K(L)$, the $K$-core of $L$, is a cone. If this is not so, it must contain $-1$, so we have

$$-1 = \sum \lambda_i f_i(\alpha)^2, \tag{3}$$

where $\lambda_i > 0$ in $K$ and each $f_i$ is a polynomial over $K$ of degree at most $n-1$. Since $p$ is the minimal polynomial of $\alpha$, we have the identity

$$-1 = \sum \lambda_i f_i(x)^2 + p(x)q(x), \tag{4}$$

for some polynomial $q$. Now $\sum \lambda_i f_i(x)^2$ has even degree, because the highest term is a sum of squares with positive coefficients and so cannot vanish. Moreover, this degree is positive, for otherwise (3) would give a contradiction, and it cannot exceed $2n-2$, because $\deg f_i \leqslant n-1$. Thus $pq$ has even degree $\leqslant 2n-2$, hence $q$ has odd degree $\leqslant n-2$. It follows that $q$ has an irreducible factor of odd degree $< n$, and if $\beta$ is a zero, then by induction the ordering extends to $K(\beta)$, but on putting $x = \beta$ in (4) we find $-1 = \sum \lambda_i f_i(\beta)^2$, a contradiction. Hence $\Gamma_K(L)$ is a cone and this shows that the ordering of $K$ can be extended to $L$.

Next let $L = K(\sqrt{a})$ and suppose again that $\Gamma_K(L)$ is not a cone. Then

$$-1 = \sum \lambda_i (x_i + y_i \sqrt{a})^2.$$

Multiplying out and observing that $1, \sqrt{a}$ are linearly independent over $K$, we obtain the relation

$$-1 = \sum \lambda_i (x_i^2 + y_i^2 a) > 0,$$

which is a contradiction. Hence $\Gamma_K(L)$ is a cone and the result follows. ■

It is useful at this point to introduce a special type of ordered field. A field $K$ is called *Euclidean* if for each $a \in K^\times$ exactly one of $a, -a$ is a square. Such a field can always be ordered by writing $a \geqslant 0$ iff $a$ is a square; clearly this is the only possible ordering of $K$. Our next result (essentially due to Artin, in a presentation following Scharlau [1985]) shows how Euclidean fields are related to real closed fields.

THEOREM 3.7 *For any field $K$ the following conditions are equivalent:*

(a) *$K$ is real closed,*

(b) *$K$ is Euclidean and every equation over $K$ of odd degree has a root in $K$,*

(c) *$K$ is not algebraically closed but $K(\sqrt{-1})$ is algebraically closed.*

*Proof.* (a)⇒(b). If $K$ is real closed, it is formally real and so can be ordered, by Cor. 3.4. By Lemma 3.6 this ordering can be extended to $K(\sqrt{a})$ for any $a>0$, and since $K(\sqrt{a})=K$ by the maximality of $K$ as formally real field, it follows that $a$ is a square in $K$. Of course every square is $\geqslant 0$, so $K$ is indeed Euclidean, and no equation of odd degree $>1$ can be irreducible, because it would lead to a larger ordered field, by Lemma 3.6. Hence every equation of odd degree $>1$ is reducible over $K$, and by induction on the degree it has a root in $K$; thus (b) holds.

(b)⇒(c). Assume (b); then $-1$ is not a square in $K$, so $K$ is not algebraically closed. Moreover, since $K$ is Euclidean, it can be ordered by writing $a\geqslant 0$ iff $a$ is a square. Now consider $K(\sqrt{-1})$ and write $\mathrm{i}=\sqrt{-1}$; we have to show that $K(\mathrm{i})$ is algebraically closed. In the first place, every element is a square, for if $c\in K$, either $c\geqslant 0$ and $\sqrt{c}\in K$ or $c<0$ and $\sqrt{-c}\in K$. Next, if $c\in K(\mathrm{i})$, say $c=a+b\mathrm{i}$, where $a,b\in K$ and $b\neq 0$, then we have to solve

$$a+b\mathrm{i}=(u+v\mathrm{i})^2=u^2-v^2+2uv\mathrm{i}. \tag{5}$$

By (5) we have $a^2+b^2=(u^2+v^2)^2$; now $a^2+b^2>0$ in $K$, hence it is a square in $K$, say $a^2+b^2=d^2$, and it remains to solve

$$(u^2+v^2)^2=d^2, \qquad 2uv=b. \tag{6}$$

Now $d$ is only determined up to sign and we may assume that $d\geqslant 0$. Then $d$ is a square, say $d=e^2$ and the first equation (6) can be replaced by $u^2+v^2=e^2$. Either $b$ or $-b$ is a square; if $b$ is a square, then $e^2+b>0$, hence $e^2+b=h^2$, and if we now solve for $u,v$ from $u+v=h$, $u-v=a/h$, we find that (5) holds with these values. Similarly if $-b$ is a square, then $e^2-b>0$, hence $e^2-b=k^2$ and we have instead to solve $u-v=k$, $u+v=a/k$. Thus every element of $K(\mathrm{i})$ has a square root, and the usual formula for quadratic equations shows that every quadratic equation over $K(\mathrm{i})$ has a root.

Suppose now that $K(\mathrm{i})$ is not algebraically closed; then any finite extension of $K(\mathrm{i})$ is contained in a Galois extension $L/K$ (recall that $\mathrm{char}\,K=0$, so every extension is separable). Let $G=\mathrm{Gal}(L/K)$, take a Sylow 2-subgroup $S$ of $G$ and denote by $H$ the fixed field of $S$. Then $[H:K]=(G:S)$ is odd, by the definition of $S$, hence any element $\alpha$ of $H$ has odd degree over $K$, but by (b) all such elements belong to $K$, so $H=K$ and $S=G$ is a 2-group. It follows that the Galois extension $L/K(\mathrm{i})$ also has a 2-group $T$, say. Now we recall from A.1, Th. 3 of **9.8**, that any non-trivial 2-group has a subgroup of index 2. A subgroup of index 2 in $T$ corresponds to an extension of $K(\mathrm{i})$ of degree 2, but this contradicts the fact that every quadratic equation over $K(\mathrm{i})$ has a root in $K(\mathrm{i})$. Hence $T=1$ and $L=K(\mathrm{i})$, i.e. $K(\mathrm{i})$ has no proper algebraic extensions, and is therefore algebraically closed. ■

As we saw, every Euclidean field can be ordered in just one way; hence we have

COROLLARY 3.8 *Every real closed field can be ordered in just one way; with this ordering every positive element has a square root, and any automorphism necessarily preserves the ordering.* ■

With the help of Th. 3.7 we can also deduce the fundamental theorem of algebra (cf. A.2, **3.3**). We shall need a property of real numbers proved in A.2, the *intermediate value property*. This states that if a polynomial $f$ over a field $K$ is such that $f(\alpha)f(\beta)<0$ for some $\alpha<\beta$ in $K$, then there exists $\gamma\in K$ such that $\alpha<\gamma<\beta$ and $f(\gamma)=0$. This depends on the continuity property of polynomials, and is well known to hold for $\mathbf{R}$ (cf. A.2, Prop. 6.6.6).

PROPOSITION 3.9 *An ordered field is real closed if and only if it has the intermediate value property for polynomials.*

*Proof.* Let $K$ be a real closed ordered field, $f$ any polynomial over $K$ and $\alpha,\beta\in K$, say $\alpha<\beta$, such that $f(\alpha)f(\beta)<0$. Since its algebraic closure is of degree 2 over $K$ (by Th. 3.7), we can write $f$ as a product of linear and quadratic factors, irreducible over $K$. Each irreducible quadratic factor has the form

$$x^2+px+q=(x+p/2)^2+(q-p^2/4),$$

and here $q-p^2/4>0$, by irreducibility. Hence this factor is positive for all values of $x$, so $f$ can change sign only when a linear factor changes sign, and this happens only at a zero of $f$, so there must be a zero lying between $\alpha$ and $\beta$.

Conversely, when the intermediate value property holds for $K$, then we shall show that $K$ is real closed by verifying (b) of Th. 3.7. Let $a>0$ in $K$; the polynomial $f=x^2-a$ is positive for $x=1+a$ and negative for $x=0$, hence it has a zero, and this is the desired square root. Similarly if $f$ is a polynomial of odd degree, then $f$ has the same sign as its leading coefficient for large positive $x$ and the opposite sign for large negative $x$, hence it changes sign and so has a zero in $K$. Thus (b) of Th. 3.7 holds and it follows that $K$ is real closed. ■

We have seen in A.2, Prop. 6.6.6 that $\mathbf{R}$ has the intermediate value property, hence we obtain

COROLLARY 3.10 *The field* $\mathbf{R}$ *of real numbers is real closed and the field of complex numbers, defined as* $\mathbf{C}=\mathbf{R}(\sqrt{-1})$ *is algebraically closed.* ■

This proof of the fundamental theorem of algebra (and the part of Th. 3.7 on which it depends) is due to E. Artin.

In Th. 3.5 we saw that every formally real field has a real closure and we now ask to what extent this real closure is unique. The answer is at first

disappointingly vague, but it can be made precise once the problem is reformulated: by Th. 3.5 and Cor. 3.8 every formally real field can be ordered. If we now take an ordered field $K$ as our starting point, we shall find that its real closure is unique, up to a unique $K$-isomorphism. One method of proof is by Sturm's theorem, which determines the exact number of real roots of an equation over $\mathbf{R}$ in any interval (cf. Ex. (12)). We shall instead use the proof of Knebusch 1972, as simplified by Becker and Spitzlay 1975, cf. Scharlau [1985].

Let $K$ be any ordered field. A quadratic form $q$ over $K$ is called *positive-definite* if $q(x) > 0$ for all $x \neq 0$; if $q(x) < 0$ for all $x \neq 0$, $q$ is said to be *negative-definite*. Given any quadratic space $(V, q)$ with a regular form $q$, we can write $V$ as an orthogonal sum

$$V = V^+ \perp V^- \tag{7}$$

of two subspaces, $V^+$ where $q$ is positive-definite and $V^-$ where $q$ is negative-definite. These subspaces are not unique, but their dimensions are unique, and the difference

$$\operatorname{sgn}(q) = \dim(V^+) - \dim(V^-) \tag{8}$$

is called the *signature* of $q$. For the case $K = \mathbf{R}$ this just reduces to Sylvester's law of inertia (A.1, **8.3**) and the proof is the same as in that case: if $V = W^+ \perp W^-$ is another such decomposition, then $W^+ \cap V^- = 0$, hence $\dim(W^+) \leqslant \dim(V) - \dim(V^-) = \dim(V^+)$; by symmetry we have equality and similarly for $\dim(V^-)$.

The definition (8) can still be used for singular spaces, by taking the signature of the regular part.

Consider a finite-dimensional $K$-algebra $A$. We have the trace form on $A$, given by the trace of the regular representation. If $\rho_a: x \mapsto xa$, then the *trace* of $a$, $\operatorname{tr}(a)$, is defined as the trace of the linear mapping $\rho_a$. This provides us with a quadratic form $\operatorname{tr}(x^2)$ on $A$, which is regular whenever $A$ is semisimple. For consider the set

$$\mathfrak{N} = \{y \in A \mid \operatorname{tr}(xy) = 0 \text{ for all } x \in A\}. \tag{9}$$

This set $\mathfrak{N}$ is an ideal in $A$, because $\operatorname{tr}(x(ay)) = \operatorname{tr}((xa)y)$, $\operatorname{tr}(x(ya)) = \operatorname{tr}(ax)y)$. By the definition (9), for any $a \in \mathfrak{N}$ all powers of $\rho_a$ have zero trace, hence $\rho_a$ is nilpotent, and so $a$ is nilpotent, because $\rho$ is faithful. Hence $\mathfrak{N}$ is a nilpotent ideal and so $\mathfrak{N} = 0$, because $A$ is semisimple. This shows the quadratic form $\operatorname{tr}(x^2)$ on $A$ to be regular.

We shall need the following lemma, which goes back to Sylvester.

LEMMA 3.11 *Let $K$ be an ordered field, with real closure $\bar{K}$. If $f$ is a polynomial in $x$ over $K$, $A = K[x]/(f)$ and $q$ is the quadratic form on $A$ defined by $q(x) = \operatorname{tr}(x^2)$, then $\operatorname{sgn}(q)$ equals the number of distinct zeros of $f$ in $\bar{K}$. Hence the number of these zeros depends only on the ordering of $K$, not on $\bar{K}$.*

*Proof.* Over $\bar{K}$ the polynomial $f$ splits into factors that are either linear or quadratic:

$$f(x) = p_1(x)^{\alpha_1} \cdots p_r(x)^{\alpha_r},$$

hence

$$A_{\bar{K}} = \bar{K}[x]/(f) = E_1 \times \cdots \times E_r, \qquad \text{where } E_i = \bar{K}[x]/(p_i^{\alpha_i}). \tag{10}$$

It is clear that the signature of $q$ is the same over $K$ and over $\bar{K}$, and to compute it over $\bar{K}$ we may use a basis adapted to the decomposition (10). The polynomial $p_i$ has degree 1 or 2; we consider these cases separately: In $\bar{K}[x]/((x-a)^\alpha)$ we can take as our basis $1, x-a, \ldots, (x-a)^{\alpha-1}$. The quadratic form has rank 1 and $\operatorname{tr}(1) = 1$, $\operatorname{tr}((x-a)^i) = 0$ for $i > 0$, hence the signature is 1 in this case.

If $E = \bar{K}[x]/(p^\alpha)$, where $p = x^2 + \lambda x + \mu$, and $4\mu > \lambda^2$ (because $p$ is irreducible over $\bar{K}$), then on writing $y = (2x+\lambda)(4\mu - \lambda^2)^{-1/2}$, we can express this as $\bar{K}[y]/((y^2+1)^\alpha)$. Taking as our basis $1, y, y^2+1, y(y^2+1), (y^2+1)^2, \ldots$, we find that our quadratic form now has rank 2 and $\operatorname{tr}(y^i(y^2+1)^j) = 0$ if $j \geqslant 1$, $\operatorname{tr}(1) = 2$, $\operatorname{tr}(y) = 0$, $\operatorname{tr}(y^2) = -2$, therefore its signature is 0. Thus in each case the signature equals the number of distinct zeros of $f$ in $\bar{K}$, as claimed. ■

We shall use this result to examine the extensions of the ordering of $K$ to a simple extension:

PROPOSITION 3.12 *Let $K$ be an ordered field and $L = K(\alpha)$ a simple algebraic extension, with quadratic form defined by the trace on $L$: $q(x) = \operatorname{tr}_{L/K}(x^2)$. If $\operatorname{sgn}(q) = r > 0$, then $L$ has an ordering extending that of $K$ and there are at most $r$ distinct orderings of $L$ extending the ordering on $K$, obtained by taking $r$ $K$-homomorphisms of $L$ into a real closure of $K$.*

*Proof.* Denote by $\bar{K}$ a real closure of $K$; by Lemma 3.11, $r$ is the number of zeros of the minimal polynomial of $\alpha$ in $\bar{K}$, and each zero $\alpha_i$ gives rise to an embedding $\sigma_i$ of $L$ in $\bar{K}$ by mapping $\alpha \mapsto \alpha_i$. The ordering induced in $L$ in this way clearly extends the ordering of $K$. If there is an ordering $>_0$ of $L$ distinct from these $r$, then for $i = 1, \ldots, r$ there exists $c_i \in L$ such that $c_i >_0 0$, but $c_i^{\sigma_i} < 0$ in $\bar{K}$. By repeated application of Lemma 3.6 the ordering $>_0$ extends to

$$L(\sqrt{c_1}, \ldots, \sqrt{c_r}) = K(\gamma), \tag{11}$$

for a primitive element $\gamma$ over $K$ with minimal polynomial $g$. Now $K(\gamma)$ has a real closure $M$, which is a real closed algebraic extension of $K$, and so is a real closure for $K$. Since $g$ has a zero in $M$, $g$ also has a zero in $\bar{K}$, by Lemma 3.10. Hence there is an embedding $\tau: K(\gamma) \to \bar{K}$; the restriction of $\tau$ to $K$ must equal one of the $\sigma_i$, say $\sigma_1$. But then $c_1^{\sigma_1} = c_1^\tau = ((\sqrt{c_1})^\tau)^2 > 0$, which contradicts the fact that $c_1^{\sigma_1} < 0$. ■

If $L$ is an algebraic extension of an ordered field $K$, then the ordering cannot necessarily be extended to $L$; but when it can, then the bound given in Prop. 3.12 is always attained. This can be proved at the same time as the uniqueness of the real closure:

THEOREM 3.13 *Let $K$ be an ordered field, $L$ a finite algebraic extension, say $L = K(\gamma)$, and let $r = \mathrm{sgn}(q)$, where $q(x) = \mathrm{tr}_{L/K}(x^2)$. If the ordering of $K$ can be extended to $L$, then there are exactly $r$ such extensions and for each such ordering $L$ has an order-preserving embedding in the real closure $\bar{K}$ of $K$.*

*Further, any two real closures of $K$ are isomorphic by a unique isomorphism and this isomorphism preserves the ordering.*

*Proof.* We first show that any ordered algebraic extension $L$ of $K$ has an order-embedding in $\bar{K}$. For by Zorn's lemma there is a maximal subfield $L_0$ of $L$ with an order-embedding in $\bar{K}$. If $L_0 \neq L$, say $\alpha \in L \setminus L_0$, then $L_0(\alpha)$ is an ordered algebraic extension of $L_0$ and by Prop. 3.12, $L_0(\alpha)$ has an order-embedding in $\bar{K}$; this contradicts the maximality of $L_0$, hence $L_0 = L$ and so $L$ has an order-embedding in $\bar{K}$.

Next we prove the last part. Suppose that $L_1, L_2$ are real closures of $K$; by what has been shown there exists a $K$-homomorphism $\sigma: L_1 \to L_2$. Since the image $L_1^\sigma$ is real closed, we have $L_1^\sigma = L_2$ and so $\sigma$ is an isomorphism; since $(\alpha^2)^\sigma = (\alpha^\sigma)^2$ for any $\alpha \in L_1$, we have $\alpha^\sigma > 0$ in $L_2$ iff $\alpha > 0$ in $L_1$, thus $\sigma$ preserves the ordering. Let $\tau: L_1 \to L_2$ be another isomorphism; we have to show that $\alpha^\sigma = \alpha^\tau$ for any $\alpha \in L_1$. Denote by $f$ the minimal polynomial of $\alpha$ over $K$ and by $\alpha_1, \ldots, \alpha_r$ its conjugates in $L_1$, where $\alpha_1 < \cdots < \alpha_r$ say. Since $\sigma, \tau$ are order-preserving, we have

$$\alpha_1^\sigma < \cdots < \alpha_r^\sigma, \qquad \alpha_1^\tau < \cdots < \alpha_r^\tau.$$

Now $\alpha_1^\sigma, \ldots, \alpha_r^\sigma$ are the zeros of $f$ in $L_2$ and likewise $\alpha_1^\tau, \ldots, \alpha_r^\tau$, hence $\alpha_i^\sigma = \alpha_i^\tau$ for $i = 1, \ldots, r$, and so $\alpha^\sigma = \alpha^\tau$, as we had to show.

Finally let $L = K(\gamma)$ be as stated; then there are $r$ embeddings $\sigma_1, \ldots, \sigma_r$ in $\bar{K}$ and we have to show that they define distinct orderings of $L$. If $\sigma_1, \sigma_2$ define the same ordering on $L$, take a real closure $M$ of $L$ for this ordering. By the first part $\sigma_1$ and $\sigma_2$ can be extended to $K$-isomorphisms $M \to \bar{K}$ and we now have distinct $K$-isomorphisms between the real closed fields $M, \bar{K}$; but we have seen that this is impossible. Hence there are $r$ distinct orderings on $L$. ■

In A.1, **6.8** we found several ways of estimating the number of roots of an equation, such as the Harriot–Descartes rule, which dates from 1637. However, this left the question of finding the exact number of real roots open, and it was not until nearly 200 years later that this problem was solved, by Sturm (his solution, announced in 1829, was first published in 1835). We shall prove Sturm's theorem here using only the intermediate value property of **R** established in

A.2, Prop. 6.6.6. It holds for any real closed field and can be used to give another proof of the uniqueness of the real closure of a formally real field, but we have preferred the above more direct method. The result, though of independent interest, will not be needed later and can be omitted without loss of continuity.

Our object is to determine the exact number of real roots of a given equation over **R**. More precisely, given real numbers $\alpha < \beta$, we shall describe a method of calculating the number of roots $\gamma$ satisfying $\alpha < \gamma < \beta$.

Let $F$ be a polynomial over **R**; by a *Sturm sequence* of polynomials for $f$ in the interval $[\alpha, \beta]$ we understand a sequence of polynomials over **R**:

$$f_0 = f, f_1, \ldots, f_r \tag{12}$$

such that

**S.1** *$f_r$ has no zeros in* $[\alpha, \beta]$,
**S.2** $f_0(\alpha), f_0(\beta) \neq 0$,
**S.3** *For* $i = 1, \ldots, r-1$, *if* $f_i(\gamma) = 0$, *where* $\alpha < \gamma < \beta$, *then* $f_{i-1}(\gamma) f_{i+1}(\gamma) < 0$.
**S.4** *If* $f(\gamma) = 0$, *where* $\alpha < \gamma < \beta$, *then* $f_0(x) f_1(x)$ *is an increasing function of* $x$ *at* $x = \gamma$, *thus* $f_0 f_1$ *changes from negative to positive values in a sufficiently small interval about* $\gamma$.

We shall soon find how to construct such sequences; for the moment we show how a Sturm sequence can be used to answer the question raised above. As in A.1, **6.8** we define the number of sign changes in a sequence of real numbers $\lambda_1, \ldots, \lambda_n$ as the number of times there is a change from a positive to negative value or vice versa, ignoring zeros; e.g. $2, 0, 3, -1, 0, 0, 4$ has two changes.

THEOREM 3.14 *Let $f$ be a real polynomial of degree $n$, with Sturm sequence* (12) *for the interval* $[\alpha, \beta]$, *and for any* $\gamma \in \mathbf{R}$ *denote by* $\delta(\gamma)$ *the number of sign changes in the sequence*

$$f_0(\gamma), f_1(\gamma), \ldots, f_r(\gamma). \tag{13}$$

*Then the number of distinct zeros of $f$ in* $[\alpha, \beta]$ *is* $\delta(\alpha) - \delta(\beta)$.

*Proof.* Consider how the sequence (13) changes as $\gamma$ varies from $\alpha$ to $\beta$. The only sign changes happen when $\gamma$ passes through a zero of some $f_i$ (by the intermediate value property for **R**). We first take the case $i > 0$. By **S.1** we also have $i < r$, thus $f_i(\gamma) = 0$ and by **S.3**, $f_{i-1}(\gamma) f_{i+1}(\gamma) < 0$, say $f_{i-1}(\gamma) < 0 < f_{i+1}(\gamma)$. Then in the section $f_{i-1}, f_i, f_{i+1}$ of (13) there is one change of sign. We still have $f_{i-1}(\gamma') < 0 < f_{i+1}(\gamma')$ for any $\gamma'$ near $\gamma$; for such $\gamma'$ the number of sign changes in the sequence $f_{i-1}, f_i, f_{i+1}$ is still 1, whatever the value of $f_i(\gamma')$. Hence the number of sign changes in (13) remains constant as the argument passes through a zero of $f_i (i > 0)$. Next take the case when the argument passes through a zero $\gamma$ of $f_0$; then $f_1(\gamma) \neq 0$ by **S.3** for $i = 1$. By taking a sufficiently small

interval about $\gamma$ we may assume that $\gamma$ is the only zero of $f_0$ in this interval and that $f_1$ has no zeros there. Assume first that $f_0(\gamma') < 0$ for $\gamma' < \gamma$; then $f_1(\gamma') > 0$ by **S.4**, and either $f_0(\gamma'') > 0$, $f_1(\gamma'') > 0$ for $\gamma'' > \gamma$, or $f_0(\gamma'') < 0$, $f_1(\gamma'') < 0$ for $\gamma'' > \gamma$. In the second case $f_1$ changes from positive to negative values and hence vanishes, against the hypothesis. Thus the first alternative must hold, and as $x$ increases from $\gamma' < \gamma$ to $\gamma'' > \gamma$, the sequence $f_0(x), f_1(x)$ changes from $-, +$ to $+, +$, so one sign change is lost as $x$ passes through a zero of $f$. The same argument applies if $f_0(\gamma') > 0$ for $\gamma' < \gamma$. By addition we find that the total number of sign changes in $[\alpha, \beta]$ is $\delta(\alpha) - \delta(\beta)$. ■

Now let $f$ be any polynomial and define the *standard sequence* for $f$ as follows: $f_0 = f, f_1 = f'$ (the derivative) and for $i \geqslant 1, f_{i+1}$ is defined as the remainder in the division algorithm, with the sign changed:

$$f_{i-1} = f_i q_i - f_{i+1}, \qquad \deg f_{i+1} < \deg f_i. \tag{14}$$

Let $f_r$ be the last non-zero remainder; by the Euclidean algorithm, $f_r$ is the highest common factor of $f$ and $f'$. Thus if $f$ has no repeated zeros, $f_r$ is a non-zero constant. We claim that in this case $f_0, f_1, \ldots, f_r$ is a Sturm sequence for $f$, for any interval $[\alpha, \beta]$ such that $f(\alpha) f(\beta) \neq 0$. **S.1** holds since $f_r$ is constant, and **S.2** holds by assumption; further no two successive $f_i$'s can vanish for the same value of $x$, for if they did, then by (14) all succeeding $f_i$'s must vanish and this contradicts **S.1**. Now if $f_i(\gamma) = 0$, then by (14), $f_{i-1}(\gamma) f_{i+1}(\gamma) < 0$, so **S.3** holds. Finally **S.4** holds because if $f(\gamma) = 0$, then $f^2$ has a minimum at $x = \gamma$ and so $(f^2)' = 2 f_0 f_1$ is increasing at $x = \gamma$.

Suppose now that $f$ is an arbitrary polynomial and let $d$ be the highest common factor of $f$ and $f'$, say $f = d g_0$, $f' = d g_1$. Clearly $d$ is a factor of each $f_i$, so writing $f_i = d g_i$, we have

$$g_{i-1} = g_i q_i - g_{i+1}, \qquad \deg g_{i+1} < \deg g_i. \tag{15}$$

We claim that $g_0, g_1, \ldots, g_r$ is a Sturm sequence for $g_0$. **S.1–3** follow as before; to establish **S.4** we note that when $g_0(\gamma) = 0$, then $f(\gamma) = 0$, $f^2$ has a minimum and so $(f^2)' = 2d^2 g_0 g_1$ is increasing at $x = \gamma$. Clearly $2d^2$ is non-negative, so $g_0 g_1$ is still increasing at $x = \gamma$.

We thus find that

$$g_0, g_1, \ldots, g_r \tag{16}$$

is a Sturm sequence for $g_0$. Now the sequence

$$f_0, f_1, \ldots, f_r \tag{17}$$

is obtained by multiplying each term of (16) by $d$; hence the numbers of sign changes of (16) and (17) are the same at any point $\gamma$ not a zero of $d$. We can therefore use (17) to determine the zeros of $g_0$. But they are just the distinct zeros of $f$, so we have proved

THEOREM 3.15 (Sturm's theorem) *Let $f$ be a polynomial over* $\mathbf{R}$ *and let $f_0 = f, f_1 = f', f_2, \ldots, f_r$ be the standard sequence for $f$. Given $\alpha, \beta \in \mathbf{R}$, not zeros of $f$, such that $\alpha < \beta$, the number of distinct zeros of $f$ (not counting multiplicities) in $[\alpha, \beta]$ is $\delta(\alpha) - \delta(\beta)$.* ■

We note that by taking $\alpha, \beta$ sufficiently large negative and positive numbers respectively, we obtain a formula for the total number of real zeros of $f$. As we saw in A.1, **6.8**, if $f = x^n + a_1 x^{n-1} + \cdots + a_n$ and $M = \max\{|a_1|, \ldots, |a_n|\} + 1$, then all the real zeros of $f$ lie in the interval $[-M, M]$.

## Exercises

(1) Verify that any cone satisfying (2) defines an ordering.

(2) Let $R$ be any ring and $P$ a maximal cone in $R$. Show that $P \cup -P = R$ and $P \cap -P$ is a prime ideal in $R$.

(3) Let $K$ be a field of characteristic not two which is not formally real. Show that the positive core of $K$ is the whole of $K$. Deduce that every element is a sum of squares.

(4) Show that there is a least Euclidean subfield $E$ of $\mathbf{R}$ and verify that $E$ is not real closed.

(5) Show that there is a least subfield $F$ of $\mathbf{R}$ over which every equation of odd degree has a root in $F$; verify that $F$ is not Euclidean.

(6) Show that a field is real closed iff its algebraic closure is of finite degree $> 1$ over it.

(7) Let $f = 0$ be an equation with real coefficients, without repeated roots. Show that if the number of pairs of conjugate complex roots of $f = 0$ is $s$, then the discriminant of $f$ has sign $(-1)^s$.

(8) In Th. 3.7 show that (b) $\Rightarrow$ (c) without using Sylow theory by forming, for any equation $f = 0$ with roots $\alpha_1, \ldots, \alpha_n$ the equation with roots $\alpha_i + \alpha_j + \lambda \alpha_i \alpha_j (i < j)$ where $\lambda$ is chosen to make these values distinct, and using induction on the 2-power dividing $\deg f$.

(9) Let $K$ be a field such that $x^2 + 1$ is irreducible over $K$, while $K(\sqrt{-1})$ is algebraically closed. Show that any polynomial over $K$ splits into linear and quadratic factors, and that the sum of two squares in $K$ is a square. Deduce that $K$ is formally real and in fact real closed. (Hint: consider $N(\sqrt{(a + b\sqrt{-1})})$.)

(10) Apply Sturm's theorem to determine the real roots of $x^4 - 3x^3 - x^2 + 8x - 4 = 0$ to the nearest integer.

(11) The Legendre polynomials, defined as

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2 - 1)^n,$$

$(n = 0, 1, \ldots)$ satisfy the relations

$$nP_n - (2n-1)xP_{n-1} + (n-1)P_{n-2} = 0,$$
$$(1-x^2)P_n' + nxP_n - nP_{n-1} = 0 \qquad (n = 1, 2, \ldots).$$

Show that for any $n \geqslant 0$, $P_n, P_{n-1}, \ldots, P_0$ is a Sturm sequence for $P_n$ in $[-1, 1]$. Deduce that $P_n$ has $n$ distinct zeros in $[-1, 1]$ which are separated by those of $P_{n-1}$.

(12) Prove the last part of Th. 3.13 by using Th. 3.15.

## 8.4 The Witt ring of a field

In **8.1–2** we have studied a single quadratic form over a field. We now come to a construction which embodies information about the totality of all quadratic forms over a given field. Throughout this section $k$ is a field of characteristic not 2.

In A.2, **6.4** we associated with $k$ the monoid $M(k)$ whose elements were equivalence classes (under isometry) of regular quadratic spaces. The class of $(V, q)$ is denoted simply by $[V]$ and the operation in $M(k)$ is denoted by $+$ and is defined by

$$[U] + [V] = [U \perp V].$$

It is easily seen that the right-hand side depends only on $[U], [V]$ and not on $U, V$, and so is a well defined element of $M(k)$. Further, the Witt cancellation theorem (A.2, Th. 6.4.1) shows that $M(k)$ is a cancellation monoid. Let us form the group $\hat{W}(k)$ of 'fractions' of this monoid (cf. A.1, Prop. 2 of **6.2**). Formally the elements of $\hat{W}(k)$ can be written as differences of isometry classes: $[U] - [V]$, and we have $[U] - [V] = [U'] - [V']$ iff $U \perp V' \cong U' \perp V$. Since $M(k)$ satisfies cancellation, it is embedded in $\hat{W}(k)$. Our object is to define a multiplication on $\hat{W}(k)$ which will make it into a ring.

Let $V, V'$ be regular quadratic spaces with associated bilinear forms $b, b'$; we can define a bilinear form $F$ on the tensor product $V \otimes V'$ by the formula

$$F(x \otimes x', y \otimes y') = b(x, y)b'(x', y') \qquad (x, y \in V, x', y' \in V').$$

By linearity this determines $F$ completely; we observe that in terms of the corresponding quadratic forms $q, q', P$ we have

$$P(x \otimes x') = q(x)q'(x'),$$

but this formula by itself it not enough to determine $P$, because $P$ is not linear and we need to define it for sums $\sum x_i \otimes x_i'$, not merely the products $x \otimes x'$. In terms of matrices, if the forms on $V, V'$ have the matrices $A, A'$ respectively (relative to any bases), then the matrix for $V \otimes V'$ is $A \otimes A'$. From the properties of the tensor product it is clear that this operation induces a multiplication on $\hat{W}(k)$:

$$[U][V] = [U \otimes V].$$

Clearly this multiplication is associative and commutative, and on the basis of the formula

$$U \otimes (V \perp V') \cong U \otimes V \perp U \otimes V'$$

it follows easily that the distributive law holds in $\hat{W}(k)$. Further, we have $U \otimes \langle 1 \rangle \cong U$ and if $Z$ is the zero-dimensional space, then $U \otimes Z \cong Z$. Hence $\hat{W}(k)$ is a commutative ring with the zero space as 0 and $\langle 1 \rangle$ as unit element; it is called the *Witt–Grothendieck ring* of quadratic forms over $k$. In terms of a diagonal representation we have

$$\langle a_1, \ldots, a_r \rangle \otimes \langle b_1, \ldots, b_s \rangle \cong \langle a_1 b_1, a_1 b_2, \ldots, a_1 b_s, a_2 b_1, \ldots, a_r b_s \rangle.$$

For any $a, b \in k^\times$, if $a + b \neq 0$, then

$$\langle a, b \rangle \cong \langle a + b, (a + b)ab \rangle. \tag{1}$$

For if $u, v$ is an orthogonal basis for $\langle a, b \rangle$, then $u + v, bu - av$ is again an orthogonal basis and $q(u + v) = a + b, q(bu - av) = b^2 a + a^2 b = (a + b)ab$, so we obtain the right-hand side of (1). Thus in $\hat{W}(k)$, $\langle a \rangle + \langle b \rangle = \langle a + b \rangle + \langle (a + b)ab \rangle$.

We next try to find a presentation for $\hat{W}(k)$. To this end let us define an abstract ring $V(k)$ with generators $[a], a \in k^\times$ and defining relations

**R.**1 $[1] = 1$,
**R.**2 $[a][b] = [ab]$ *for all* $a, b \in k^\times$,
**R.**3 $[a] + [b] = [a + b] + [(a + b)ab]$ *for all* $a, b \in k^\times$ *such that* $a \neq -b$.

By **R.**2, $V(k)$ is commutative; we also note the following consequence of **R.**1–3:

$$[ab^2] = [a]. \tag{2}$$

To see this we put $a = b$ in **R.**3: $2[b] = [2b] + [2b^3]$. Multiplying by $[a/2b]$ we get $2[a/2] = [a] + [ab^2]$, and since the left-hand side is independent of $b$, it follows that $[ab^2] = [a]$, i.e. (2).

We have seen that **R.**1–3 hold for the generators $\langle a \rangle$ of $\hat{W}(k)$; therefore there is a unique ring homomorphism $\varphi: V(k) \to \hat{W}(k)$ defined by $[a] \mapsto \langle a \rangle (a \in k^\times)$; we claim that $\varphi$ is an isomorphism. Since $\hat{W}(k)$ is additively generated by the $\langle a \rangle$, $\varphi$ is surjective. Now any element of $V(k)$ can by **R.**2 be written in the form

$$[a_1] + \cdots + [a_n] - [b_1] - \cdots - [b_m] \qquad (a_i, b_j \in k^\times); \tag{3}$$

if $\ker \varphi \neq 0$, we may choose a non-zero element (3) of $\ker \varphi$ so as to minimize $n$. This means that $\langle a_1, \ldots, a_n \rangle \cong \langle b_1, \ldots, b_m \rangle$ and hence $m = n$, by the invariance of the rank. Here we must have $n > 1$, for otherwise $a_1 = b_1 x^2$ and so $[a_1] = [b_1]$ by (2). Now since $b_1$ is represented by $\langle a_1, \ldots, a_n \rangle$, we have an expression

$$b_1 = a_1 x_1^2 + \cdots + a_p x_p^2 \qquad \text{where } p \leqslant n. \tag{4}$$

Suppose that the $a_i$ in (3) are chosen so as to minimize $p$ in (4). If $p > 1$, then $d = a_1 x_1^2 + a_2 x_2^2$ is non-zero and by **R.**3 and (2),

$$[a_1] + [a_2] = [a_1 x_1^2] + [a_2 x_2^2] = [d] + [d \cdot a_1 x_1^2 a_2 x_2^2] = [d] + [d a_1 a_2].$$

Therefore $[a_1] + \cdots + [a_n] = [d] + [d a_1 a_2] + [a_3] + \cdots + [a_n]$ and now $b_1 = d + a_3 x_3^2 + \cdots + a_p x_p^2$, which contradicts the choice of $p$. Thus $p = 1, b_1 = a_1 x^2$, hence $[a_1] = [b_1]$ and by Witt's cancellation theorem, $\ker \varphi$ contains

$$[a_2] + \cdots [a_n] - [b_2] - \cdots [b_n].$$

This contradicts the minimality of $n$, hence $\ker \varphi = 0$ as claimed.

Thus we have proved

THEOREM 4.1 *Let $k$ be a field of characteristic not 2. The the Witt–Grothendieck ring $\hat{W}(k)$ has a presentation with generators $\langle a \rangle$, $a \in k^\times$ and defining relations*

**W.**1 $\langle 1 \rangle = 1$,

**W.**2 $\langle a \rangle \langle b \rangle = \langle ab \rangle$ *for* $a, b \in k^\times$,

**W.**3 $\langle a \rangle + \langle b \rangle = \langle a + b \rangle + \langle (a+b)ab \rangle$ *for* $a, b \in k^\times$ *such that* $a \neq -b$. ■

A closer examination reveals a further consequence. Let us call two diagonal forms $\langle a_1, \ldots, a_n \rangle$ and $\langle b_1, \ldots, b_n \rangle$ *simply related* if there exist $i, j$ such that $\langle a_i, a_j \rangle \cong \langle b_i, b_j \rangle$ and $a_k = b_k$ for all $k \neq i, j$. Here $i$ may equal $j$, in which case we interpret $\langle a_i, a_j \rangle$ as $\langle a_i \rangle$. Now the proof of Th. 4.1 shows the truth of

THEOREM 4.2 (Witt's chain equivalence theorem) *Let $q, q'$ be two regular diagonal quadratic forms (over a field of characteristic not 2) which are equivalent. Then there is a sequence of diagonal forms $q_1 = q, q_2, \ldots, q_r = q'$ such that for $i = 2, \ldots, r, q_{i-1}$ is simply related to $q_i$.* ■

Every invariant of quadratic forms gives rise to a function on $\hat{W}(k)$. Consider e.g. the rank; clearly isometric forms have the same rank, hence the rank function induces a mapping

$$\mathrm{rk}: \hat{W}(k) \to \mathbf{Z}, \tag{5}$$

and since $\mathrm{rk}(U \perp V) = \mathrm{rk}\, U + \mathrm{rk}\, V$, $\mathrm{rk}(U \otimes V) = \mathrm{rk}\, U \cdot \mathrm{rk}\, V$, $\mathrm{rk}\langle 1 \rangle = 1$, it follows that (5) is a ring homomorphism; the kernel $\hat{I}(k)$ is called the *augmentation ideal* of $\hat{W}(k)$. Since (5) is split by the natural mapping $\mathbf{Z} \to \hat{W}(k)$, we have a direct decomposition

$$\hat{W}(k) \cong \mathbf{Z} \oplus \hat{I}(k).$$

A second important ideal in $\hat{W}(k)$ consists of the split spaces. We recall (A.2, **6.4**) that a *split* quadratic space is an orthogonal sum of hyperbolic planes. Hence a space is split iff its dimension is even, say $n = 2r$, and relative to a

suitable basis its matrix is $\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$. A quadratic space is split whenever it is regular of even dimension $2r$ and has a totally isotropic $r$-dimensional subspace. For then its matrix relative to a suitable basis has the form

$$\begin{pmatrix} 0 & P \\ P^T & Q \end{pmatrix}$$

where $P, Q$ are $r \times r$ and $P$ is invertible. Now we have

$$\begin{pmatrix} I & 0 \\ A & B \end{pmatrix}\begin{pmatrix} 0 & P \\ P^T & Q \end{pmatrix}\begin{pmatrix} I & A^T \\ 0 & B^T \end{pmatrix} = \begin{pmatrix} 0 & PB^T \\ BP^T & BP^TA^T + APB^T + BQB^T \end{pmatrix},$$

and this reduces to $\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$ when $B = (P^T)^{-1}$, $A = -\frac{1}{2}BQB^T$.

We shall also need Th. 6.4.3 of A.2, which stated that in characteristic not 2 any regular quadratic space $V$ can be written as

$$V = L \perp U, \tag{6}$$

where $L$ is split and $U$ is anisotropic. If $\dim L = 2r$, then $r$, and hence the isometry type of $L$, is uniquely determined as the maximal dimension of a totally isotropic subspace of $V$, while $U$ is unique up to isometry.

The decomposition (6) is called the *Witt decomposition* of $V, r$ is the *Witt index* and $U$ the *anisotropic part* of $V$.

If $L$ is any split space, then so is $L \otimes U$, for any regular quadratic space $U$. For if $\dim L = 2r$, $\dim U = n$, then $L$ has a totally isotropic subspace $L_0$ of dimension $r$, hence $L_0 \otimes U$ is a totally isotropic subspace of $L \otimes U$ of dimension $rn$. This means that the additive subgroup spanned by the split spaces in $\hat{W}(k)$ is already an ideal. We shall often write $H$ for the element of $\hat{W}(k)$ representing the hyperbolic plane, thus $H = [\langle 1, -1 \rangle]$; the ideal of split spaces may then be written $\mathbf{Z}H$. The residue class ring $W(k) = \hat{W}(k)/\mathbf{Z}H$ is called the *Witt ring* of $k$. A more direct description of $W(k)$ is given in

PROPOSITION 4.3 *Let $k$ be any field of characteristic not 2. The correspondence*

$$V \mapsto [V] \qquad (\text{mod } \mathbf{Z}H) \tag{7}$$

*induces a bijection between the set of isometry classes of anisotropic spaces and the Witt ring $W(k)$; the negative of $(V, q)$ is $(V, -q)$. Given two anisotropic spaces $U, V$, neither $U \perp V$ nor $U \otimes V$ need be anisotropic, but if we take their anisotropic parts provided by the Witt decomposition we obtain the sum and product in $W(k)$.*

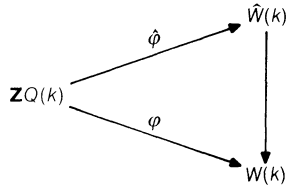
*Proof.* Let us call two spaces (or also their forms) *similar* if their anisotropic parts (obtained from the Witt decomposition) are isometric. It is clear that similar spaces have the same image in $W(k)$. Since $H = [\langle a, -a \rangle]$ for any

$a \in k^\times$, it follows that $-[\langle a\rangle] = [\langle -a\rangle]$ in $W(k)$; hence every element of $W(k)$ is represented by a space $V$. If we take a Witt decomposition $V = V_h \perp V_a$, where $V_h$ is split and $V_a$ anisotropic, then $V$ is similar to $V_a$ and so $[V] = [V_a]$ and it follows that (7) is surjective. Now let $[U] = [V]$; then in $\hat{W}(k)$ we have $[U] = [V] + nH$, where $n \in \mathbf{Z}$. If $n \geqslant 0$ say, then $U \cong V \perp nH$, hence $U$ and $V$ are similar. This shows that (7) is injective, as claimed. Now the rest is clear. ■

There is another presentation of $W(k)$, related more closely to $k$, which is often useful. Let us write $Q(k)$ for the group of square classes of $k$:

$$Q(k) = k^\times / k^{\times 2}. \tag{8}$$

We shall use multiplicative notation for $Q(k)$, writing $\langle a\rangle$ for the class of $a \in k^\times$. Each element of $Q(k)$ corresponds to a one-dimensional quadratic form and the product in $Q(k)$ corresponds to the tensor product of forms. Hence we have a homomorphism of $Q(k)$ into the multiplicative group of units of $W(k)$;likewise there is a homomorphism of $Q(k)$ into $\mathbf{U}(\hat{W}(k))$. It follows that we have ring homomorphisms from the group ring of $Q(k)$. Since every quadratic space has an orthogonal basis, it follows that $\varphi$ and $\hat{\varphi}$ are surjective.




THEOREM 4.4 *Let $k$ be a field of characteristic not 2 and define $Q(k)$ as in* (8). *Then the Witt–Grothendieck ring and the Witt ring can be described in terms of the group ring $\mathbf{Z}Q(k)$ by*

$$\hat{W}(k) \cong \mathbf{Z}Q(k)/\mathfrak{a}, \qquad W(k) \cong \mathbf{Z}Q(k)/\mathfrak{b}, \tag{9}$$

*where $\mathfrak{a}$ is the ideal generated by $(1 + \langle a\rangle)(1 - \langle 1 + a\rangle)$, $a \neq 0, -1$, and $\mathfrak{b}$ is the ideal generated by $\mathfrak{a}$ together with $1 + \langle -1\rangle$.*

*Proof.* It is clear that (9) holds with $\mathfrak{a} = \ker \hat{\varphi}, \mathfrak{b} = \ker \varphi$ and it remains to determine these ideals. Now the proof of Th. 4.1 shows that $\mathfrak{a}$ is generated by $\langle a\rangle + \langle b\rangle - \langle c\rangle - \langle d\rangle$, where $\langle a, b\rangle \cong \langle c, d\rangle$, and clearly also by $1 + \langle b\rangle - \langle c\rangle - \langle d\rangle$, where $\langle 1, b\rangle \cong \langle c, d\rangle$. Here $\langle b\rangle = \langle cd\rangle$, since the determinants are equal, so also $\langle d\rangle = \langle bc\rangle$, and hence

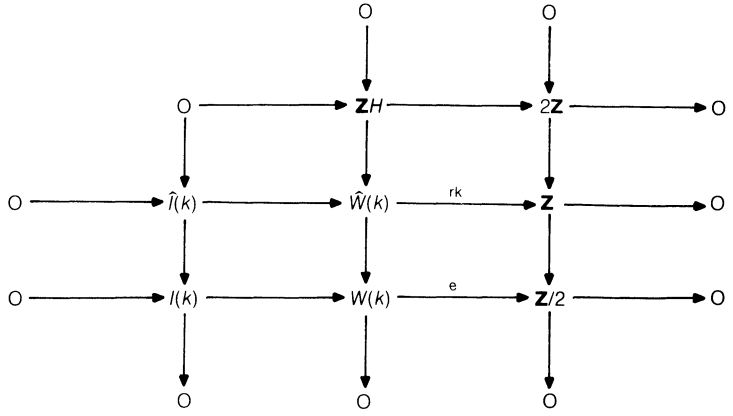
$$1 + \langle b\rangle - \langle c\rangle - \langle d\rangle = (1 + \langle b\rangle)(1 - \langle c\rangle). \tag{10}$$

Now $c = x^2 + y^2 b$; if $x = 0$, then $\langle c\rangle = \langle b\rangle$ and (10) reduces to 0. Otherwise

we have $\langle c\rangle = \langle 1 + a\rangle$, where $a = x^{-2}y^2b$ and $\langle a\rangle = \langle b\rangle$, so the right-hand side of (10) reduces to $(1 + \langle a\rangle)(1 - \langle 1 + a\rangle)$.

Now $W(k) = \hat{W}(k)/\mathbf{Z}(1 + \langle -1\rangle)$, so b is generated by a and $1 + \langle -1\rangle$. ■

The two ideals $\hat{I}(k)$ and $\mathbf{Z}H$ of $\hat{W}(k)$ lead to the following commutative diagram with exact rows and columns:




From it we see that the rank mapping (5) induces a homomorphism $e: W(k) \to \mathbf{Z}/2$, called the *dimension index*. Its kernel is $I(k) = \hat{I}(k)$, the set representing forms of even rank.

By contrast the determinant function on $W(k)$ cannot in general be factored via $\hat{W}(k)$, because $\det(H) \neq 1$ in general (i.e. when $-1$ is not a square). The discriminant $\mathrm{dis}(q)$, defined as in (15) of **8.1**, has the advantage that $\mathrm{dis}(H) = 1$, but we now have $\mathrm{dis}(q \perp q') \neq \mathrm{dis}(q)\cdot\mathrm{dis}(q')$ in general. In order to restore multiplicativity we can proceed as follows. Consider again the group $k^\times/k^{\times 2}$, now written *additively*, and define the set $E(k) = \mathbf{Z}/2 \times k^\times/k^{\times 2}$ as a group by the rule

$$(e, d) + (e', d') = (e + e', d + d' + ee'\{-1\}),$$

where $\{-1\}$ is the class of $-1$. Clearly this is an abelian group with $k^\times/k^{\times 2}$ as a subgroup of index 2. We can also define a multiplication on $E(k)$ by setting

$$(e, d)(e', d') = (ee', ed' + e'd),$$

and with this definition it is easily seen that $E(k)$ is a commutative ring. Now it can be shown that the mapping

$$q \mapsto (e(q), \{\mathrm{dis}(q)\}), \tag{11}$$

where $e(q)$ is the rank of $q \pmod 2$, defines a surjective homomorphism from

$W(k)$ to $E(k)$ with kernel $I(k)^2 = \{\sum a_i b_i | a_i, b_i \in I(k)\}$. This shows that $I(k)^2$ consists of all classes of even-dimensional forms with discriminant 1.

Returning for a moment to Clifford algebras, let us recall that for any quadratic form $q$, either $C$ or $C_0$ is central simple (depending on whether the rank is even or odd). Thus each quadratic form gives rise to an element of the Brauer group $\mathbf{B}_k$, called the *Witt invariant* of $q$. By taking the graded structure of $C$ into account, one obtains a larger group, the *Brauer–Wall group* $\mathbf{BW}_k$. We shall not enter into details here but confine ourselves to the remark that the quotient $\mathbf{BW}_k/\mathbf{B}_k$ is isomorphic to the additive group of $E(k)$ (cf. Lam [1980], Scharlau [1985]).

Another invariant of $q$, equivalent to the Witt invariant, but a little easier to handle, is obtained by writing $q$ in diagonal form: $q = \langle a_1, \ldots, a_n \rangle$ and taking the element of $\mathbf{B}_k$ corresponding to

$$\prod_{i<j} (a_i, a_j; k).$$

This is called the *Hasse invariant*; it is not hard to verify (using Th. 4.2) that this is in fact an invariant of $q$. It can be shown that two forms of the same dimension $\leqslant 3$ are isometric iff they have the same determinant and the same Hasse invariant (cf. Ex. (11)).

The ideal structure of $W(k)$ is closely related to the ways of ordering $k$, and to end this section we describe the connexion.

Let $k$ be an ordered field and $V$ a quadratic space over $k$; it is clear that the signature $\sigma(V)$ depends only on the class of $V$ in $W(k)$, so that we have a mapping

$$\sigma: W(k) \to \mathbf{Z}. \tag{12}$$

Moreover, since $\sigma(U \perp V) = \sigma(U) + \sigma(V), \sigma(\langle ab \rangle) = \sigma(\langle a \rangle)\sigma(\langle b \rangle), \sigma(\langle 1 \rangle) = 1$, it follows that (12) is a ring homomorphism. Its kernel is generated by the Witt classes of all forms $\langle 1, -a \rangle = 1 - \langle a \rangle$ such that $a > 0$. In particular, if $k$ is Euclidean, then every positive element is a square, so all such forms split and $\sigma$ is an isomorphism. Thus we have

PROPOSITION 4.5 *For any Euclidean field $k$ the signature mapping* (12) *is an isomorphism; in particular this holds for every real closed field.* ■

From Th. 3.5 we know that any ordered field $k$ has a real closure $K$ and it is clear that the inclusion $i: k \to K$ induces a homomorphism $i^*: W(k) \to W(K)$. But $W(K) \cong \mathbf{Z}$, by Prop. 4.5, hence $W(k)/\ker i^* \cong \mathbf{Z}$. Thus each ordering of $k$ determines a prime ideal of $W(k)$ with residue class ring $\mathbf{Z}$. The full connexion between orderings and prime ideals is described in Th. 4.7 below, but first we need a lemma.

LEMMA 4.6 *Let* $\mathfrak{p}$ *be a prime ideal in* $W(k)$, *for any field* $k$. *Then* $W(k)/\mathfrak{p}$ *is isomorphic to* $\mathbf{Z}$ *or to* $\mathbf{Z}/p$, *for some prime number* $p$.

*Proof.* $A = W(k)/\mathfrak{p}$ is an integral domain by definition of $\mathfrak{p}$, and for any $a \in k^\times$, $(\langle a\rangle + 1)(\langle a\rangle - 1) = \langle a^2\rangle - 1 = 0$, hence $\langle a\rangle \equiv \pm 1 \pmod{\mathfrak{p}}$. Since $W(k)$ is additively generated by $\langle a\rangle, a \in k^\times$, the additive group of $A$ is cyclic and hence is a homomorphic image of $\mathbf{Z}$, as claimed. ■

A prime ideal $\mathfrak{p}$ in $W(k)$ is said to be of *characteristic* 0 or $p$ according as $W(k)/\mathfrak{p}$ is isomorphic to $\mathbf{Z}$ or $\mathbf{Z}/p$.

THEOREM 4.7 (Harrison–Lorenz–Leicht) *Let* $k$ *be any field and* $W(k)$ *its Witt ring with augmentation ideal* $I(k)$. *Then* $I(k)$ *is the unique prime ideal of characteristic* 2 *in* $W(k)$. *For an odd prime* $p$ *there is a natural bijection between* (i) *the orderings of* $k$, (ii) *the prime ideals of characteristic* 0 *in* $W(k)$ *and* (iii) *the prime ideals of characteristic* $p$ *in* $W(k)$.

Thus when $k$ is formally real, every minimal prime ideal of $W(k)$ or equivalently, every ring homomorphism $W(k) \to \mathbf{Z}$ is the signature homomorphism for some ordering of $k$.

*Proof.* Let $\mathfrak{p}$ be a prime ideal of characteristic 2 in $W(k)$. Then $\langle a^2\rangle \equiv 1 \pmod{\mathfrak{p}}$, hence $\langle a\rangle \equiv 1 \pmod{\mathfrak{p}}$ for all $a \in k^\times$, because $\mathfrak{p}$ is prime, and so $I(k) \subseteq \mathfrak{p}$. Since $I(k)$ is maximal, it follows that $I(k) = \mathfrak{p}$, so $I(k)$ is the unique prime ideal of characteristic 2 in $W(k)$.

In the remark after Prop. 4.5 we have seen how to associate with each ordering $P$ of $k$ a prime ideal $\mathfrak{p}$ of characteristic 0 in $W(k)$. Since $W(k)/\mathfrak{p} \cong \mathbf{Z}$, the ideal $p\mathbf{Z}$ in $\mathbf{Z}$ corresponds to a prime ideal $\bar{\mathfrak{p}}$ of characteristic $p$ in $W(k)$. Now let $\mathfrak{p}$ be any prime ideal of odd characteristic $p$ in $W(k)$ and define

$$P = \{a \in k^\times \mid \langle a\rangle \equiv 1 \pmod{\mathfrak{p}}\}.$$

We claim that $P$ is the positive cone of an ordering on $k$. It is clear that $0 \notin P$ and that $P$ is closed under multiplication. Moreover, any $a \in k^\times$ satisfies $\langle a\rangle \equiv \pm 1 \pmod{\mathfrak{p}}$, so either $a$ or $-a$ is in $P$, but not both. Further, if $a, b \in P$ and $c = a + b$, then $c \neq 0$, for otherwise $\langle b\rangle = \langle -a\rangle \equiv -1 \pmod{\mathfrak{p}}$, i.e. $2 \equiv 0 \pmod{\mathfrak{p}}$, which contradicts the fact that $p$ is odd. Now $\langle a\rangle + \langle b\rangle = \langle c\rangle(1 + \langle ab\rangle)$ by **W**.3, hence $2\langle c\rangle = 2$, and so $\langle c\rangle \equiv 1 \pmod{\mathfrak{p}}$; this shows that $P$ is closed under addition, and therefore defines an ordering on $k$.

It only remains to show that the mappings $P \mapsto \mathfrak{p}, \mathfrak{p} \mapsto \bar{\mathfrak{p}}, \bar{\mathfrak{p}} \mapsto P$ when carried out in succession lead back to the same ordering resp. prime ideal; this is straightforward and may be left to the reader. ■

Let $k$ be a field such that $W(k) \cong \mathbf{Z}$; then $k$ can be ordered and the signature

$\sigma$ is an isomorphism, hence $\langle a \rangle = 1$ for all $a > 0$. Thus every positive element of $k$ is a square, i.e. $k$ is Euclidean. Together with Prop. 4.5 this proves

COROLLARY 4.8 *A field $k$ is Euclidean if and only if $W(k) \cong \mathbf{Z}$.* ■

## Exercises

(1) Verify the homomorphism (11) of $W(k)$ into $E(k)$ and show that the kernel is $I(k)^2$. (Hint: define a map $E(k) \to W(k)/I^2$ by $(0, \{a\}) \mapsto \langle 1, -a \rangle + I^2, (1, \{a\}) \mapsto \langle a \rangle + I^2$.)

(2) Show that $E(k)$ splits over $Q(k)$ (cf. (8)) iff $-1$ is a square in $k$.

(3) Show that the inverse image of $Q(k)$ in the homomorphism (11) is $I = I(k)$, hence that $I/I^2$ is a $W(k)/I$-module. Deduce that $W(k)$ is Noetherian iff $Q(k)$ is finite. (Hint: recall that $W(k)$ is additively generated by the $\langle a \rangle, a \in k^\times$.)

(4) Show that the Witt ring $W(k)$ has no idempotents apart from $0, 1$.

(5) Show that if $k$ is real closed then $\hat{W}(k) \cong \mathbf{Z} \oplus \mathbf{Z}$, and determine the multiplication. Determine $\hat{W}(k)$ when (i) $k$ is finite, (ii) every element of $k$ is a square.

(6) Show that $k$ is not formally real iff $W(k)$ is a local ring, and the unique maximal ideal is then $I(k)$.

(7) Show that $k$ is formally real iff $W(k)$ is not a torsion group.

(8) A field is said to be *pythagorean* if every sum of squares is a square. Show that $k$ is pythagorean iff $W(k)$ is torsion free.

(9) Let $A$ be an anisotropic space. Show that if $U$ and $U \perp A$ are both split, then $A = 0$. Deduce that $V$ corresponds to the 0 of $W(k)$ iff $V$ is split.

(10) Show that the Witt ring of the $p$-adic number field $\mathbf{Q}_p$ is given by $W(\mathbf{Q}_p) = W(\mathbf{F}_p) \oplus W(\mathbf{F}_p)$ for $p \neq 2$, $W(\mathbf{Q}_2) = \mathbf{C}_8 \oplus \mathbf{C}_2 \oplus \mathbf{C}_2$.

(11) Show that any regular ternary (i.e. three-dimensional) quadratic form of determinant $-1$ can be written $\langle a, b, -ab \rangle$. Verify that its Hasse invariant is $(a, b; k)$ and deduce that two ternary forms of the same determinant and the same Hasse invariant are isometric.

(12) Show that a field $k$ is Euclidean iff it is formally real and $|Q(k)| = 2$.

## 8.5 The symplectic group

Let $k$ be a field. By a *symplectic space* $V$ over $k$ we understand a vector space $V$ over $k$ with a regular bilinear form $b$ which is *alternating*, i.e. such

that

$$b(x,x)=0 \qquad \text{for all } x\in V. \tag{1}$$

By linearization we find that $b$ is antisymmetric: $b(y,x)=-b(x,y)$, and relative to a basis of $V$, $b$ is described by an *alternating* matrix, i.e. a matrix $A$ such that $A^T=-A$.

Symplectic spaces are of some importance in the Hamiltonian theory of dynamics. Their theory is considerably simpler than that of inner product spaces, for as we saw in A.1, **8.6**, every symplectic space is even-dimensional, say $n=2m$, and it has a basis of the form $u_1,\dots,u_m,v_1,\dots,v_m$ such that $b(u_i,v_j)=\delta_{ij}, b(u_i,u_j)=b(v_i,v_j)=0$; such a basis is said to be *symplectic*. Any pair of vectors $u,v$ satisfying $b(u,v)=1$ is called *hyperbolic*; thus a symplectic basis consists of mutually orthogonal hyperbolic pairs. Relative to a symplectic basis the matrix of a symplectic form becomes

$$J=\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{2}$$

Let us again derive the existence of a symplectic basis in a slightly more general form. Any two-dimensional symplectic space clearly has a basis consisting of a hyperbolic pair; it is called a *hyperbolic plane.*

LEMMA 5.1 *Let $V$ be a symplectic space, $U$ any subspace and $U_0$ a maximal totally isotropic subspace of $U$. Then* $\dim U\leqslant 2\dim U_0$ *and any basis $u_1,\dots,u_r$ of $U_0$ can (after suitable renumbering) be completed to a basis $u_1,\dots,u_r,v_1,\dots,v_s$ of $U$ such that $(u_i,v_i)$ $(i=1,\dots,s\leqslant r)$ are mutually orthogonal hyperbolic pairs. The basis of $U$ can be completed to a symplectic basis of $V$; in particular, $V$ is an orthogonal sum of hyperbolic planes.*

*Proof.* Let $U$, $U_0$ be as stated and let $u_1,\dots,u_r$ be a basis of $U_0$. If $U_0\neq U$, then any vector of $U\backslash U_0$ cannot be orthogonal to all of $u_1,\dots,u_r$, so we can choose $v_1\in U$ such that $b(u_i,v_1)=0$ for all $i$ except one, say (by renumbering the $u$'s), $b(u_1,v_1)\neq 0$. On multiplying $v_1$ by a suitable scalar we may assume that $b(u_1,v_1)=1$. If $\langle U_0,v_1\rangle\neq U$, we can repeat the process and after a finite number of steps we reach the required basis of $U$. By the maximality of $U$ we have $s\leqslant r$, hence $\dim U=r+s\leqslant 2r$.

Now if $s<r$ we can in $V$ find $v_r$ such that $b(u_i,v_r)=\delta_{ir}$ (because $V$ is regular), and continuing in this way we obtain $u_1,\dots,u_r,v_1,\dots,v_r$, a symplectic basis for a subspace $W$ of $V$. If $W\neq V$, we have $V=W\perp W^\perp$ and by induction on $\dim V$ we find a symplectic basis of $W^\perp$ which together with the basis of $W$ already found is a symplectic basis of $V$. ■

The isometries of $V$ are also called *symplectic mappings*; they form a group

$\mathbf{Sp}_{2m}(k)$, called the *symplectic group*. If we refer all our mappings to a fixed symplectic basis, we see that the symplectic mappings are described by matrices $P$ such that

$$PJP^T = J, \qquad \text{where } J \text{ is as in (2).} \tag{3}$$

A matrix $P$ satisfying (3) is said to be *symplectic*.

As for linear groups we can define the *projective* symplectic group $\mathbf{PSp}_{2m}(k) = \mathbf{Sp}_{2m}(k)/C$, where $C$ is the centre; we shall see in Cor. 5.7 below that $C = \{\pm 1\}$. Our main task in this section will be to show that $\mathbf{PSp}_{2m}(k)$ is simple (again with a few exceptions).

As examples of symplectic matrices we have

$$S_P = \begin{pmatrix} P & 0 \\ 0 & (P^T)^{-1} \end{pmatrix}, \qquad P \in \mathbf{GL}_m(k), \tag{4}$$

and

$$R_Q = \begin{pmatrix} I & Q \\ 0 & I \end{pmatrix} \qquad \text{where } Q \in k_m,\ Q^T = Q. \tag{5}$$

It is easily verified that $S_P$, $R_Q$, $R_Q^T$ all lie in $\mathbf{Sp}_{2m}(k)$. In fact, these matrices form a generating set:

LEMMA 5.2 *The symplectic group* $\mathbf{Sp}_{2m}(k)$ *is generated by the matrices* $S_P (P \in \mathbf{GL}_m(k))$ *and* $R_Q$, $R_Q^T$ $(Q \in k_m,\ Q^T = Q)$.

*Proof.* Consider the general matrix in $\mathbf{Sp}_{2m}(k)$:

$$F = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad A, B, C, D \in k_m. \tag{6}$$

The condition $F \in \mathbf{Sp}_{2m}(k)$ means that $FJF^T = J$. On writing this out, we obtain the equations

$$AB^T = BA^T, \qquad CD^T = DC^T, \qquad AD^T - BC^T = I. \tag{7}$$

Suppose first that $A$ is non-singular. Then it is easily checked that*

$$F = R^T_{CA^{-1}} R_{BA^T} S_A. \tag{8}$$

In the general case we use induction on the nullity of $A$ (i.e. $m - r$, where $r$ is the rank of $A$). By left and right multiplication by suitable matrices $S_P$ we can reduce $F$ to the form

$$F_1 = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \qquad \text{where } A_1 = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}.$$

*It follows from (8) that $CA^{-1}$ is symmetric, but this is also easily proved directly (cf. Ex. (4)).

The condition that $B_1 A_1^T$ is symmetric means that $B_1$ has the form

$$B_1 = \begin{pmatrix} G & H \\ 0 & K \end{pmatrix}, \qquad \text{where } G \text{ is symmetric.}$$

Since $F_1$ is non-singular, it follows that $K \neq 0$. If we multiply $F_1$ on the right by $R_Q^T$, where $Q = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}$, partitioned in conformity with $A_1$, we obtain

$$F_2 = \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix}, \qquad \text{where } A_2 = \begin{pmatrix} I & H \\ 0 & K \end{pmatrix},$$

and this has smaller nullity than $A$. Now the result follows by induction. ■

For any $u \neq 0$ in $V$ and any $c \in k$ the linear mapping defined by

$$\tau_{u,c}: x \mapsto x + cb(x,u)u \tag{9}$$

is a transvection along $u$, which is easily verified to be symplectic; its properties are given by

LEMMA 5.3 *Let $V$ be a symplectic space and $u$ any non-zero vector in $V$. Then any transvection which is symplectic, with kernel $u^\perp$, has the form $\tau_{u,c}$ for some $c \in k$. Moreover,* (i) *for fixed $u$ the mapping $c \mapsto \tau_{u,c}$ is an injective homomorphism of the additive group of $k$ into* $\mathbf{Sp}_{2m}(k)$, (ii) *for any $\sigma \in \mathbf{Sp}_{2m}(k)$, $\sigma^{-1}\tau_{u,c}\sigma = \tau_{u\sigma,c}$ and* (iii) *for $a \in k^\times$, $\tau_{au,c} = \tau_{u,a^2c}$.*

*Proof.* Any linear functional on $V$ with kernel $u^\perp$ is a multiple of $b(-,u)$, hence the symplectic transformation with kernel $u^\perp$ has the form $\tau_{u,c}$ for some $c \in k$. The other properties are verified without difficulty. ■

In a special case the symplectic group is easily identified:

LEMMA 5.4 *For any field $k$, $\mathbf{Sp}_2(k) \cong \mathbf{SL}_2(k)$.*

*Proof.* Let $u, v$ be a symplectic basis. Any linear transformation has the form

$$u \mapsto u' = au + bv,$$
$$v \mapsto v' = cu + dv,$$

and this will be symplectic iff $b(u', v') = 1$, i.e. $ad - bc = 1$. Hence $\mathbf{Sp}_2(k)$ consists precisely of the matrices with determinant 1, as claimed. ■

As a first step we find another generating set for $\mathbf{Sp}_{2m}(k)$:

THEOREM 5.5 *For any field $k$ and any $m \geqslant 1$, $\mathbf{Sp}_{2m}(k)$ is transitive on the hyperbolic pairs, and is generated by the set of all symplectic transvections.*

*Proof.* We denote by $T$ the subgroup generated by all symplectic transvections and divide the proof into three parts:

(i) $T$ is transitive on the non-zero vectors of $V$. For if $u_1, u_2 \neq 0$ and $b(u_1, u_2) \neq 0$, let $c \in k$ be such that $cb(u_1, u_2) = 1$ and put $u = u_1 - u_2$. Then $\tau_{u,c}$ maps $u_1$ to $u_1 + cb(u_1, u_1 - u_2)(u_1 - u_2) = u_1 - (u_1 - u_2) = u_2$. If $b(u_1, u_2) = 0$ and $v \in V$ is such that $b(u_i, v) \neq 0$ $(i = 1, 2)$, then the result just proved yields transvections to map $u_1$ to $v$ and $v$ to $u_2$, so it only remains to find $v$. If $u_1, u_2$ are linearly dependent, we can take any $v$ not in $u^{\perp}$. Otherwise $u_1, u_2$ are linearly independent and such that $b(u_1, u_2) = 0$; then we can, by Lemma 5.1 find $v_1, v_2$ such that $(u_1, v_1)$ and $(u_2, v_2)$ are orthogonal hyperbolic pairs, and now $v = v_1 + v_2$ is the required vector. Thus $T$ has been shown to be transitive on the non-zero vectors of $V$.

(ii) Next we show that $T$ is transitive on the hyperbolic pairs in $V$. Let $(u_i, v_i)$ $(i = 1, 2)$ be two such pairs. By what has been proved we may take $u_1 = u_2 = u$. If $b(v_1, v_2) \neq 0$, then (as we saw in (i)) there is a symplectic transvection $\tau$ along $v_2 - v_1$ with $v_1\tau = v_2$; since $b(u, v_1 - v_2) = b(u, v_1) - b(u, v_2) = 0$, we have $u\tau = u$, so $\tau$ maps $(u, v_1)$ to $(u, v_2)$. If $b(v_1, v_2) = 0$, we use the hyperbolic pair $(u, v_1 + u)$; since $-b(v_1, v_1 + u) = b(v_1 + u, v_2) = 1$, we can find symplectic transvections to map $(u, v_1)$ to $(u, v_1 + u)$ and $(u, v_1 + u)$ to $(u, v_2)$.

(iii) We now use induction on $m$ to show that $T = \mathbf{Sp}_{2m}(k)$. For $m = 1$ the result follows by Lemma 5.4 and Th. 4.7.1, so we may assume $m > 1$. Let $\sigma \in \mathbf{Sp}_{2m}(k)$ and take any hyperbolic pair $(u, v)$ in $V$. By (ii) there exists $\tau \in T$ such that $(u, v)\sigma = (u, v)\tau$, hence $\sigma\tau^{-1}$ leaves $u, v$ fixed. Therefore it maps $W = \langle u, v \rangle^{\perp}$ into itself and defines an isometry there. By the induction hypothesis, $\sigma\tau^{-1}|W = \prod \tau_i'$, where $\tau_i'$ is a symplectic transvection on $W$. We can extend $\tau_i'$ to a symplectic transvection $\tau_i$ on $V$ by defining it as the identity on $\langle u, v \rangle$. Then $\sigma = \prod \tau_i \cdot \tau$, and this shows that $\sigma \in T$. ■

Since a symplectic transvection clearly has determinant 1, we obtain a result which was already proved (by a quite different method) in A.1., Th. 2 of **8.6**; we note that the result also follows easily from Lemma 5.2.

COROLLARY 5.6 *Every symplectic transformation has determinant* 1. ■

Th. 5.5 also allows us to determine the centre of $\mathbf{Sp}_{2m}(k)$:

COROLLARY 5.7 *The centre of $\mathbf{Sp}_{2m}(k)$ consists of the transformations* 1 *and* $-1$.

*Proof.* If $\tau$ is a symplectic transvection along $u$, then $x\tau - x$ is proportional to $u$, hence $x\tau \cdot \sigma - x\sigma$ is proportional to $u\sigma$, for any $\sigma \in \mathbf{Sp}_{2m}(k)$. Now $x\sigma \cdot \tau - x\sigma$ is

proportional to $u$ and not always zero, so if $\sigma$ and $\tau$ commute, then $u\sigma$ must be proportional to $u$. For a given $\sigma$ this can happen for all $u$ only if $\sigma$ is a scalar, say $\sigma = c.1$. Now the condition (3) shows that $c^2 = 1$, hence $c = \pm 1$. ■

We next determine the commutator structure of the symplectic group; we recall that a group is said to be *perfect* if it is equal to its derived group.

THEOREM 5.8 $\mathbf{Sp}_{2m}(k)$ *is perfect except when* $m = 1$ *and* $|k| \leqslant 3$ *or* $m = 2$ *and* $|k| = 2$.

*Proof.* For $m = 1$ this follows from Lemma 5.4 and Prop. 4.7.2, so we may take $m \geqslant 2$ for the rest of the proof. Suppose first that $|k| > 3$; we shall show that every symplectic transvection $\tau_{u,a}$ is a commutator. In $k$ there exists $c \neq 0$ such that $c^2 \neq 1$. Write $b = (1 - c^2)^{-1}a$, $d = -c^2 b$; then $b + d = a$, hence $\tau_{u,a} = \tau_{u,b}\tau_{u,d}$. If $\sigma$ is any symplectic mapping such that $u\sigma = cu$ (Th. 5.5), then

$$\sigma^{-1}\tau_{u,b}^{-1}\sigma = \sigma^{-1}\tau_{u,-b}\sigma = \tau_{u\sigma,-b} = \tau_{cu,-b} = \tau_{u,-c^2 b} = \tau_{u,d}.$$

Hence

$$\tau_{u,a} = \tau_{u,d}\tau_{u,b} = \sigma^{-1}\tau_{u,b}^{-1}\sigma\tau_{u,b}, \tag{10}$$

and this is the required expression.

There remains the case where $k$ has two or three elements. It will be enough to find a transvection $\tau_{u,a} \neq 1$ in the derived group, for since $\mathbf{Sp}_{2m}(k)$ is transitive on the non-zero vectors of $V$, we then have $\sigma^{-1}\tau_{u,a}\sigma = \tau_{u\sigma,a}$ in the derived group, and in case $|k| = 3$, $\tau_{u,2a} = \tau_{u,a}^2$, so it follows that the derived group contains all transvections and so by Th. 5.5, coincides with $\mathbf{Sp}_{2m}(k)$. We shall write our transformations as matrices relative to a symplectic basis. Let $A$, $B \in k_m$; then if $A$ is invertible, $S_A$ is symplectic and if $B$ is symmetric $R_B$ is symplectic (cf. (4) and (5)). When this is so, we have

$$(S_A, R_B) = R_C, \qquad \text{where } C = B - A^{-1}B(A^T)^{-1}, \tag{11}$$

and for suitable choice of $A, B$ this is a symplectic transvection.

Suppose first that $|k| = 3$, $m = 2$. Taking $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$, we find that $B - A^{-1}B(A^T)^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, so we have obtained a transvection which is a commutator. The same argument applies for $m > 2$, taking $A$ and $B$ as the identity and 0 respectively on the remaining coordinates.

There remains the case $|k| = 2$, $m \geqslant 3$. When $m = 3$, we take

$$A = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

It is easily verified that $B - A^{-1}B(A^T)^{-1} = E_{11}$, so we have again a symplectic transvection which is a commutator; for $m > 3$ we again take $A$, $B$ to be $1, 0$ respectively on the new coordinates. So we have in all cases expressed a transvection as a commutator, and the result follows. ■

Finally we come to the simplicity proof, which runs along similar lines to that for the general linear group (cf. **4.7**).

THEOREM 5.9 *The projective symplectic group* $\mathbf{PSp}_{2m}(k)$ *is simple for all fields* $k$ *and all integers* $m \geqslant 1$, *except when* $m = 1$ *and* $|k| \leqslant 3$ *or* $m = 2$ *and* $|k| = 2$.

*Proof.* By Lemma 5.4 we may assume that $m > 1$. We consider the action of $G = \mathbf{PSp}_{2m}(k)$ on the space $P = \mathbf{P}^{2m-1}(k)$ and begin by showing that $G$ is primitive. Let $Q$ be a set in a partition of $P$ compatible with the $G$-action and containing more than one point. Suppose first that $Q$ contains a pair of points $\langle x\rangle$, $\langle y\rangle$ defined by a hyperbolic pair of vectors $(x, y)$. Given any other point $\langle z\rangle$ in $P$, if $b(x, z) \neq 0$, we may assume that $b(x, z) = 1$; since $G$ is transitive on the hyperbolic pairs, by Th. 5.5, there exists $\sigma \in G$ mapping $(x, y)$ to $(x, z)$, hence $Q\sigma \cap Q \neq \varnothing$, so $Q\sigma = Q$, but $\langle z\rangle \in Q\sigma = Q$. If $b(x, z) = 0$, we may assume that $\langle z\rangle \neq \langle x\rangle$, because otherwise $\langle z\rangle = \langle x\rangle \in Q$. Then there exists $w \in V$ such that $b(x, w) = b(z, w) = 1$, and by what has been shown, $\langle w\rangle \in Q$. Further there exists $\sigma \in G$ mapping $(x, w)$ to $(z, w)$ hence $Q\sigma \cap Q \neq \varnothing$ so again $\langle z\rangle \in Q\sigma = Q$ and it follows that $Q = P$. The alternative is that the subspace defining $Q$ is totally isotropic. Let $\langle x, y\rangle$ be a plane in $Q$ and choose $w \in V$ such that $b(x, w) = 1$, $b(y, w) = 0$. Writing $H = \langle x, w\rangle$, we have $V = H \perp H^{\perp}$. Further, $y \in H^{\perp}$ and for any $0 \neq z \in H^{\perp}$ there exists a symplectic transformation $\sigma$ leaving $H$ fixed and mapping $y$ to $z$. Since $\langle x\rangle \in Q$, it follows that $Q\sigma = Q$ and since $\langle y\rangle \in Q$, we have $\langle z\rangle \in Q\sigma = Q$. Hence $Q$ contains all points defined by vectors in $H^{\perp}$. Since $m > 1$, $H^{\perp}$ contains a hyperbolic pair, so the first part of the argument can be applied to show that $Q = P$, and this shows $G$ to be primitive.

In order to apply Lemma 4.7.4, we need to find a normal abelian subgroup of a stabilizer whose conjugates generate $G$. Take a point $\langle x\rangle$, let $S$ be its stabilizer and let $A$ be the subgroup of transvections $\tau_{x,a}(a \in k)$. Then (10) shows that $A \lhd S$ and by Th. 5.5, $A$ and its conjugates generate $G$. Hence $G$ is indeed simple, with the exceptions listed. ■

Again the exceptions actually occur (cf. Ex. (2), (3)).

It still remains to compute the order of $\mathbf{Sp}_{2m}(k)$ when $k$ is finite.

PROPOSITION 5.10 *If* $k$ *is a field of* $q$ *elements, then*

$$|\mathbf{Sp}_{2m}(k)| = q^{2m-1}(q^{2m} - 1)\cdot|\mathbf{Sp}_{2m-2}(k)|, \tag{12}$$

*hence*

$$|\mathbf{Sp}_{2m}(k)| = q^{2m-1}(q^{2m}-1)q^{2m-3}(q^{2m-2}-1)\cdots q(q^2-1). \tag{13}$$

*For* $\mathbf{PSp}_{2m}(k)$ *the order is the same when* $q$ *is even and has one half this value when* $q$ *is odd.*

*Proof.* Let $V$ be a two-dimensional symplectic space over $k$; we first determine the number of hyperbolic pairs in $V$. For a hyperbolic pair $(x, y)$, $x$ may be any non-zero vector in $V$, so there are $q^{2m}-1$ choices. Given a particular hyperbolic pair $(x, y_0)$, any other hyperbolic pair with first vector $x$ has the form $(x, y)$, where $y = y_0 + z$ for a vector $z \in x^{\perp}$, and here we have $q^{2m-1}$ choices for $z$. Hence there are $q^{2m-1}(q^{2m}-1)$ pairs in all. By Th. 5.5, $\mathbf{Sp}_{2m}(k)$ is transitive on the set of these pairs, and the stabilizer of a particular pair $(x, y)$ is isomorphic to the symplectic group of $\langle x, y\rangle^{\perp}$, which is $\mathbf{Sp}_{2m-2}(k)$. Hence we obtain (12) and now (13) is an easy consequence. The final remark follows because the centre is $\{\pm 1\}$, as we saw in Cor. 5.7, and $-1 = 1$ when $q$ is even. ■

## Exercises

(1) Show that in the action of $\mathbf{PSp}_{2m}(k)$ on $\mathbf{P}^{2m-1}(k)$, the stabilizer of a point has exactly three orbits.

(2) Show that $\mathbf{PSp}_2(\mathbf{F}_2) \cong S_3$, $\mathbf{PSp}_2(\mathbf{F}_3) \cong A_4$.

(3) Let $V$ be a four-dimensional symplectic space over $\mathbf{F}_2$ and define a *pentagon* as a 5-tuple of vectors $u_1, \ldots, u_5$ satisfying $b(u_i, u_j) = 1$ for $i \neq j$. Show that (i) every pentagon contains 20 hyperbolic pairs, and (ii) every hyperbolic pair occurs in exactly one pentagon. Deduce that there are exactly 6 pentagons. Use the result to show that $\mathbf{Sp}_4(\mathbf{F}_2) \cong S_6$.

(4) Verify directly that if $F \in \mathbf{Sp}_{2m}(k)$, where $F$ is as in (6), and $A$ is invertible, then $CA^{-1}$ is symmetric. (This is implicit in the formula (8) for $F$.)

(5) Show that $\mathbf{Sp}_{2m}(k)$ contains a subgroup isomorphic to $\mathbf{GL}_m(k)$. (Hint: consider a maximal totally isotropic subspace.)

## 8.6 The orthogonal group

The orthogonal group leads to a class of simple groups, just as the symplectic groups do, but this time the result is dependent on the precise structure of the underlying quadratic space. Our treatment follows Iwasawa and Tamagawa, as in the account of Jacobson [1985]. The field $k$ is again assumed to have characteristic not equal to 2. We begin by determining the centre of $\mathbf{O}(V)$.

PROPOSITION 6.1 *Let* $V$ *be a regular quadratic space. Then the centre of* $\mathbf{O}(V)$ *is* $\{\pm 1\}$.

*Proof.* If $u$ is any anisotropic vector in $V$ and $\sigma_u$ is the symmetry defined by $u$, then $\langle u\rangle = \{x \in V | x\sigma_u = -x\}$; it follows that any $\alpha$ in the centre of $\mathbf{O}(V)$ has the property that $u\alpha \in \langle u\rangle$ for every anisotropic vector $u$. Let $u_1, \ldots, u_n$ be an orthogonal basis of $V$; then $u_i\alpha = \pm u_i$, so by suitable numbering we may assume that $u_i\alpha = u_i$ for $i = 1, \ldots, s$ and $u_i\alpha = -u_i$ for $i > s$. If $u_i + u_j$ is anisotropic, we have $(u_i + u_j)\alpha = \pm(u_i + u_j)$, and it follows that either $i, j \leqslant s$ or $i, j > s$. Thus if $1 \leqslant s < n$, then $u_1 + u_n$ is isotropic; hence for $1 < i < n$, $q(u_1 + u_i + u_n) = q(u_i) \neq 0$, so $(u_1 + u_i + u_n)\alpha = \varepsilon(u_1 + u_i + u_n)$, where $\varepsilon = \pm 1$, but also $(u_1 + u_i + u_n)\alpha = u_1 + \varepsilon' u_i - u_n$ and this leads to a contradiction. Therefore $s$ is 0 or $n$ and $\alpha = 1$ or $-1$. ■

The symmetry $\sigma_u$ is defined only for anisotropic vectors $u$; in the isotropic case one has the following replacement, going back to C. L. Siegel. In what follows we shall assume that $\dim V \geqslant 3$; the case $\dim V = 2$ is that of a hyperbolic plane and is easily dealt with separately.

Let $u$ be an isotropic vector and choose $v$ so that $u$, $v$ is a hyperbolic pair, and $w \in \langle u, v\rangle^\perp$, $w \neq 0$. We have $V = \langle v\rangle \oplus u^\perp$, hence the equations

$$\begin{cases} x\rho_{u,w} = x + b(x, w)u & \text{for } x \in u^\perp, \\ v\rho_{u,w} = v - q(w)u - w, \end{cases} \tag{1}$$

define the linear transformation $\rho_{u,w}$ completely and it is easily checked that $\rho_{u,w}$ is orthogonal. Moreover, it is proper (i.e. of determinant 1), in fact it is *unipotent*, i.e. $1 - \rho_{u,w}$ is nilpotent.

We remark that $\rho_{u,w}$ is uniquely determined as the orthogonal transformation which maps $x$ to $x + b(x, w)u$ for all $x \in u^\perp$. For this mapping can be extended to an orthogonal transformation of $V$, by Witt's theorem (A.2, Th. 6.4.5) and if there were two such mappings, their quotient $\alpha$ would leave $u^\perp$ fixed. Now $v\alpha$ has the form $\lambda u + \mu v + z$, where $z \in \langle u, v\rangle^\perp$. For any $x \in \langle u, v\rangle^\perp$ we have $0 = b(x, v) = b(x\alpha, v\alpha) = b(x, \lambda u + \mu v + z) = b(x, z)$; hence $z$ is also orthogonal to $\langle u, v\rangle^\perp$ and so $z = 0$. Further, $1 = b(u, v) = b(u\alpha, v\alpha) = b(u, \lambda u + \mu v) = \mu$, thus $\mu = 1$ and finally $0 = q(v) = q(v\alpha) = q(\lambda u + v) = \lambda$, hence $\lambda = 0$, so $v\alpha = v$ and therefore $\alpha = 1$. This shows $\rho_{u,w}$ to be uniquely determined by its effect on $u^\perp$.

We shall use the transformation $\rho_{u,w}$ to construct an abelian normal subgroup of the stabilizer of $u$.

LEMMA 6.2 *Let $V$ be a regular quadratic space of dimension $\geqslant 3$, containing a hyperbolic pair $u, v$ and put $W = \langle u, v\rangle^\perp$. Then the set*

$$A_u = \{\rho_{u,w} | w \in W\} \tag{2}$$

*is a subgroup of $\mathbf{O}(V)$ which is abelian and normal in the stabilizer of $u$. Moreover,*

*the mapping*

$$w \mapsto \rho_{u,w} \tag{3}$$

*is an isomorphism of* $W$ *with* $A_u$.

*Proof.* Given $w, w' \in W$, we see from (1) that $\rho_{u,w}\rho_{u,w'}$ and $\rho_{u,w+w'}$ both map $x$ to $x + b(x, w + w')u$, for any $x \in u^\perp$, hence they agree on the whole of $V$. This shows (3) to be a homomorphism. If $w$ lies in the kernel of (3), then by (1), $w \in u^{\perp\perp} = \langle u \rangle$ (cf. A.2, Prop. 6.1.3), but the only multiple of $u$ in $W$ is 0; this shows (3) to be injective. It follows that $A_u$ is an abelian subgroup of $\mathbf{O}(V)$. Clearly it is contained in the stabilizer of $u$. Moreover, for any $\sigma \in \mathbf{O}(V)$ we have $\sigma^{-1}\rho_{u,w}\sigma = \rho_{u\sigma,w\sigma}$, hence $A_u$ is mapped into itself by any $\sigma$ leaving $u$ fixed. ■

Our next aim will be to show that under some mild restriction on $V$, the derived group $\mathbf{O}(V)'$ is generated by all the $A_u$. That some restriction is needed is clear since there are no $A_u$ unless the Witt index of $V$ is positive. It will be convenient to write $\Omega$ for the subgroup of $\mathbf{O}(V)$ generated by all the $A_u$, where $u$ ranges over the isotropic vectors. We begin by establishing some transitivity properties.

LEMMA 6.3 *Let* $V$ *be a regular quadratic space,* $\dim V \geqslant 3$. *Then* (i) *for any isotropic* $u \in V$, $A_u$ *is transitive on the one-dimensional isotropic subspaces not orthogonal to* $u$, (ii) *for any two linearly independent isotropic vectors* $u_1, u_2$ *there is a vector* $v$ *and* $\lambda_1, \lambda_2 \in k$ *such that* $(\lambda_i u_i, v)$ *is a hyperbolic pair,* (iii) *the subgroup* $\Omega$ *of* $\mathbf{O}(V)$ *generated by the* $A_u$ *is transitive on the hyperbolic pairs.*

*Proof.* (i) Let $x, y$ be isotropic vectors not orthogonal to $u$; we may assume that $b(u, x) = b(u, y) = 1$. Write $y = \lambda u + \mu x + z$, where $z \in \langle u, x \rangle^\perp$; since $b(u, y) = 1$, we have $\mu = 1$ and the relation $q(y) = 0$ shows that $\lambda + q(z) = 0$. Hence $y = x - q(z)u + z = x\rho_{u,-z}$ and (i) follows.

(ii) Let $u_1, u_2$ be as stated; if $b(u_1, u_2) \neq 0$, we may assume that $b(u_1, u_2) = 1$. The space $\langle u_1, u_2 \rangle^\perp$ contains an anisotropic vector $w$, by the regularity of $V$. We put $v = u_1 - q(w)u_2 + w$; then $q(v) = -q(w) + q(w) = 0$, $b(u_1, v) = -q(w) \neq 0$, $b(u_2, v) = 1$, so with $\lambda_1 = -1/q(w)$, $\lambda_2 = 1$, $v$ satisfies the required conditions. If $b(u_1, u_2) = 0$, then since $u_1, u_2$ are linearly independent, there is a linear functional equal to 1 on $u_1, u_2$, hence by the regularity of $b$ there exists $v$ such that $b(u_i, v) = 1$, and by subtracting a suitable multiple of $u_1$ from $v$ we can ensure that $q(v) = 0$. Then $(u_i, v)$ are again two hyperbolic pairs.

(iii) Let $(u_i, v_i)$ $(i = 1, 2)$ be any two hyperbolic pairs. If $u_1, u_2$ are linearly independent, we can by (ii) find $v$ and $\lambda_i$ such that $(\lambda_i u_i, v)$ are hyperbolic pairs; then by (i) there is an element of $A_v$ mapping $\langle u_1 \rangle$ to $\langle u_2 \rangle$. Thus for some $\sigma \in \Omega$, $u_1\sigma = cu_2$. We now have the hyperbolic pairs $(u_1\sigma, v_1\sigma) = (cu_2, v_1\sigma)$ and

$(u_2, v_2)$; applying (i) again, we find $\tau \in A_{u_2}$ such that $v_1\sigma\tau = v_2$, $u_2\tau = u_2$, so $\sigma\tau$ maps $(u_1, v_1)$ to $(cu_2, v_2)$, but $1 = b(u_1, v_1) = b(cu_2, v_2) = c$, hence $c = 1$. If $u_1, u_2$ are linearly dependent, we can by (i) apply an element of $\Omega$ to $(u_1, v_1)$ to obtain a hyperbolic pair with first vector linearly independent of $u_2$ and now proceed as before. ■

We recall from A.2, Cor. 6.3.3 that $\mathbf{O}(V)$ is generated by symmetries (Cartan–Dieudonné theorem). It follows that the square of every orthogonal transformation lies in $\mathbf{O}(V)'$. For if $\alpha = \sigma_1 \cdots \sigma_r$, then $\alpha^2 = \sigma_1 \cdots \sigma_r \sigma_1 \cdots \sigma_r \equiv \sigma_1^2 \cdots \sigma_r^2 \equiv 1 \pmod{\mathbf{O}(V)'}$. We also note that for $\dim V > 2$ we have $\mathbf{O}(V)' = \mathbf{SO}(V)'$ (A.2, Th. 6.3.4). In two dimensions this need not hold; as an example (which is used later) let us determine $\mathbf{O}(V)$ for a hyperbolic plane.

Let $H$ be a hyperbolic plane with the hyperbolic pair $u, v$ as basis. Any linear mapping has the form

$$u' = au + bv, \qquad v' = cu + dv,$$

and this is an isometry iff $q(u') = q(v') = 0$, $b(u', v') = 1$. Thus $ab = cd = 0$, $ad + bc = 1$; hence either $b = c = 0$ or $a = d = 0$, and the isometries of $H$ have the following forms:

$$\lambda_a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \qquad \text{or} \qquad \lambda_a\tau = \begin{pmatrix} 0 & a \\ a^{-1} & 0 \end{pmatrix}, \qquad \text{where } \tau = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (4)$$

The $\lambda_a$ form a group isomorphic to $k^\times$ and $\mathbf{O}(H)$ is an extension by a cyclic group of order 2, acting by inversion: $\tau\lambda_a\tau = \lambda_a^{-1}$. In particular,

$$\lambda_{a^2} = (\tau, \lambda_a). \quad (5)$$

We can now determine the subgroup $\Omega$:

THEOREM 6.4 *Let $V$ be a regular quadratic space of dimension $\geqslant 3$ and positive Witt index $\nu$. Then $\mathbf{SO}(V) = \mathbf{O}(V)'$; more precisely, the subgroup $\Omega$ generated by the $A_u$ ($u$ isotropic) coincides with the derived group: $\Omega = \Omega' = \mathbf{O}(V)'$, except when $n = 4$, $\nu = 2$ and $n = 3$, $|k| = 3$.*

*Proof.* As a first step we show that $\Omega \supseteq \mathbf{O}(V)'$. We fix a hyperbolic pair $(u, v)$ and write $O_1 = \mathbf{O}(\langle u, v\rangle)$, $W = \langle u, v\rangle^\perp$. We claim that every symmetry is conjugate under $\Omega$ to one in $O_1$. For if $x$ is any anisotropic vector, then $x_1 = u + q(x)v$ satisfies $q(x_1) = q(x)$, hence there exists $\rho \in \mathbf{O}(V)$ such that $x_1\rho = x$. By Lemma 6.3(iii) there exists $\alpha \in \Omega$ mapping $(u\rho, v\rho)$ to $(u, v)$; since $x_1 \in \langle u, v\rangle$, we have $x = x_1\rho \in \langle u\rho, v\rho\rangle$ and so $x\alpha = x_1\rho\alpha \in \langle u, v\rangle$. Hence $\alpha^{-1}\sigma_x\alpha = \sigma_{x\alpha}$, where $x\alpha \in \langle u, v\rangle$, as we wished to show.

Now $O_1 = \mathbf{O}(\langle u, v\rangle)$ is generated by symmetries $\sigma_x$, $x \in \langle u, v\rangle$ and the restriction $\sigma_x | W$ is the identity; it follows that the mapping $\sigma_x \mapsto \sigma_x | \langle u, v\rangle$ defines

an isomorphism $O_1 \cong \mathbf{O}(H)$, and clearly $SO_1$ corresponds to $\mathbf{SO}(H)$ in this isomorphism. Now any $\rho \in \mathbf{SO}(H)$ can be written as a product of an even number of symmetries:

$$\rho = \sigma_1 \cdots \sigma_{2r},$$

and we have seen that $\sigma_i = v_i^{-1}\tau_i v_i$, where $\tau_i$ is a symmetry in $O_1$ and $v_i \in \Omega$. Thus we have

$$\rho = (v_1^{-1}\tau_1 v_1)\cdots(v_{2r}^{-1}\tau_{2r}v_{2r}).$$

Since $\Omega$ is normal in $\mathbf{O}(V)$, this relation can be written $\rho = \mu\tau_1 \cdots \tau_{2r}$, where $\mu \in \Omega$, so we have $\mathbf{SO}(V) \subseteq \Omega \cdot SO_1$. But as we saw, $\Omega \subseteq \mathbf{SO}(V)$, and we conclude that

$$\mathbf{SO}(V) = \Omega \cdot SO_1. \tag{6}$$

Therefore $\mathbf{SO}(V)/\Omega \cong SO_1/SO_1 \cap \Omega$, but $SO_1 \cong k^\times$ is abelian, by the earlier remarks, so $\Omega \supseteq \mathbf{SO}(V)' = \mathbf{O}(V)'$ and we find that

$$\Omega \supseteq \mathbf{SO}(V)'. \tag{7}$$

To complete the proof we show that $\Omega$ is perfect, for then $\Omega = \Omega' = \mathbf{SO}(V)'$. It will be enough to show that $\rho_{u,w} \in \Omega'$ for all isotropic $u$ and all $w$ orthogonal to a hyperbolic plane containing $u$. So we may take $u, v, W, O_1$ as before. Let $\lambda_a, \tau$ be the transformations in $O_1$ defined as in (4), so that $\lambda_{a^2} \in \Omega$ by (5) and (7). For any $w \in W$ we have

$$\lambda_{a^2}^{-1}\rho_{u,w}^{-1}\lambda_{a^2}\rho_{u,w} = \rho_{a^2u,-w}\rho_{u,w} = \rho_{u,-a^2w}\rho_{u,w} = \rho_{u,(1-a^2)w}.$$

When $|k| > 3$, we can choose $a \in k^\times$ such that $a^2 \neq 1$ and replacing $w$ by $(1-a^2)^{-1}w$ we see that $\rho_{u,w} \in \Omega'$, hence $\Omega = \Omega' = \mathbf{O}(V)'$ when $|k| > 3$.

Suppose now that $|k| = 3$; then $n \geqslant 4$ and for $n = 4$, $v = 1$. Thus $\dim W \geqslant 2$ and we have an orthogonal basis $w_1 = w, w_2, \ldots, w_r$ for $W$. If $q(w_1) = q(w_2)$, then the map $\alpha: w_1, w_2, \ldots, w_r \mapsto -w_2, w_1, w_3, \ldots, w_r$ is an isometry, hence $\alpha^2 \in \mathbf{O}(V)' \subseteq \Omega$ and $\alpha^2$ maps $w$ to $-w$. Thus

$$\rho_{u,w}^{-1}\alpha^{-2}\rho_{u,w}\alpha^2 = \rho_{u,-w}\rho_{u,-w} = \rho_{u,-2w} = \rho_{u,w},$$

and this shows that $\rho_{u,w} \in \Omega'$.

It remains to justify the assumption $q(w_1) = q(w_2)$. Since $q(w_1) \in k^\times$ and $|k| = 3$, $q(w_1)$ is 1 or $-1$. For $n = 4$, $W$ is isotropic, so $q(w_1)$, $q(w_2)$ have the same sign and hence must be equal. When $n \geqslant 5$, $W \cap w^\perp$ is regular and at least two-dimensional, so $q$ restricted to this space is universal (A.2, Th. 6.2.7), and it follows that $W$ contains $w'$ orthogonal to $w$ such that $q(w') = q(w)$. Thus we can in all cases find a basis of the required form. ■

We now have the means at our disposal to prove the main structure theorem

for orthogonal groups. The result was first established, with a restriction on the index, by Dickson [1901], and in its full generality by Dieudonné in 1940.

THEOREM 6.5 *Let $V$ be a regular quadratic space of dimension $n \geqslant 3$ and of positive Witt index $v$. Denote by $C$ the centre of $\mathbf{SO}(V)$. Then $\mathbf{SO}(V)/C$ is simple, except when $n = 4$, $v = 2$ and $n = 3$, $|k| = 3$.*

*Proof.* We remark that $C$ has order 2 when $n$ is even and is trivial when $n$ is odd. For by Prop. 6.1 the centre can only contain 1 and $-1$ and $\det(-1) = (-1)^n$.

Consider the quadric cone $Q$ defined as the set of points $\langle x \rangle$ in $\mathbf{P}^{n-1}(k)$ satisfying $q(x) = 0$. By Lemma 6.3(i), $\Omega$ acts transitively on $Q$. We first show that the action is primitive except when $n = 4$, $v = 2$. If $b(x, y) \neq 0$ for any two linearly independent isotropic vectors $x, y$, then by Lemma 6.3(iii), $\Omega$ is 2-fold transitive, and hence primitive. In particular this always holds for $v = 1$, so we may assume henceforth that $v \geqslant 2$ and hence $n \geqslant 5$. Let $S$ be a subset with more than one point in a partition of $Q$ stable under $\Omega$; we have to show that $S = Q$. Given $\langle x_1 \rangle, \langle x_2 \rangle \in S$, if $b(x_1, x_2) = 0$, then there exist $y_1, y_2 \in Q$ such that $(x_1, y_1)$, $(x_2, y_2)$ are orthogonal hyperbolic pairs. By Lemma 6.3(i) applied to $\langle x_2, y_2 \rangle^{\perp}$ there exists $\alpha \in \Omega$ mapping $x_1$ to $y_1$ and leaving $x_2, y_2$ fixed. Since $\langle x_2 \rangle \in S$, we have $S\alpha = S$ and $\langle x_1 \rangle \in S$, therefore $\langle y_1 \rangle \in S$. Now if $\langle z \rangle$ is any point of $Q$ different from $\langle x_1 \rangle$, we can, by Lemma 6.3(ii) find $v$ such that $(x_1, v)$, $(z, v)$ are hyperbolic pairs. By Lemma 6.3(iii) there exists $\beta \in \Omega$ mapping $(x_1, y_1)$ to $(x_1, v)$, hence $S\beta = S$ and $\langle v \rangle \in S\beta = S$. Similarly there is $\gamma \in \Omega$ mapping $(x_1, v)$ to $(z, v)$, hence $\langle v \rangle \in S\gamma = S$ and $z = x_1\gamma$, so $\langle z \rangle \in S$. Since $z$ was arbitrary, this means that $S = Q$.

We may therefore assume that for any distinct points $\langle x_1 \rangle, \langle x_2 \rangle$ in $S$, $b(x_1, x_2) \neq 0$. Given $\langle z \rangle \neq \langle x_1 \rangle$ in $Q$ as before, we find $v$ such that $(x_1, v)$, $(z, v)$ are hyperbolic pairs, and $\beta \in \Omega$ maps $(x_1, x_2)$ to $(x_1, v)$. It follows that $S\beta = S$ and $\langle v \rangle \in S$. If now $\gamma \in \Omega$ is chosen so as to map $(x_1, v)$ to $(z, v)$, then since $v\gamma = v$, we have $S\gamma = S$ and $z = x_1\gamma$, hence $\langle z \rangle \in S$ and we again find that $S = Q$.

We thus see that $\Omega$ acts primitively on $Q$ and $\Omega$ is perfect, by Th. 6.4. Moreover, if $u$ is isotropic, then $A_u$ is a normal abelian subgroup of the stabilizer of $\langle u \rangle$ and the conjugates of $A_u$ generate $\Omega$, by the proof of Th. 6.4. Hence by Lemma 4.7.4, $\Omega$ is simple, and now the result follows by Th. 6.4. ∎

Some of the exceptions in Th. 6.5 will be considered in the exercises. Let us now take up some particular cases to show that the hypotheses on the Witt index cannot be omitted, without striving for full generality. We take a quadratic form over $\mathbf{R}$; if its index is 0, the form must be definite, say positive definite, and in suitable coordinates it will have an orthonormal basis. For simplicity consider the case $n = 3$, thus $\mathbf{SO}(V)$ is the group of rotations in 3-space. We claim that $\mathbf{SO}(V)$ acts primitively on the unit sphere $S$. For let $T$ be a subset of $S$ with more than one point, stable under all rotations. Given $p, q \in T$, $T$ must

include all points of the circle through $q$ about $p$ as axis. If the points at opposite ends of a diameter of this circle are a spherical distance $d$ apart, then $T$ will include points at any distance $\leqslant d$ from $q$, and by repetition, points at any finite (spherical) distance, hence $T = S$ and the action is primitive. The rotations about a point form an abelian subgroup whose conjugacy classes generate $\mathbf{SO}(V)$, and this shows $\mathbf{SO}(V)$ to be simple. The same argument applies for any odd dimension $\geqslant 3$, while for even dimensions $\geqslant 6$, $\mathbf{PSO}(V)$ is simple. When $\dim V = 4$, we have $PO \cong G \times G$, where $G = \mathbf{PSL}_2(k)$ when $V$ has index 2, and $G = \mathbf{SO}(\mathbf{R}^3)$ for a Euclidean 3-space when $V = \mathbf{R}^4$ is Euclidean. When $\mathbf{R}^4$ has index 1 (e.g. the Lorentz metric of relativity theory), then $PO \cong \mathbf{PSL}_2(\mathbf{C})$; in this case $\Omega$ consists of the rotations which do not reverse the time direction.

The argument just used to show that for a definite quadratic form $\mathbf{PSO}(\mathbf{R}^n)$ is simple depended essentially on the fact that $\mathbf{R}$ is Archimedean ordered. For an ordered field $K$ which is non-Archimedean it can be shown that $\mathbf{PSO}(K^n)$ is not simple: the infinitesimal rotations generate a proper normal subgroup (cf. Ex. (4)). This happens e.g. for the field of Laurent series $\mathbf{R}((x))$, ordered by the sign of its first coefficient.

For a finite field it is again possible to calculate the order of the orthogonal group, but this depends on the quadratic character of the determinant as well as the parity of the dimension, cf. Ex. (5)–(7).

## Exercises

(1) Give the details of the proof that for a real Euclidean space $V$ of dimension $\geqslant 5$, $\mathbf{PSO}(V)$ is simple.

(2) Verify that $\rho_{u,w}$ defined by (1) satisfies $(1 - \rho_{u,w})^3 = 0$. When is $(1 - \rho_{u,w})^2 = 0$?

(3) Let $V$ be an $n$-dimensional quadratic space. Given two anisotropic vectors $x, y$, show that $\sigma_x\sigma_y$ is a rotation, in the plane $\langle x, y\rangle$, leaving $\langle x, y\rangle^\perp$ fixed. Verify that when $V$ is Euclidean, then the angle of rotation is twice the angle between $x$ and $y$.

(4) Let $V$ be a Euclidean space over an ordered field $K$ which is non-Archimedean. Show that the rotations through an infinitesimal angle generate a proper normal subgroup ($\alpha$ is *infinitesimal* if $n\alpha < 1$ for all $n \in \mathbf{Z}$).

(5) Show that over a finite field of odd characteristic every regular quadratic form has the form $\langle 1^{n-1}, d\rangle$ where $d$ is the determinant and that $|k^\times/k^{\times 2}| = 2$, so that there are just two classes of forms in each dimension. (Hint: recall (A.2, **6.2**) that a quadratic form of rank $\geqslant 2$ over a finite field is universal.)

(6) Let $|k| = q$ be odd. Show that the number of solutions of $\sum_1^m(x_i^2 - y_i^2) = b$ is $q^{2m-1} - q^{m-1} + \delta_{0b}q^m$, the number of solutions of $\sum_1^m(x_i^2 - y_i^2) - (d-1)y_m^2 = b$ is $q^{2m-1} + q^{m-1} - \delta_{0b}q^m$, and the number of solutions of $\sum_1^m(x_i^2 - y_i^2) - z^2 = b$ is

$q^{2m}+(-b/q)q^m$, where $(-b/q)$ is $0, 1$ or $-1$ according as $-b$ is 0, a non-zero square or not a square in $k$.

(7) Show that for regular forms over a finite field $\mathbf{F}_q$ ($q$ odd), $|\mathbf{O}_{2m}(\mathbf{F}_q)| = (q^{2m-1}-q^{m-1})|\mathbf{O}_{2m-1}(\mathbf{F}_q)|, |\mathbf{O}_{2m+1}(\mathbf{F}_q)| = (q^{2m}+(-d/q)q^m)|\mathbf{O}_{2m}(\mathbf{F}_q)|$, where $d$ is the determinant of the form. Hence calculate the order of the orthogonal group.

(8) What form do the equations (1) take when $u$ is replaced by $-u$? Show that $\rho_{-u,-w} = \rho_{u,w}$ and deduce that the normalizer of $A_u$ includes any $\sigma$ mapping $u$ to $-u$.

## 8.7 Quadratic forms in characteristic two

In the treatment of quadratic forms we have at various points used the possibility of dividing by 2, and we shall now consider the case of characteristic 2, when this is no longer possible. In particular, there is now no longer an equivalence between quadratic forms and symmetric bilinear forms.

Let us first suppose that we have a space $V$ over a field of characteristic 2, with a symmetric bilinear form $B$. If $B$ is also alternating, i.e. $B(x,x)=0$ for all $x \in V$, then the rank of $B$ is even, equal to $2m$ say, and in a suitable basis the form has the matrix

$$\begin{pmatrix} 0 & I_m & 0 \\ I_m & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{1}$$

as we saw in **8.5**. Since the property of being alternating is independent of the choice of basis, it is clear that we cannot transform $B$ to diagonal form unless it is identically zero. But for any form that is non-alternating such a transformation exists.

PROPOSITION 7.1 *Let $V$ be a vector space over a field $k$ of characteristic 2, with a symmetric bilinear form $B$ which is not alternating. Then $V$ has an orthogonal basis.*

*Proof.* By hypothesis there exists $u_1 \in V$ such that $B(u_1,u_1) \neq 0$. We have $V = \langle u_1 \rangle \perp u_1^{\perp}$ and we can use induction on $\dim V$ unless $B$ is alternating on $u_1^{\perp} = U$. In that case we can take a basis in $U$ so that the matrix takes the form (1). Consider the subspace of $V$ spanned by $u_1$ and a hyperbolic plane in $U$ (which will exist unless the form is zero on $U$). Its matrix is

$$\begin{pmatrix} c & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

where $c = B(u_1,u_1) \neq 0$. Put $v_1 = u_1 + u_2 + u_3$; then $B(v_1,v_1) = c \neq 0$ and $v_1^{\perp}$ has

the basis $v_2 = u_1 + cu_2, v_3 = u_1 + cu_3$. The matrix in this basis is

$$\begin{pmatrix} c & 0 & 0 \\ 0 & c & c+c^2 \\ 0 & c+c^2 & c \end{pmatrix}$$

and the earlier argument shows that this can be diagonalized; now the diagonalization can be completed by induction on $\dim V$. ■

This result, together with the earlier remark on alternating forms deals with the case of bilinear forms, and we now turn to the case of a quadratic form $q$. From it we obtain as before a symmetric bilinear form $B$, given by

$$B(x, y) = q(x + y) - q(x) - q(y), \tag{2}$$

but in characteristic 2 this is always alternating: $B(x, x) = 0$, so it cannot be used to reconstruct $q$. Nevertheless it is possible to define $q(x)$ in terms of a bilinear form $b(x, y)$, but this form cannot always chosen to be symmetric.

Given a quadratic form $q$ and its derived bilinear form $B$ given by (2), we define a bilinear form $b$ in terms of a basis $u_1, \ldots, u_n$ of $V$ by

$$b(u_i, u_j) = \begin{cases} B(u_i, u_j) & \text{if } i < j, \\ q(u_i) & \text{if } i = j, \\ 0 & \text{if } i > j. \end{cases} \tag{3}$$

Together with bilinearity this defines $b$ on $V$ and we have

$$b(x, x) = q(x), \qquad \text{for all } x \in V, \tag{4}$$

as is easily checked. We define orthogonality for $q$ in terms of the alternating form $B$ given by (2). If this form $B$ is regular, the form $q$ is said to be *non-defective*. In that case the dimension is even: $n = 2m$, and we can find a symplectic basis $u_1, \ldots, u_m, v_1, \ldots, v_m$ such that $B(u_i, v_i) = B(v_i, u_i) = 1$ $(i = 1, \ldots, m)$, while all other products are zero. If the subspace spanned by the pair $u_i, v_i$ is denoted by $V_i$, then we have

$$V = V_1 \perp \cdots \perp V_m,$$

and the form $b$ defined by (3) has the matrix $P = P_1 \oplus \cdots \oplus P_m$, where

$$P_i = \begin{pmatrix} a_i & 1 \\ 0 & b_i \end{pmatrix}, \qquad a_i = q(u_i),\ b_i = q(v_i).$$

We remark that the matrix $P$ is not uniquely determined by the given basis, since $P + N$, for any alternating matrix $N$, determines the same form. However, $P$ is the unique upper triangular matrix for this basis.

Since the matrix of a quadratic form is only determined up to a summand which is alternating, its determinant is not an invariant, But there is an invariant

to take its place, the Arf invariant. To define it we recall that for any field $k$ of characteristic 2 the set $\{x^2 + x | x \in k\}$ is a subgroup of its additive group, denoted by $\wp(k)$ (cf. **5.8**).

THEOREM 7.2 (C. Arf 1940) *Let $V$ be a vector space over a field $k$ of characteristic 2, with a non-defective quadratic form $q$. Then for any symplectic basis $u_1, \dots, u_m, v_1, \dots, v_m$ of $V$ the expression*

$$\delta(q) = \sum q(u_i)q(v_i) \tag{5}$$

*is unique modulo $\wp(k)$.*

The element of $k/\wp(k)$ defined by $\delta(q)$ is called the *Arf invariant* of the form $q$.

*Proof.* The matrix of $q$ relative to the given basis is $\begin{pmatrix} A & I \\ 0 & B \end{pmatrix}$, where $A, B$ are diagonal matrices, and $\delta(q) = \operatorname{tr}(AB)$. If we add an alternating matrix, $A$, $B$ are replaced by $A' = A + M$, $B' = B + N$, where $M, N$ are alternating. Now $M, N$ have zeros on the main diagonal; so do $AN$, $MB$ and $\operatorname{tr}(MN) = 0$, hence $\operatorname{tr}(AB) = \operatorname{tr}(A'B')$.

Thus $\delta$ is independent of the matrix chosen for $q$ (in the given basis) and it remains to show that it is unchanged mod $\wp(k)$ by a change of symplectic basis. It is enough to verify this for the generating set of $\mathbf{Sp}_{2m}(k)$ given in Lemma 5.2. Consider first $S_P$; we have

$$\begin{pmatrix} P & 0 \\ 0 & P^{-T} \end{pmatrix}\begin{pmatrix} A & I \\ 0 & B \end{pmatrix}\begin{pmatrix} P^T & 0 \\ 0 & P^{-1} \end{pmatrix} = \begin{pmatrix} PAP^T & I \\ 0 & P^{-T}BP^{-1} \end{pmatrix},$$

and $\operatorname{tr}(PAP^T \cdot P^{-T}BP^{-1}) = \operatorname{tr}(AB)$. For $R_Q$ we have

$$\begin{pmatrix} I & Q \\ 0 & I \end{pmatrix}\begin{pmatrix} A & I \\ 0 & B \end{pmatrix}\begin{pmatrix} I & 0 \\ Q & I \end{pmatrix} = \begin{pmatrix} A + Q + QBQ & I + QB \\ BQ & B \end{pmatrix}.$$

Bearing in mind the relation $\operatorname{tr}(C^2) = \operatorname{tr}(C)^2$ (in characteristic 2), we find that $\operatorname{tr}((A + Q + QBQ)B) = \operatorname{tr}(AB + QB + (QB)^2) = \operatorname{tr}(AB) + \operatorname{tr}(QB) + \operatorname{tr}(QB)^2$. Hence $R_Q$ does not affect the residue class of $\delta$, and a similar argument applies to $R_Q^T$. Since $\mathbf{Sp}_{2m}(k)$ is generated by all $S_P, R_Q$ and $R_Q^T$, this shows $\delta(q)$ to be invariant. ■

## Exercises

(1) Verify that $\delta(q_1 \perp q_2) = \delta(q_1) + \delta(q_2)$.

(2) Show that a two-dimensional form is isotropic iff $\delta = 0$.

(3) Show that a two-dimensional form is determined up to isometry by its Arf invariant

and a single value, i.e. two forms $q_1, q_2$ are isometric iff $\delta(q_1) = \delta(q_2)$ and $q_1(x_1) = q_2(x_2)$ for some vectors $x_1, x_2 \neq 0$.

(4) Let $q$ be a quadratic form in characteristic 2. Show that the Arf invariant $\delta(q)$ is non-zero iff the even Clifford algebra $C_0(q)$ is simple.

## Further exercises on Chapter 8

(1) Show that the Clifford algebra of an even-dimensional regular quadratic space over a finite field $k$ is a full matrix ring over $k$. What happens in odd dimensions?

(2) Show that any Clifford algebra $C(V)$ forms an algebra for the multiplication

$$a \wedge b = ab - (-1)^{mn}ba, \qquad \text{where } m = \deg a,\ n = \deg b,$$

and that the subalgebra generated by $V$ is just the exterior algebra on $V$.

(3) Show that an orthogonal transformation of a split space which is the identity on a maximal totally isotropic subspace is a rotation.

(4) Let $L/K$ be a field extension. Show that an ordering on $K$ has an extension to $L$ iff the signature homomorphism on $W(K)$ maps the kernel of the natural homomorphism $W(K) \to W(L)$ to 0. Deduce another proof of Lemma 3.6.

(5) Let $L = K(\alpha)$ be a quadratic extension, where $\alpha^2 = a \in K$. If $V$ is an anisotropic $K$-space such that $V_L$ is isotropic, show that $V$ splits off a subspace isometric to $\langle c, -ac \rangle$, for some $c \in K$. (Hint: take an isotropic vector $x + \alpha y$ and consider the space spanned by $x, y$.)

(6) (A. Pfister) Let $L = K(\alpha)$ be a quadratic extension, where $\alpha^2 = a \in K$. Show that the kernel of the natural homomorphism $W(K) \to W(L)$ is the ideal generated by $\langle 1, -a \rangle$.

(7) Show that for extensions $L/K$ of odd degree, the natural map $W(K) \to W(L)$ is injective.

(8) Let $I$ be the augmentation ideal in $W(K)$. Show that there is a homomorphism from $I^2$ to the Sylow 2-subgroup of $\mathbf{B}_K$ (the *symbol homomorphism*; Merkuryev showed in 1981 that $I^2/I^3$ is isomorphic to the Sylow 2-subgroup of $\mathbf{B}_K$).

(9) Let $A$ be a symmetric matrix over $\mathbf{R}$ and denote by $f_i(x)$ the principal minor of order $i$ (i.e. formed from the first $i$ rows and columns) of $xI - A$. Show that $f_i g_i - h_i^2 = f_{i+1} f_{i-1}$, where $g_i, h_i$ are certain polynomials in $x$. (Hint: use Ex. (16) of Ch. 2.)

Deduce that $f_n, \ldots, f_0$ form a Sturm sequence for the characteristic polynomial of $A$ in any interval in which no two successive terms have a common zero.

(10) Show that the number of real roots of $f = 0$ can be determined from the leading

terms of a Sturm sequence. Show that the number of non-real roots is the number of sign changes of $1, s_0, \begin{vmatrix} s_0 & s_1 \\ s_1 & s_2 \end{vmatrix}, \ldots$, where the $s_i$ are the power sums of the roots.

(11) Let $r = r(x)$ be a rational function of $x$. At a pole $\xi$ of $r$ define the *index* of $r$ to be 1 if $r(x)$ changes from negative to positive values as $x$ increases through $\xi$, $-1$ if $-r$ has index 1 at $\xi$ and 0 otherwise. Write $r = f_1/f_0$, where $f_0, f_1$ are polynomials without a common factor, and let $f_0, f_1, f_2, \ldots$ be the sequence of negative remainders in the Euclidean algorithm for $f_0, f_1$. If $\alpha < \beta$ are not poles of $r$, show that the sum of the indices of $r(x)$ in $[\alpha, \beta]$ is $\delta(\alpha) - \delta(\beta)$, where $\delta$ is defined as in Th. 3.14. Deduce Th. 3.15 as a corollary.

(12) Show that in a field $k$ of characteristic not 2 every element is a square iff $W(k)$ is of order 2.

(13) Let $L/K$ be a finite extension of degree $d$. Show that there is an injective homomorphism $\mathbf{Sp}_{2m}(L) \to \mathbf{Sp}_{2dm}(K)$.

(14) Describe $\mathbf{O}(V)^{ab}$ for a two-dimensional anisotropic space $V$.

(15) Show that over a perfect field of characteristic 2, a non-defective quadratic form of dimension $2m$ has Witt index at least $m-1$ and is determined up to isometry by its dimension and its Arf invariant.

# 9
# Noetherian rings and polynomial identities

The Artinian condition on rings leads to a very satisfactory theory, at least in the semisimple case, yet it excludes such familiar examples as the ring of integers. This ring is included in the wider class of Noetherian rings, which has been much studied in recent years. We shall present some of the highlights, such as localization and Goldie's theorem, and illustrate the theory by examples from skew polynomial rings and power series rings.

Another condition which helps to make rings amenable is the existence of a polynomial identity. The topics treated include generic matrix rings, central polynomials and the theorems of Amitsur and Regev, while Kaplansky's basic theorem of *PI*-rings is reserved for Ch. 10.

## 9.1 Rings of fractions

In A.2, **9.3** we constructed a ring of fractions $R_S$ for a commutative ring $R$ with respect to a multiplicative subset $S$. This construction can be carried out quite generally, as we shall see in Th. 1.1 below, which should be compared with Th. 9.3.1 of A.2. It is not even necessary to assume $R$ commutative. But for the result to be of use, we must have a means of comparing expressions in $R_S$ and this will require further hypotheses to be imposed. This will be done in Th. 1.3 below. For a general treatment it is necessary to invert matrices rather than just elements, but this will not be needed here (cf. Cohn [1985]).

Given any ring $R$ and a subset $X$ of $R$, we say that a homomorphism $f: R \to R'$ is *$X$-inverting* if it maps each element of $X$ to an invertible element of $R'$. For fixed $R$ and $X$, the pairs $(R, f)$, where $f$ is $X$-inverting, can be regarded as objects of a category whose morphisms are commutative triangles, and an initial object in this category is called a *universal $X$-inverting homomorphism.* An element $c$ in a ring $R$ is called *left regular* if $xc = 0$ implies $x = 0$, i.e. $c$ is not a right zerodivisor; *right regular* elements are defined similarly, and a *regular* element is one which is left and right regular.

THEOREM 1.1 *Let $R$ be any ring and $S$ a subset of $R$. Then there exists a ring $R_S$ and a homomorphism $\lambda: R \to R_S$ which is universal $S$-inverting.*

In detail the assertion means that $\lambda$ is $S$-inverting and every $S$-inverting homomorphism can be factored uniquely by $\lambda$. To construct $R_S$ we take for each $a \in R$ a sympol $p_a$ and for each $s \in S$ an additional symbol $q_s$ and form the ring $R_S$ on all these symbols as generators, with the defining relations

$$p_1 = 1, p_a + p_b = p_{a+b}, p_a p_b = p_{ab}, p_s q_s = q_s p_s = 1, \qquad \text{for all } a, b \in R, s \in S. \quad (1)$$

The first three sets of equations ensure that the mapping $\lambda: a \mapsto p_a$ is a homomorphism of $R$ into $R_S$, while the fourth shows that $\lambda$ is $S$-inverting. Now given any $S$-inverting homomorphism $f: R \to R'$, we can define a homomorphism $g$ of the free ring $F$ on the $p$'s and $q$'s into $R'$ by the rules $p_a \mapsto af, q_s \mapsto (sf)^{-1}$. It is clear that $g$ preserves the relations (1); hence it induces a homomorphism $g_1: R_S \to R'$, and it is easily checked that $\lambda g_1 = f$. Moreover, $g_1$ is uniquely determined by this equation, since its value on $p_a$ is given, while its value on $q_s$ is determined by (1): if $q_s g_1 = c$, then $sf \cdot c = (p_s q_s) g_1 = 1 = (q_s p_s) g_1 = c \cdot sf$, so $c = (sf)^{-1}$. Thus $g_1$ is determined on a generating set of $R_S$ and hence is unique. ■

COROLLARY 1.2 *Let $R, R'$ be any rings with subsets $S, S'$ respectively. Then any homomorphism $f: R \to R'$ which maps $S$ into $S'$ can be extended in a unique way to a homorphism of $R_S$ into $R'_{S'}$.*

For the composition $R \xrightarrow{f} R' \xrightarrow{\lambda'} R'_{S'}$ is $S$-inverting and so can be factored uniquely by $\lambda: R \to R_S$ to give the required homomorphism $f_1: R_S \to R'_{S'}$. ■

The ring $R_S$ constructed in Th. 1.1, together with the homomorphism $\lambda: R \to R_S$ is called the *universal S-inverting ring* for $R$ or also the *localization* of $R$ at the set $S$. To study it in more detail we need to make some simplifying assumptions. A very simple but most fruitful idea, due to O. Ore (1931) (and independently, to E. Noether, unpublished) is to look at the case where all the elements of $R_S$ can be written as simple fractions $(a\lambda)(s\lambda)^{-1}$. If this is to be possible, we must in particular be able to express $(s\lambda)^{-1}(a\lambda)$ in this form: $(s\lambda)^{-1}(a\lambda) = (a_1\lambda)(s_1\lambda)^{-1}$, hence on multiplying up, we obtain

$$(as_1)\lambda = (sa_1)\lambda, \quad (2)$$

and this provides a clue to the condition needed. Of course, if every element is to be expressed as a fraction with denominator in $S$, we must also assume $S$ to be *multiplicative*, i.e. to contain 1 and be closed under products.

THEOREM 1.3 *Let $R$ be a ring and $S$ a multiplicative subset of $R$ such that*

**O.1** *$aS \cap sR \neq \varnothing$ for all $a \in R$, $s \in S$,*

**O.2** *for each $a \in R$, $s \in S$, if $sa = 0$, then $at = 0$ for some $t \in S$.*

*Then the elements of the universal S-inverting ring $R_S$ can be constructed as*

*fractions $a/s$, where*

$$a/s = a'/s' \Leftrightarrow au = a'u' \text{ and } su = s'u' \in S \qquad \text{for some } u, u' \in R. \quad (3)$$

*Moreover, the kernel of the natural homomorphism $\lambda: R \to R_S$ is*

$$\ker \lambda = \{a \in R \mid at = 0 \text{ for some } t \in S\}. \quad (4)$$

Condition **O.**1 is called the *right Ore condition.* A multiplicative subset $S$ of $R$ satisfying **O.**1 is called a *right Ore set*; if $S$ *also satisfies* **O.**2, it is called (right) *reversible* or also a *right denominator set.* Such a set allows the construction of *right* fractions $a/s = (a\lambda)(s\lambda)^{-1}$, by Th. 1.3. They must be carefully distinguished from *left* fractions $(s\lambda)^{-1}(a\lambda)$. By symmetry we have the notion of a (reversible) *left Ore set*, which allows us to construct all the elements of $R_S$ as left fractions, and the set $S$ in Th. 1.3 may well be a right but not left Ore set.

The proof of Th. 1.3 is similar to the commutative case (Th. 9.3.1 of A.2), though more care is needed, owing to the lack of commutativity. Guided by (3), we define a relation on $R \times S$ by writing

$$(a, s) \sim (a', s') \Leftrightarrow \text{there exist } u, u' \in R \text{ such that } au = a'u', su = s'u' \in S.$$

We claim that this is an equivalence. Clearly it is reflexive and symmetric. To prove transitivity, let $(a, s) \sim (a', s')$, $(a', s') \sim (a'', s'')$; say $au = a'u'$, $su = s'u' \in S$, $a'v = a''v'$, $s'v = s''v' \in S$. By **O.**1 there exist $z \in S$, $z' \in R$ such that $s'u'z = s'vz'$, hence $s'u'z \in S$ (by multiplicativity) and moreover, $s'(u'z - vz') = 0$, therefore by **O.**2 there exists $t \in S$ such that $u'zt = vz't$. Now we have $auzt = a'u'zt = a'vz't = a''v'z't$, $suzt = s'u'zt = s'vz't = s''v'z't$, and this lies in $S$ because $s'u'z \in S$ and $t \in S$. Thus $(a, s) \sim (a'', s'')$.

We thus have an equivalence on $R \times S$; let us write $a/s$ for the class containing $(a, s)$ and call $a$ the *numerator* and $s$ the *denominator* of this expression. We note that (3) holds now by definition, and it may be interpreted as saying that two fractions are equal iff when they are brought to a common denominator, their numerators agree. Of course it follows from **O.**1 that any two expressions can be brought to a common denominator. For this reason we can define the addition of fractions by the rule

$$a/s + b/s = (a + b)/s. \quad (5)$$

Here it is necessary to check that the expression on the right depends only on $a/s, b/s$ and not on $a, b, s$, a task which may be left to the reader. To define the product of $a/s$ and $b/t$ we determine $b_1 \in R$ and $s_1 \in S$ such that $bs_1 = sb_1$ and then put

$$(a/s)(b/t) = ab_1/ts_1.$$

Again the proof that this is well defined is left to the reader. Now it is easy to check that with these operations the classes $a/s$ form a ring $T$ say, and the

mapping $a \mapsto a/1$ is an $S$-inverting homomorphism from $R$ to $T$. Moreover, if $f: R \to R'$ is any $S$-inverting homomorphism, then the mapping $f_1: R \times S \to R'$ given by $(a, s) \mapsto (af)(sf)^{-1}$ is constant on each equivalence class and so can be factored via $T$ to provide a homomorphism $f': T \to R'$ such that $f = \lambda f'$. Here $f'$ is unique, because it is determined on $a/1$ and $1/s$, so by uniqueness, $T$ is indeed the universal $S$-inverting ring. Finally, $\ker \lambda$ consists of all $a/1 = 0/1$, i.e. by (3), all $a$ such that $at = 0$ for some $t \in S$. ■

An important case is that where $S$ lies in the centre of $R$. Then **O.**1–2 are automatic and we have

COROLLARY 1.4 *Let $R$ be a ring and $S$ any multiplicative subset of the centre of $R$. Then $S$ is a reversible Ore set and the universal $S$-inverting ring $R_S$ consists of all fractions $a/s$ $(a \in R, s \in S)$, where*

$$a/s = a'/s' \Leftrightarrow as' = sa'. \quad ■$$

The conditions of Th. 1.3 simplify slightly when $S$ consists entirely of regular elements. The **O.**2 is superfluous and $\ker \lambda = 0$. We state this as

COROLLARY 1.5 *Let $R$ be a ring and $S$ a multiplicative right Ore subset of regular elements. Then the natural homomorphism $\lambda: R \to R_S$ is injective.* ■

The subset $T$ of all regular elements in $R$ is always multiplicative and satisfies **O.**2. When it satisfies **O.**1, we can form $R_T$; this is called the *total* (or *classical*) quotient ring. Generally one understands by a *quotient ring* a ring in which every regular element is a unit.

Finally we note the special case of integral domains. Writing $R^\times = R \setminus \{0\}$, we have

COROLLARY 1.6 *Let $R$ be an integral domain such that*

$$aR \cap bR \neq 0 \qquad \text{for all } a, b \in R^\times. \tag{6}$$

*Then $R^\times$ is a regular Ore set, $K = R_{R^\times}$ is a skew field and the natural mapping $\lambda: R \to K$ is an embedding. Conversely, if $R$ is an integral domain with an embedding in a skew field whose elements all have the form $ab^{-1}$ $(a \in R, b \in R^\times)$, then (6) holds.* ■

The skew field $K$ occurring here is called the *field of fractions* of $R$. The special case (6) of **O.**1 was used by Ore (1931) in his proof of Cor. 1.6; since then there have been many papers generalizing Ore's construction to the case of Th. 1.3 or a special case. An integral domain satisfying (6) is called a *right Ore domain*; *left* Ore domains are defined similarly and an *Ore domain* is a domain which is left and right Ore.

The following property of fractions is often useful.

PROPOSITION 1.7 *Let $R$ be ring with a reversible right Ore set $S$ and universal $S$-inverting ring $R_S$. Then any finite set in $R_S$ can be brought to a common denominator.*

*Proof.* We shall use induction on the number of elements. Let $a_i/s_i \in R_S$ be given, $i = 1, \ldots, n$. For $n = 1$ there is nothing to prove; so by induction we may assume these fractions to be in the form $a_1/s_1$, $a_2/s, \ldots, s_n/s$. By **O.1** there exist $t \in S$, $c \in R$ such that $sc = s_1 t = u \in S$, hence the fractions can be written $a_1 t/u, a_2 c/u, \ldots, a_n c/u$. ■

In any integral domain $R$ (commutative or not) the additive subgroup generated by 1 is a commutative subring, necessarily a homomorphic image of $\mathbf{Z}$, hence it is either $\mathbf{Z}$ or $\mathbf{Z}/p$ for some prime $p$. Accordingly we say that $R$ has *characteristic* 0 or $p$; this generalizes the customary usage for fields. If $K$ is any commutative ring, a $K$-algebra $A$ is said to be *faithful* if the natural homomorphism $K \to A$ is injective. Thus we can say that every integral domain can be regarded as a faithful $K$-algebra, where $K = \mathbf{Z}$ or $\mathbf{F}_p$.

Suppose now that $R$ is an integral domain, hence a faithful $K$-algebra ($K = \mathbf{Z}$ or $\mathbf{F}_p$), but not right Ore. Then $R$ contains $a, b \neq 0$ such that $aR \cap bR = 0$. We claim that the $K$-subalgebra of $R$ generated by $a$ and $b$ is free on $a, b$. For let $f(x, y)$ be a non-zero polynomial in the non-commuting variables $x, y$ with coefficients in $K$, such that $f(a, b) = 0$, and choose $f$ to have the least possible degree. We can write

$$f(x, y) = \alpha + xf_1(x, y) + yf_2(x, y), \qquad \text{where } \alpha \in K,$$

and $f_1, f_2$ are not both zero, for otherwise $\alpha = 0$ and $f$ would be the zero polynomial. Suppose that $f_1 \neq 0$; then $f_1(a, b) \neq 0$, by the minimality of $\deg f$, but $f(a, b) = 0$, hence $f(a, b)b = 0$ and so

$$af_1(a, b)b = b(-\alpha - f_2(a, b)b),$$

and this is non-zero because the left-hand side is non-zero. This contradicts the fact that $aR \cap bR = 0$, and it proves

PROPOSITION 1.8 *Any integral domain $R$ is a faithful $K$-algebra, where $K$ is $\mathbf{Z}$ or $\mathbf{F}_p$, according as $R$ is of characteristic 0 or $p$. Further, $R$ is either a left and right Ore domain, or it contains a free $K$-algebra on two free generators.* ■

We observe that the last two possibilities are not mutually exclusive: an Ore domain may well contain a free algebra. We also recall from Ex. (9) of Ch. 2 that an algebra containing a free algebra of rank 2 contains a free algebra of countable rank.

It is an interesting observation made by Goldie (1958) (actually the simplest case of Th. 4.8 below) that the Ore condition is a consequence of the Noetherian condition:

PROPOSITION 1.9 *Any integral domain is either a right Ore domain or it contains free right ideals of infinite rank. In particular, any right Noetherian domain is right Ore.*

*Proof.* Let $R$ be an integral domain and suppose that $R$ is not right Ore. Then there exist $a, b \in R^\times$ such that $aR \cap bR = 0$; now the conclusion will follow if we show that the elements $b, ab, a^2b, \ldots$ are right linearly independent. Suppose that $\sum a^i b c_i = 0$ and let $c_\nu$ be the first non-zero coefficient. We can cancel $a^\nu$ and obtain the relation $bc_\nu + abc_{\nu+1} + \cdots + a^{n-\nu}bc_n = 0$, i.e.

$$a(bc_{\nu+1} + \cdots + a^{n-1-\nu}bc_n) = -bc_\nu \neq 0,$$

and this contradicts the assumption on $a, b$. ■

A ring $R$ is said to be *prime* if $R \neq 0$ and $xRy = 0$ implies $x = 0$ or $y = 0$; Such rings will be discussed in **10.6**; for the moment let us apply the results found to prime rings.

PROPOSITION 1.10 *Let $R$ be a prime ring and $S$ a right Ore set consisting of regular elements of $R$. Then $R$ is embedded in $R_S$ and $R_S$ is again prime.*

*Proof.* The mapping $R \to R_S$ is an embedding, by Prop. 1.5. Suppose that $uR_S v = 0$, where $u = as^{-1}$, $v = bt^{-1}$ $(a, b \in R, s, t \in S)$. Then for any $x \in R$, $axb = as^{-1} \cdot sx \cdot bt^{-1} \cdot t = 0$, hence $a = 0$ or $b = 0$ and accordingly $u = 0$ or $v = 0$. ■

COROLLARY 1.11 *Let $R$ be a prime ring with centre $C$. Then $C^\times$ consists of regular elements in $R$; in particular, $C$ is an integral domain, and if its field of fractions is denoted by $K$, then $R$ is embedded in $R_{C^\times} \cong R \otimes_C K$, and the latter ring is again prime, with centre $K$.*

*Proof.* We first show that $C^\times$ consists of regular elements. If $c \in C^\times$ and $ca = 0$ $(a \in R)$, then for all $x \in R$, $cxa = xca = 0$, hence $a = 0$. Since $ac = ca$ for $a \in R$, this shows $c$ to be regular. Since $K$ is universal $C^\times$-inverting, it follows that $R \otimes K$ is universal $C^\times$-inverting and by Prop. 1.10, the mapping $R \to R_{C^\times} \cong R \otimes K$ is an embedding. Clearly $K$ is contained in the centre of $R \otimes K$; conversely, if $ac^{-1}$ is in the centre, then for any $x \in R$, $ax = ac^{-1} \cdot cx = cx \cdot ac^{-1} = xa$, because $c \in C$. Hence $a \in C$ and so $ac^{-1} \in K$. ■

## Exercises

(1) Show that every right Artinian ring is a quotient ring.

(2) Show that an integral domain in which every finitely generated right ideal is principal is a right Ore domain (such a ring is called a *right Bezout domain*).

(3) Show that for any reversible right Ore set $S$ in a ring $R$, the ring $R_S$ is flat as left $R$-module.

(4) Show that for any ring $R$ and subset $S$, the natural mapping $\lambda: R \to R_S$ is an epimorphism in the category of rings.

(5) Show that the characteristic can be defined for simple rings as for integral domains, and that a simple ring is a $\mathbf{Q}$-algebra or an $\mathbf{F}_p$-algebra according as the characteristic is 0 or $p$.

(6) Let $R$ be a simple ring and $S$ a reversible right Ore set. Show that the centre of $R_S$ coincides with the centre of $R$.

(7) Let $R$ be a ring, $X$ a subset consisting of regular elements and $S$ the submonoid of $R^\times$ generated by $X$. Show that $S$ is a right Ore set provided that the Ore condition holds in the form: $xR \cap rS \neq \varnothing$ for all $x \in X$, $r \in R$.

(8) Let $F$ be the free group on $x$ and $y$. Show that the group ring $\mathbf{Z}F$ is an integral domain, but not left or right Ore.

(9) Let $R$ be a ring with right total quotient ring. Show that any right factor of a regular element is regular.

(10) Let $R$ be a right Ore domain and $K$ its field of fractions. Show that the centre of $K$ consists of all $ab^{-1}$ $(b \neq 0)$ such that $axb = bxa$ for all $x \in R$.

## 9.2 Principal ideal domains

In A.1 we saw that over a commutative principal ideal domain (PID) every matrix is associated to a diagonal matrix; this is the Smith normal form, proved in Th. 4, **10.5** of A.1. Surprisingly there is an analogue for non-commutative PIDs, with a rather strong uniqueness condition which is sometimes useful. To state the result, let us define an *invariant* element in a ring $R$ as a regular element $c$ such that $cR = Rc$, thus the left (or right) ideal generated by $c$ is two-sided. If $a, b$ are regular elements of $R$, then $a$ is called a *total divisor* of $b$, in symbols $a \parallel b$, if there exists an invariant element $c$ such that $a|c|b$. We observe that an element is not generally a total divisor of itself; in fact $a \parallel a$ iff $a$ is invariant. A simple ring has no non-unit invariant elements and $a \parallel b$ implies that either $a$ is a unit or $b = 0$. Further, in a principal ideal domain $R$, every two-sided ideal is of the form $cR = Rc$, where $c$ is 0 or an invariant element.

We shall write $\mathrm{diag}(d_1, \ldots, d_r)$ for a matrix with $d_1, \ldots, d_r$ along the main diagonal and zeros elsewhere; this notation will be used even for matrices that are not square. The exact size will be indicated explicitly unless it is clear from the context.

THEOREM 2.1 *Let $R$ be a principal ideal domain and $A \in {}^mR^n$. Then there exist an integer $r$ and $P \in \mathbf{GL}_m(R)$, $Q \in \mathbf{GL}_n(R)$ such that*

$$PAQ = \operatorname{diag}(e_1, \ldots, e_r, 0, \ldots, 0), \qquad e_i \| e_{i+1},\ e_r \neq 0. \tag{1}$$

*Proof.* As in the commutative case there are four types of operations on the columns of $A$, of which the first three are the well known elementary operations:

(i) *interchange two columns,*
(ii) *multiply a column on the right by a unit factor,*
(iii) *add a right multiple of one column to another,*
(iv) *multiply two columns on the right by an invertible $2 \times 2$ matrix.*

Each of (i)–(iii) corresponds to multiplying on the right by an invertible matrix. The object of (iv) is to replace the first two elements in the columns by their highest common left factor (HCLF) and 0 respectively. Thus if these elements are $a, b$, not both 0, then we have $aR + bR = kR$, say $a = ka_1$, $b = kb_1$ and there exist $c_1', d_1'$ such that $a_1 d_1' - b_1 c_1' = 1$. By hypothesis $Rc_1' \cap Rd_1'$ is principal, with generator $d_1 c_1' = c_1 d_1'$ and $c_1, d_1$ have no common left factor, so there exist $a_1', b_1' \in R$ such that $d_1 a_1' - c_1 b_1' = 1$. Thus we have

$$\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \begin{pmatrix} d_1' & -b_1' \\ -c_1' & a_1' \end{pmatrix} = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}.$$

This shows that the first matrix on the left, $C$ say, is invertible. It follows that $(k, 0)C = (a, b)$ and $(a, b)C^{-1} = (k, 0)$, as required. Corresponding operations can of course be carried out on the rows, acting on the left.

We can now proceed with the reduction. If $A = 0$, there is nothing to prove; otherwise we bring a non-zero element to the $(1, 1)$-position in $A$, by permuting rows and permuting columns, using (i). Next we use (iv) to replace $a_{11}$ successively by the HCLF of $a_{11}$ and $a_{12}$, then by the HCLF of the new $a_{11}$ and $a_{13}$, etc. After $n - 1$ steps we have transformed $A$ to a form where $a_{12} = a_{13} = \cdots = a_{1n} = 0$. By symmetry the same process can be applied to the first column of $A$; in the course of the reduction the first row of $A$ may again become non-zero, but this can happen only if the length (i.e. number of factors) of $a_{11}$ is reduced; therefore by induction on the length of $a_{11}$ we reach the form $a_{11} \oplus A_1$, where $A_1$ is $m - 1 \times n - 1$. We now apply the same process to $A_1$ and by induction on $\max(m, n)$ we reach the form

$$\operatorname{diag}(a_1, a_2, \ldots, a_r, 0, \ldots, 0).$$

Consider $a_1$ and $a_2$; for any $d \in R$ we have

$$\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} = \begin{pmatrix} a_1 & da_2 \\ 0 & a_2 \end{pmatrix};$$

so we can further diminish the length of $a_1$ unless $a_1$ is a left factor of $da_2$ for

all $d \in R$, i.e. unless $a_1 R \supseteq Ra_2$. But in that case $a_1 R \supseteq Ra_2 R \supseteq Ra_2$; thus $a_1 | c | a_2$, where $c$ is the invariant generator of the ideal $Ra_2 R$. Hence $a_1 \| a_2$, and by repeating the argument we obtain the expression on the right of (1). The totality of column operations amount to right multiplication by an invertible matrix $Q$, the row operations yield $P$ and we thus have the equation (1). ■

We remark that most of the PIDs we encounter will be Euclidean domains (possibly non-commutative). In that case we can instead of (iv) use the Euclidean algorithm, with an induction on the degree instead of the length, to accomplish the reduction in Th. 2.1 (cf. the proof of Th. 4.7.1). Further we can dispense with (ii), so $P, Q$ can in this case be taken to be products of elementary matrices.

In the case of simple rings Th. 2.1 still simplifies, since then every invariant element is a unit:

COROLLARY 2.2 *If $R$ is a simple principal ideal domain, then every matrix over $R$ is associated to a matrix of the form $I \oplus a \oplus 0$ $(a \in R)$.*

For if $a \| b$, then either $b = 0$ or $a$ is a unit. Now any unit can be transformed to 1 by applying (ii), so there can only be one diagonal element not 1 or 0. ■

We note the consequences for modules over a PID:

PROPOSITION 2.3 *Let $R$ be a principal ideal domain. Then every submodule of a free module of finite rank $n$ is free of rank at most $n$. Further, any finitely generated module is a direct sum of cyclic modules:*

$$M = M_1 \oplus \cdots \oplus M_r, \qquad \text{where } M_{i+1} \text{ is a quotient of } M_i.$$

*If moreover, $R$ is simple, then every finitely generated module is the direct sum of a cyclic module and a free module of finite rank.*

*Proof.* Let $F$ be a free right $R$-module of finite rank $n$; we shall use induction on $n$. If $\alpha_1$ denotes the projection on the first component $R$ in $F$, then $\alpha_1 M$ is a right ideal, principal by hypothesis, and we have the exact sequence

$$0 \to M' \to M \to \alpha_1 M \to 0.$$

Here $M'$ is a submodule of $\ker \alpha_1$, while $\alpha_1 M$ is free (of rank 1 or 0), hence $M \cong M' \oplus \alpha_1 M$, and since $M'$ is free of rank $\leqslant n - 1$ by the induction hypothesis, the first assertion follows. Now the rest is clear from Th. 2.1 and Cor. 2.2. ■

In a skew field every non-zero element is a unit, so then Cor. 2.2 yields the well known result that every matrix over a skew field is associated to $I_r \oplus 0$, where $r$ is the rank of the matrix. We can also use Th. 2.1 to describe the rank of a matrix over the rational function field $K(t)$, defined as field of fractions of

$K[t]$, where $K$ is any skew field and $t$ is an indeterminate. We remark that if $C$ is the centre of $K$, then for any polynomial $f$ of degree $n$ over $K$ and any $\lambda \in C$ such that $f(\lambda) = 0$, we can write $f = (t - \lambda)g$, where $g$ has degree $n - 1$. By induction it follows that $f$ cannot have more than $n$ zeros in $C$.

LEMMA 2.4 *Let $K$ be a skew field with infinite centre $C$, and consider the polynomial ring $K[t]$, with field of fractions $K(t)$. If $A = A(t)$ is a matrix over $K[t]$, then the rank of $A$ over $K(t)$ is the supremum of the ranks of $A(\alpha)$, $\alpha \in C$. In fact, this supremum is assumed for all but a finite number of values of $\alpha$.*

*Proof.* Since $K[t]$ is a PID (A.1, **6.5**) we can by Th. 2.1 find invertible matrices $P, Q$ such that

$$PAQ = \operatorname{diag}(f_1, \ldots, f_r, 0, \ldots, 0), \qquad \text{where } f_i \in K[t]. \tag{2}$$

The product of the non-zero diagonal terms on the right gives us a polynomial $f$ whose zeros in $C$ are the only points of $C$ at which $A = A(t)$ falls short of its maximum rank, and the number of these values cannot exceed the degree of $f$. ■

## Exercises

(1) Show that over a PID any submodule of a free module (even of infinite rank) is free.

(2) Give the details of the proof that for a Euclidean domain (i), (iii) suffice to accomplish the reduction to the form (1).

(3) Let $K$ be a skew field with centre $C$. Show that a polynomial over $K$ may well have infinitely many zeros in $K$. How is this to be reconciled with the remark before Lemma 2.4?

(4) Let $R$ be a PID and $M$ a right $R$-module. An element $x$ of $M$ is called a *torsion element* if $xc = 0$ for some $c \in R^\times$. Verify that the set $tM$ of all torsion elements is a submodule of $M$ (called the *torsion submodule*).

(5) If $tM = 0, M$ is called *torsion free*. Show that any finitely generated torsion free module over a PID $R$ is free and deduce that a finitely generated $R$-module splits over its torsion submodule.

(6) Let $R$ be any ring and $M$ a right $R$-module. Show that for any invariant element $c$, $Mc$ is a submodule. If $R$ is a PID and $M$ is presented by a matrix $\operatorname{diag}(a_1, \ldots, a_r, 0, \ldots, 0)$, $a_i \| a_{i+1}$, then for any invariant element $c$ such that $a_i | c | a_{i+1}$ we have $M/Mc \cong R/a_1R \oplus \cdots \oplus R/a_iR \oplus R/cR \oplus \cdots \oplus R/cR$. Deduce that the $a_i$ are unique up to similarity, where $a, b$ are *similar* if $R/aR \cong R/bR$.

## 9.3 Skew polynomials and Laurent series

In commutative ring theory the polynomial ring $R[x]$ in an indeterminate $x$ plays a basic role. The corresponding concept for general rings is the ring freely generated by an indeterminate $x$ over $R$; in the notation of **2.2** this is the tensor ring $R_A\langle x\rangle$, where $A$ is the subring of $R$ generated by 1. This looks very different from $R[x]$; its elements are not at all like the usual polynomials, but we can simplify matters by taking the special case of those rings whose elements can be written in the form of polynomials. Thus for a given ring $R$ we consider a ring $P$ whose elements can be written uniquely in the form

$$f = a_0 + xa_1 + \cdots + x^n a_n, \qquad \text{where } a_i \in R. \tag{1}$$

As usual we define the *degree* of $f$ as the highest power of $x$ which occurs with a non-zero coefficient:

$$d(f) = \max\{i \mid a_i \neq 0\}. \tag{2}$$

We shall characterize the ring $P$ under the assumption that the degree has the usual properties:

**D.1** $d(f) \geqslant 0$ *for* $f \neq 0, d(0) = -\infty$,
**D.2** $d(f-g) \leqslant \max\{d(f), d(g)\}$,
**D.3** $d(fg) = d(f) + d(g)$.

An integer-valued function $d$ on a ring satisfying **D.**1–3 is called a *degree function* (essentially this means that $-d$ is a valuation, cf. **11.4** below or Ch. 8 of A.2). Leaving aside the trivial case $R = 0$, we see from **D.**3 that $P$ is an integral domain and moreover, for any $a \in R^\times$, $ax$ has degree 1, so there exist $a^\alpha, a^\delta$ such that

$$ax = xa^\alpha + a^\delta. \tag{3}$$

By the uniqueness of the form (1), the elements $a^\alpha, a^\delta$ are uniquely determined by $a$, and $a^\alpha = 0$ iff $a = 0$. By (3) we have $(a+b)x = x(a+b)^\alpha + (a+b)^\delta$, $ax + bx = xa^\alpha + a^\delta + xb^\alpha + b^\delta$, hence on comparing the right-hand sides we find

$$(a+b)^\alpha = a^\alpha + b^\alpha, \qquad (a+b)^\delta = a^\delta + b^\delta, \tag{4}$$

so $\alpha$ and $\delta$ are additive mappings of $R$ into itself. Next we have $(ab)x = x(ab)^\alpha + (ab)^\delta$, $a(bx) = a(xb^\alpha + b^\delta) = (xa^\alpha + a^\delta)b^\alpha + ab^\delta$, hence

$$(ab)^\alpha = a^\alpha b^\alpha, \qquad (ab)^\delta = a^\delta b^\alpha + ab^\delta. \tag{5}$$

Finally, $1 \cdot x = x \cdot 1 = x \cdot 1^\alpha + 1^\delta$, so

$$1^\alpha = 1, \qquad 1^\delta = 0. \tag{6}$$

The first equation in (4)–(6) shows that $\alpha$ is a ring homomorphism, and by the

remark following (3) it is injective. The remaining equations now show $\delta$ to be a $(1, \alpha)$-derivation (cf. **2.2**); here we shall refer to $\delta$ more briefly as an *$\alpha$-derivation.*

We next observe that the commutation rule (3), with the uniqueness of (1) is enough to determine the multiplication in $P$. For by the distributive law we need only know $x^m a \cdot x^n b$, and by (3) the effect of moving $a$ past a factor $x$ is

$$x^m a x^n b = x^{m+1} a^\alpha x^{n-1} b + x^m a^\delta x^{n-1} b.$$

Now an induction on $n$ allows us to write $x^m a x^n b$ as a polynomial in $x$. Thus $P$ is completely determined when $\alpha, \delta$ are given. We shall call $P$ a *skew polynomial ring* in $x$ over $R$ relative to the endomorphism $\alpha$ and $\alpha$-derivation $\delta$, and write it as $P = R[x; \alpha, \delta]$. Thus we have proved the first part of

THEOREM 3.1 *Let $P$ be a ring whose elements can be expressed uniquely as polynomials in $x$ with coefficients in a non-trivial ring $R$, as in* (1), *with a degree function defined by* (2), *and satisfying the commutation rule* (3). *Then $R$ is an integral domain, $\alpha$ is an injective endomorphism, $\delta$ is an $\alpha$-derivation and $P = R[x; \alpha, \delta]$ is the skew polynomial ring in $x$ over $R$, relative to $\alpha, \delta$. Conversely, given an integral domain $R$ with an injective endomorphism $\alpha$ and an $\alpha$-derivation $\delta$, there exists a skew polynomial ring $R[x; \alpha, \delta]$.*

It only remains to prove the converse. Consider the set $R^{\mathbb{N}}$ of all sequences $(a_i) = (a_0, a_1, \ldots)(a_i \in R)$, as right $R$-module. Besides the right multiplication by $R$ we have the additive group endomorphism

$$x:(a_i) \mapsto (a_i^\delta + a_{i-1}^\alpha), \qquad \text{where } a_{-1} = 0. \tag{7}$$

Clearly $R$ acts faithfully on $R^{\mathbb{N}}$ by right multiplication, so we may identify $R$ with its image in $E = \mathrm{End}(R^{\mathbb{N}})$. Let $P$ be the subring of $E$ generated by $R$ and $x$; we claim that $P$ is the required skew polynomial ring. To verify (3) we have

$$\begin{aligned}(c_i)ax = (c_i a)x &= ((c_i a)^\delta + (c_{i-1} a)^\alpha) \\ &= (c_i^\delta a^\alpha + c_i a^\delta + c_{i-1}^\alpha a^\alpha) \\ (c_i)(xa^\alpha + a^\delta) &= (c_i^\delta a^\alpha + c_{i-1}^\alpha a^\alpha + c_i a^\alpha).\end{aligned}$$

Hence $ax = xa^\alpha + a^\delta$ in $P$, and (3) holds. It follows that every element of $P$ can be written as a polynomial (1), and this expression is unique, for we have

$$(1, 0, 0, \ldots)(a_0 + xa_1 + \cdots + x^n a_n) = (a_0, a_1, \ldots, a_n, 0, \ldots),$$

so distinct polynomials represent different elements of $P$. Finally it is clear that $d(f)$ defined as in (2) is a degree function, because $R$ is an integral domain and $\alpha$ is injective. So $P$ is indeed a skew polynomial ring. ■

Restricting attention to polynomial rings is analogous to singling out Ore domains among general integral domains, and in fact the skew polynomial rings

were first considered by O. Ore 1933; the general form of Th. 3.1 is due to Jacobson 1934.

It is important to observe that the construction of the skew polynomial ring is *not* left–right symmetric, and besides $R[x; \alpha, \delta]$ we can also introduce the *left* skew polynomial ring, in which the coefficients are written on the left. The commutation rule (3) then has to be replaced by

$$xa = a^{\alpha}x + a^{\delta}. \tag{8}$$

In general the left and right skew polynomial rings are distinct, but when $\alpha$ is an automorphism of $R$, with inverse $\beta$ say, then on replacing $a$ by $a^{\beta}$ we can write (3) as $a^{\beta}x = xa + a^{\beta\delta}$, i.e.

$$xa = a^{\beta}x - a^{\beta\delta},$$

which is of the form (8). Hence we obtain

PROPOSITION 3.2 *The ring $R[x; \alpha, \delta]$ is a left skew polynomial ring whenever $\alpha$ is an automorphism of $R$.* ■

The Hilbert basis theorem extends to skew polynomial rings relative to an automorphism (for endomorphisms it need not hold, cf. Ex. (2)).

THEOREM 3.3 *Let $R$ be a right Noetherian domain, $\alpha$ an automorphism and $\delta$ an $\alpha$-derivation of $R$. Then the skew polynomial ring $R[x:\alpha, \delta]$ is again a right Noetherian domain.*

The proof is essentially the same as in the commutative case (Th. 9.4.2 of A.2) and will be left to the reader. Here it will be found more convenient to write the coefficients on the left of $x$; since $\alpha$ is an automorphism, this is possible. ■

In the construction of skew polynomial rings it was necessary to start from an integral domain because we insisted on a degree function; this is not essential, but it is the case mostly used in applications. Frequently the coefficient ring will even be a field, possibly skew. In that case the skew polynomial ring is a principal right ideal domain; this follows again as in the commutative case, using the division algorithm.

In particular, for a skew field $K$, the skew polynomial ring $K[x; \alpha, \delta]$ is right Noetherian (because it is right principal), and hence it is right Ore, by Prop. 1.9, so we can form its skew field of fractions. This is denoted by $K(x; \alpha, \delta)$; its elements are fractions $fg^{-1}$, where $f, g$ are polynomials (1) with coefficients in $K$.

PROPOSITION 3.4 *Any skew polynomial ring over a right Ore domain is again right Ore.*

*Proof.* Let $R$ be a right Ore domain and $K$ its skew field of fractions. Since $\alpha$ is an injective endomorphism of $R$, it can by Cor. 1.2 be extended to an endomorphism of $K$, again denoted by $\alpha$. Further, $\delta$ gives rise to an $R^\times$-inverting homomorphism from $R$ to $\mathfrak{M}_2(K)$, and this gives an $\alpha$-derivation of $K$, again written $\delta$. Now we have the inclusions

$$R[x;\alpha,\delta] \subseteq K[x;\alpha,\delta] \subseteq K(x;\alpha,\delta).$$

Any element $u \in K(x;\alpha,\delta)$ has the form $fg^{-1}$, where $f, g \in K[x;\alpha,\delta]$. By Prop. 1.7 we can bring the finite set of coefficients of $f, g$ to a common denominator, say $f = f_1 c^{-1}, g = g_1 c^{-1}$, where $f_1, g_1 \in R[x;\alpha,\delta]$, and $c \in R^\times$. Now $u = f_1 c^{-1}(g_1 c^{-1})^{-1} = f_1 g_1^{-1}$, so every element of $K(x;\alpha,\delta)$ can be written as a right fraction of elements of $R[x;\alpha,\delta]$, and hence the latter is right Ore, by Cor. 1.6. ■

We list some examples of skew polynomial rings. When the derivation is 0, we write $R[x;\alpha]$ in place of $R[x;\alpha,0]$.

1. $\alpha = 1, \delta = 0$. We obtain the ordinary polynomial ring $R[x]$ in a central indeterminate (although $R$ need not be commutative).

2. The *complex-skew* polynomial ring $\mathbf{C}[x;-]$ is the ring of polynomials with complex coefficients and commutation rule

$$ax = x\bar{a}, \qquad \text{where } \bar{a} \text{ is the complex conjugate of } a.$$

The centre of this ring is the ring $\mathbf{R}[x^2]$ of all real polynomials in $x^2$, and $\mathbf{C}[x;-]/(x^2+1)$ is the division algebra of real quaternions. More generally, let $k$ be a field of characteristic not 2 with a quadratic extension $K = k(\sqrt{b})$; this has an automorphism $\alpha$ given by $(x + y\sqrt{b})^\alpha = x - y\sqrt{b}$. For any $a \in k^\times$, $K[x;\alpha]/(x^2 - a)$ is the quaternion algebra $(a, b; k)$ (cf. **7.4**).

3. Let $K$ be any commutative ring and denote by $A_1[K]$ the $K$-algebra generated by $u, v$ over $K$ with the relation

$$uv - vu = 1. \tag{9}$$

This ring is called the *Weyl algebra* on $u, v$ over $K$. It may also be defined as the skew polynomial ring $R[v;1,']$, where $R = K[u]$ and $'$ denotes differentiation with respect to $u$. We observe that when $K$ is a Noetherian domain, then so is $A_1[K]$.

From (9) we obtain by induction on $n$,

$$u^n v - vu^n = nu^{n-1},$$

hence $u^m v^n \cdot v - v \cdot u^m v^n = mu^{m-1} v^n = \partial(u^m v^n)/\partial u$. A similar formula holds for commutation by $u$ and by linearity it follows that for any $f \in A_1[K]$,

$$fv - vf = \frac{\partial f}{\partial u}, \qquad uf - fu = \frac{\partial f}{\partial v}. \tag{10}$$

From these formulae it is easy to show that for a field $k$ of characteristic 0, $A_1[k]$ is a simple ring. For if $\mathfrak{a}$ is a non-zero ideal in $A_1[k]$, pick an element $f(u,v) \neq 0$ in $\mathfrak{a}$ of least possible degree in $u$. Then $\partial f/\partial u = fv - vf \in \mathfrak{a}$, but this has lower $u$-degree and so must be 0. Hence $f = f(v)$ is a polynomial in $v$ alone. If its $v$-degree is taken minimal, then $\mathrm{d}f/\mathrm{d}v = uf - fu = 0$ and so $f = c \in k$. Thus $\mathfrak{a}$ contains a non-zero element of $k$ and so must be the whole ring, i.e. $A_1[k]$ is simple, as claimed.

We observe that $A_1[k]$ is an example of a simple Noetherian domain, not a field. For a field $k$ of finite characteristic $p$, $A_1[k]$ is no longer simple, since it has the centre $k[u^p, v^p]$.

4. The *translation ring* $k\langle x, y | xy = y(x+1)\rangle$ may be described as $R = A[y;\sigma]$, where $A$ is the polynomial ring $k[x]$ with the shift automorphism $\sigma: x \mapsto x+1$.

5. Let $k$ be a field of prime characteristic $p$ and $F: a \mapsto a^p$ the Frobenius endomorphism. Then $k[x;F]$ is a skew polynomial ring whose field of fractions $k(x;F)$ has an inner automorphism inducing $F$, namely conjugation by $x$.

More generally, if $k$ is any field, even skew, with an endomorphism $\alpha$, then $k(x;\alpha)$ is an extension with an inner automorphism inducing $\alpha$ on $k$, because (3) now reads $ax = xa^\alpha$. Similarly, if $\delta$ is an $\alpha$-derivation, then $k[x;\alpha,\delta]$ is a ring with an inner $\alpha$-derivation inducing $\delta$, as we see by writing (3) in the form

$$a^\delta = ax - xa^\alpha.$$

6. Let $K$ be a commutative field with an automorphism $\alpha$ of order $n$, and consider the skew field of fractions $E = K(x;\alpha)$. If $k$ is the fixed field of $\alpha$, then $F = k(x^n)$ is contained in the centre of $E$, as is easily checked. Moreover, $K(x^n)$ is a commutative subfield, a Galois extension of $F$ of degree $n$, and provided that $K$ contains a primitive $n$th root of 1, the structure of $E$ is given by the equations

$$ax^i = x^i a^{\alpha^i} \qquad \text{for all } a \in K, i = 0, 1, \ldots, n-1.$$

It follows that $k(x^n)$ is the precise centre of $E$ and $E$ is of dimension $n^2$ over its centre, in fact a crossed product (cf. **7.5**).

7. If $D$ is a skew field not algebraic over its centre $k$, then $D \otimes k(t)$ is a simple Noetherian domain (by Th. 7.1.2), but not Artinian, hence not a skew field.

8. Let $R$ be an integral domain with an automorphism $\alpha$. In the skew polynomial ring $R[x;\alpha]$ the powers of $x$ form an Ore set, and the ring of fractions consists of all polynomials $\sum_{-r}^{s} x^i a_i$, involving negative as well as positive powers of $x$. Such an expression is called a *skew Laurent polynomial* and the resulting ring may be written $R[x, x^{-1};\alpha]$.

For each polynomial $f$ of the form (1) we can also define its *order* $\mathrm{o}(f)$ as the lowest power of $x$ occurring with a non-zero coefficient:

$$\mathrm{o}(f) = \min\{i | a_i \neq 0\}.$$

This function has the properties of a valuation on $P$:

**O.1** $o(f) \geqslant 0$ *for all* $f \in P$, $o(0) = \infty$,
**O.2** $o(f-g) \geqslant \min\{o(f), o(g)\}$,
**O.3** $o(fg) = o(f) + o(g)$.

Taking first the case $\delta = 0$, we can form the ring $R[\![x;\alpha]\!]$ of formal power series over $R$ as the set of all infinite series

$$f = a_0 + xa_1 + x^2a_2 + \cdots, \tag{11}$$

with componentwise addition and with multiplication based on the commutation rule $ax = xa^{\alpha}$. There is of course no question of convergence here; we regard (11) as a series in a purely formal sense. We can describe $f$ equally well as an infinite sequence $(a_i) = (a_0, a_1, \ldots)$, with addition $(a_i) + (b_i) = (a_i + b_i)$ and with multiplication

$$(a_i)(b_j) = (c_k), \qquad \text{where } c_k = \sum a_{k-j}^{\alpha^j} b_j. \tag{12}$$

Alternatively we can regard $R[\![x;\alpha]\!]$ as the completion of the skew polynomial ring $R[x;\alpha]$ with respect to the powers of the ideal generated by $x$; these powers define a topology called the *x-adic* topology. This topological viewpoint is not essential for the construction, but it helps in understanding the situation.

Let $R$ be a ring and $\alpha$ an automorphism of $R$. The powers of $x$ form an Ore set in $R[\![x;\alpha]\!]$ and by taking fractions we obtain the ring $R((x;\alpha))$ of all formal Laurent series

$$\sum_{-r}^{\infty} x^i a_i = x^{-r}a_{-r} + \cdots + x^{-1}a_{-1} + a_0 + xa_1 + x^2a_2 + \cdots. \tag{13}$$

This is again a ring, with the same multiplication (12); here the restriction to finitely many negative powers is necessary to ensure that the multiplication rule (12) has a sense. This is also the reason for taking $\alpha$ to be an automorphism, since now $j$ may take negative values in (12).

Let us now consider a skew polynomial ring $R[x;\alpha,\delta]$, where $\delta$ may be non-zero, but $\alpha$ is still an automorphism, and ask whether a power series ring can be formed. If we attempt to define the multiplication of power series by means of the commutation formula (3), we shall find that (apart from a more complicated form for the coefficients of the product), the product $cf$, where $c \in R$, cannot always be expressed as a power series, because there will in general be contributions to the coefficient of a given power $x^r$ from each term $cx^n a_n (n \geqslant r)$ and so we may have infinitely many such terms to consider. In terms of the $x$-adic topology we can express this by saying that left multiplication by $c \in R$ is not continuous; this follows from (3), because when $a^{\delta} \neq 0$, we have $o(ax) < o(x)$.

One way to overcome this difficulty is to introduce $y = x^{-1}$ and rewrite (3) in terms of $y$. We find

$$\begin{aligned} ya &= a^{\alpha}y + ya^{\delta}y \\ &= a^{\alpha}y + a^{\delta\alpha}y^2 + ya^{\delta^2}y^2 = \cdots, \end{aligned}$$

hence by induction we obtain

$$ya = a^{\alpha}y + a^{\delta\alpha}y^2 + a^{\delta\alpha^2}y^3 + \cdots. \tag{14}$$

With the help of this commutation formula we can multiply power series in $y$ and even Laurent series. We observe that in passing from $x$ to $y = x^{-1}$ we have also had to change the side on which the coefficients are put; of course this is immaterial as long as $\alpha$ is an automorphism. To be precise, from any skew polynomial ring $R[x; \alpha, \delta]$ we can form a skew power series ring in $x^{-1}$, with coefficients on the left; in order to define Laurent series in $x^{-1}$ we need to assume that $\alpha$ is an automorphism. We shall not pursue this point but note one result which illustrates the usefulness of power series.

THEOREM 3.5 *Let $K$ be a skew field and $\alpha$ an automorphism of $K$. Further denote the centre of $K$ by $C$ and the subfield of $C$ fixed under $\alpha$ by $C_0$. If no positive power of $\alpha$ is an inner automorphism of $K$, then the centre of $K(x; \alpha)$ is $C_0$.*

*Proof.* Every element of $K(x; \alpha)$ can be written as a Laurent series $f = \sum x^i a_i$. If this lies in the centre, then $fc = cf$ for all $c \in K$, i.e. $\sum x^i(a_i c - c^{\alpha^i} a_i) = 0$, hence

$$a_i c = c^{\alpha^i} a_i \qquad \text{for all } i \in \mathbf{Z} \text{ and all } c \in K. \tag{15}$$

If $a_i \neq 0$ for some $i > 0$, then $\alpha^i$ is inner, by (15); in case $i < 0$, $\alpha^{-i}$ is inner, but this contradicts the hypothesis. Hence $a_i = 0$ for $i \neq 0$ and $f = a_0 \in K$. Now (15) reads $a_0 c = c a_0$, so $a_0 \in C$, and since $x a_0 = a_0 x = x a_0^{\alpha}$, we have $f = a_0 \in C_0$. Conversely, every element of $C_0$ commutes with every element of $K(x; \alpha)$. ■

Finally we note a rationality criterion for power series, which applies in the skew case:

PROPOSITION 3.6 *Let $K$ be a skew field with an automorphism $\alpha$, and consider the natural embedding of $K(x; \alpha)$ in the skew field $K((x; \alpha))$ of formal Laurent series. A given series $f = \sum x^i a_i$ lies in $K(x; \alpha)$ if and only if there exist integers $r, n_0$ and elements $c_1, \ldots, c_r \in K$ such that*

$$a_n = a_{n-1}^{\alpha} c_1 + a_{n-2}^{\alpha^2} c_2 + \cdots + a_{n-r}^{\alpha^r} c_r \qquad \text{for all } n > n_0. \tag{16}$$

*Proof.* The series $f$ lies in $K(x; \alpha)$ iff there is a polynomial $g$ with constant term 1 such that $fg$ is a Laurent polynomial. Writing $g = 1 - xc_1 - \cdots - x^r c_r$, we require that in the product $(\sum x^i a_i)(1 - \sum x^j c_j)$ all coefficients of powers

beyond a given one, say $x^{n_0}$, vanish. On equating the coefficient of $x^n$ to 0 we just obtain (16), and the conclusion follows. ■

## Exercises

(1) Supply the details of the proof of Th. 3.3, and point out where the fact that $\alpha$ is an automorphism is used.

(2) Let $K$ be a skew field and $\alpha$ an endomorphism. Show that $K[x;\alpha]$ is a left Ore domain iff $\alpha$ is surjective (and hence an automorphism). Using Prop. 1.8, obtain an embedding of the free algebra $K\langle x, y\rangle$ in a right Ore domain, and hence an embedding of $K\langle x, y\rangle$ in a skew field. (Jategaonkar, Koshevoi).

(3) Find a localization of the Weyl algebra over a field of finite characteristic, which is a crossed product.

(4) Show that for any ring $K$, a Weyl $K$-ring $A_1[K]$ on $u, v$ may be defined by (9), where $u, v$ commute with the elements of $K$. Show that if $K$ is a simple ring of characteristic 0, then so is $A_1[K]$.

(5) Show that if $R$ is Noetherian and $\alpha$ is an automorphism, then $R[x, x^{-1};\alpha]$ is again Noetherian.

(6) Let $K$ be a ring and $\alpha$ an injective endomorphism of $K$. Put $K_0 = K$ and take an ascending chain of $K$-rings $K_n$, all isomorphic to $K$, such that $K_{n-1}$ is identified with the image of $K_n$ under $\alpha$. Show that their union is a ring $K^{[\alpha]}$ to which $\alpha$ can be extended as an automorphism. Verify that within the ring $K[x, x^{-1};\alpha]$ the subring $\cup x^m K x^{-m}$ is isomorphic to $K^{[\alpha]}$.

(7) Let $F \subseteq E$ be any fields and $\alpha$ an automorphism of $E$ mapping $F$ into itself. Show that $E(x;\alpha) \cap F((x;\alpha)) = F(x;\alpha)$.

(8) Let $K$ be a skew field with centre $C, \alpha$ an automorphism of $K$ and $C_0$ the subfield of $C$ left fixed by $\alpha$. Suppose further that $\alpha^n = 1$, but no power $\alpha^i (0 < i < n)$ is inner. Show that the centre of $K((x;\alpha))$ is $C_0((x^n))$ and the centre of $K(x;\alpha)$ is $C_0(x^n)$.

(9) (P. Draxl) Let $k$ be a field of characteristic not equal to 2 and $K = k(u_1, u_2, u_3, u_4)$, where the $u_i$ are independent indeterminates. Put $K_1 = k((x_1))$, $K_2 = K_1((x_2;\alpha))$, $K_3 = K_2((x_3))$, $K_4 = K_3((x_4;\beta))$, where $\alpha: x_1 \mapsto -x_1$, $\beta: x_3 \mapsto -x_3$, and by identifying $u_i$ with $x_i^2$ show that the $K$-subalgebra $D$ of $K_4$ generated by $x_1, \ldots, x_4$ is a central division algebra of degree 4 over $K$, of the form $D = (u_1, u_2; K) \otimes (u_3, u_4; K)$. Using the representation as Laurent series, show that any multiplicative commutator in $D$ has the form $1 + f$, where $f$ involves only positive powers of the $x$'s. Deduce that if $k$ contains a primitive 4th root of 1, then $\mathbf{SK}_1(D)$ is non-trivial. (Hint: use Ex. (8) to show that $D$ has centre $K$ and verify that the 4th root of 1 has reduced norm 1. For more details of this solution to the Tannaka–Artin problem see Draxl [1983].)

(10) Show that the trace group, defined in **6.5**, vanishes on the Weyl algebra over a field of characteristic 0.

## 9.4 Goldie's theorem

The main structure theorem for Artinian rings states that a semisimple ring is a direct product of simple rings and each of the latter is a full matrix ring over a skew field (Wedderburn theorems, **5.3** of A.2). Such precise information is not to be expected for Noetherian rings, but in this case there is a reduction theorem due to Goldie, which implies that a Noetherian semiprime ring has a quotient ring which is semisimple (hence Artinian) and here prime Noetherian rings correspond to simple Artinian rings. Our aim in this section is a proof of these results, but some preparation will be necessary. We recall that a ring $R$ is said to be *prime* if $R \neq 0$ and $aRb = 0$ implies $a = 0$ or $b = 0$; a ring $R$ in which $aRa = 0$ implies $a = 0$ is called *semiprime*. The precise relation between these concepts will be described in Ch. 10.

We shall also need a substitute for the dimension of a vector space which can be applied to general modules; this is the notion of uniform rank. Let $R$ be a ring, $M$ an $R$-module and $M'$ a submodule of $M$. Then $M'$ is said to be *large* in $M$ and $M$ is called an *essential extension* of $M'$ if

$$N \neq 0 \Rightarrow N \cap M' \neq 0, \qquad \text{for any submodule } N \text{ of } M. \tag{1}$$

In particular, $R$ itself may be considered as a left or right $R$-module; we then obtain large left or right ideals, also called *left large* or *right large*, respectively.

We list some properties of large submodules, used later:

**L.1** *If $M'$ is large in $M$ and $M''$ large in $M'$ then $M''$ is large in $M$.*

For if $0 \neq N \subseteq M$, then $N \cap M' \neq 0$, hence $N \cap M'' = (N \cap M') \cap M'' \neq 0$. ■

**L.2** *Let $M$ be a right $R$-module and $M'$ a large submodule of $M$. For any $m \in M$, the set $\mathfrak{c} = \{x \in R \mid mx \in M'\}$ is a large right ideal in $R$, and if $m \neq 0$, then $0 \neq m\mathfrak{c} \subseteq M'$.*

Clearly $\mathfrak{c}$ is a right ideal; for any non-zero right ideal $\mathfrak{a}$ in $R$, either $m\mathfrak{a} = 0$, in which case $\mathfrak{a} \subseteq \mathfrak{c}$, or $m\mathfrak{a} \neq 0$; then $mR \cap M' \neq 0$ and for any $a \in \mathfrak{a}$ such that $0 \neq ma \in M'$ we have $0 \neq aR \subseteq \mathfrak{c} \cap \mathfrak{a}$, which shows $\mathfrak{c}$ to be right large. Moreover, if $ma \in M'$, then $a \in \mathfrak{c}$, so $0 \neq m\mathfrak{c} \subseteq M'$. ■

**L.3** *For any nilpotent ideal $\mathfrak{n}$ of $R$, the left annihilator $(\mathfrak{n})_l = \{x \in R \mid x\mathfrak{n} = 0\}$ is a right large ideal of $R$.*

For, given $c \neq 0$, there exists $s \geqslant 1$ such that $c\mathfrak{n}^{s-1} \neq 0$, $c\mathfrak{n}^s = 0$, and $c\mathfrak{n}^{s-1} \subseteq cR \cap (\mathfrak{n})_l$. ■

**L.4** *A module $M$ is semisimple if and only if it has no proper large submodules.*

For in a semisimple module every proper submodule has a non-zero complement and so cannot be large. Conversely, if $M$ has no proper large submodules and $M' \subseteq M$ is given, we can by Zorn's lemma find a submodule $N$ which is maximal subject to $N \cap M' = 0$. If $P$ is a non-zero submodule such that $P \cap (N + M') = 0$, then the sum $P + N + M'$ is direct and so $(P + N) \cap M' = 0$, but this contradicts the maximality of $N$. Hence $P \cap (N + M') \neq 0$ for all $P \neq 0$, so $N + M'$ is large and therefore equal to $M$. Thus $N$ is a complement of $M'$ and this shows $M$ to be semisimple. ■

A module $M$ is said to be *uniform* if $M \neq 0$ and every non-zero submodule of $M$ is large, thus any two non-zero submodules of $M$ have a non-zero intersection. For example, an integral domain is right uniform (i.e. uniform as right $R$-module) iff it is a right Ore domain.

With the help of uniform modules we can define a weak form of dependence relation which leads to a notion of rank in general modules. Let $M$ be an $R$-module (for any ring $R$) and denote by $\mathscr{U}$ the collection of all its uniform submodules. On $\mathscr{U}$ we introduce a dependence relation as follows. If $Q, P_1, \ldots, P_r \in \mathscr{U}$, then $Q$ is said to be *dependent* on $P_1, \ldots, P_r$ if $Q \cap \sum P_i \neq 0$. Generally $Q$ is said to be dependent on a (possibly infinite) family of uniform submodules if it is dependent on a finite subfamily. A set of uniform modules is *independent* if no member is dependent on the rest. We claim that this dependence relation satisfies **D.**0, 1′ and 2 (though not **D.**1) of **1.4**. **D.**0 is clear; to prove **D.**1′, we may take our family to be finite, say $Q$ is dependent on the independent family $\{P_1, \ldots, P_r\}$ and each $P_i$ is dependent on $\{N_1, \ldots, N_s\}$. By hypothesis, $Q \cap \sum P_i \neq 0$, so there exists $q \in Q$, $q \neq 0$, such that

$$q = p_1 + \cdots + p_r, \qquad \text{where } p_i \in P_i. \tag{2}$$

Writing $N = \sum N_j$, we have to show that $Q \cap N \neq 0$. If $p_i \in N$ for all $i$, the conclusion follows from (2); otherwise we choose an equation (2) with $q \neq 0$ such that the least number of $p_i$ are *not* in $N$. If $p_1 \notin N$ say, then since $P_1$ is uniform, $p_1 R \cap (P_1 \cap N) \neq 0$, say $0 \neq p_1 c \in N$. Now in

$$qc = p_1 c + \cdots + p_r c \tag{3}$$

there are fewer terms outside $N$ than in (2), and $qc \neq 0$, because $p_1 c \neq 0$ and the sum $\sum P_i$ is direct. This contradiction shows that $Q \cap N \neq 0$, as claimed.

The exchange axiom **D.**2 follows easily: if $Q$ is dependent on $P_1, \ldots, P_r$ but not on $P_2, \ldots, P_r$ then there exists $q \in Q, q \neq 0$, such that (2) holds with $p_1 \neq 0$,

by hypothesis. If we now rewrite (2) as $p_1 = q - p_2 - \cdots - p_r$, we have a relation showing $P_1$ to be dependent on $Q, P_2, \ldots, P_r$, so **D**.2 holds. ■

Let us also note that if $Q$ is dependent on a family $\mathscr{F}$ of uniform submodules, then so is any non-zero submodule of $Q$. For if $Q \cap P \neq 0$, where $P$ is a sum of terms in $\mathscr{F}$, and $0 \neq Q' \subseteq Q$, then $Q' \cap P = Q' \cap (Q \cap P) \neq 0$, because $Q$ is uniform.

We can now apply the results of **1.4**. A *basis* of $M$ will be a family of uniform submodules whose sum is direct and large in $M$. Such bases need not exist, but if every non-zero submodule of $M$ contains a uniform submodule, then we can by Zorn's lemma find a direct sum of uniform submodules which is large in $M$, and hence obtain a basis. We shall mainly be interested in modules with a finite basis, and we shall find conditions for this to exist in Prop. 4.1 below.

From Cor. 1.4.6 we see that if $M$ has a finite basis, then any two bases have the same number of terms. This number is called the *uniform rank* or simply the *rank* of $M$, and is written rk $M$. It is clear that rk $M = 1$ iff $M$ is uniform, and in a module of rank $n$, any direct sum of non-zero submodules has at most $n$ terms; it has exactly $n$ terms iff each term is uniform and the sum is large in $M$. The conditions for a module to have finite rank are easily stated:

PROPOSITION 4.1 *Let $R$ be any ring and $M$ an $R$-module which contains no infinite direct sums of non-zero submodules. Then there is a direct sum of uniform submodules which is large in $M$, so that $M$ has a rank, and in fact the rank is then finite. Conversely, a module of finite rank contains uniform submodules, but no infinite direct sum of submodules.*

*Proof.* We begin by showing that every non-zero submodule $N$ of $M$ contains a uniform submodule. For if $N$ is not itself uniform, then it contains a direct sum $N_1 \oplus N_1'$; now either $N_1'$ is uniform or it contains a direct sum $N_2 \oplus N_2'$ and continuing in this way we obtain in $M$ the direct sum

$$N_1 \oplus N_2 \oplus \cdots.$$

But $M$ contains no infinite direct sums, so the process must break off, which can only happen when we reach an $N_i'$ which is uniform. Hence $N$ contains a uniform submodule.

Now let $\sum_1^r U_i$ be a direct sum of uniform submodules in $M$; such a sum exists, e.g. for $r = 0$. Either it is large in $M$ or we can find $V \neq 0$ such that $V \cap \sum U_i = 0$. By the first part of the proof $V$ contains a uniform submodule $U_{r+1}$ and now $\sum_1^{r+1} U_i$ is a direct sum of $r+1$ terms. If we continue in this way, the process must break off, since $M$ has no infinite direct sums, and it can end only when we have a direct sum of uniform submodules which is large in $M$. This shows that $M$ has a rank and that this rank is finite. Conversely, if rk $M = n$, we know that any direct sum contains part of a basis and so cannot have more than $n$ terms. ■

This result may be applied to $R$ itself, as left or right $R$-module, and in this way we obtain the notion of *left rank* and *right rank* of $R$. For example, an integral domain $R$ has right rank 1 iff it is a right Ore domain. In fact, by Prop. 1.9 we obtain

COROLLARY 4.2 *An integral domain which has finite right rank, necessarily has right rank* 1. ■

We remark that for a submodule $M'$ of $M$ we have $\operatorname{rk} M' \leqslant \operatorname{rk} M$, with equality iff $M'$ is large in $M$. On the other hand, going over to a quotient module may well raise the rank, e.g. $\operatorname{rk} \mathbf{Z} = 1$, but $\operatorname{rk}(\mathbf{Z}/m) > 1$ unless $m$ is a prime power.

Let $R$ be any ring and $M$ a right $R$-module. For any subset $S$ of $M$ we define the *right annihilator* of $S$ in $R$ as

$$(S)_r = \{x \in R \mid Sx = 0\}.$$

It is clear that $(S)_r$ is a right ideal; if $S$ is a submodule, $(S)_r$ is even a two-sided ideal. When $S \neq 0$, $(S)_r$ cannot contain 1 and so will be proper. If $S = \{a\}$, we write $(a)_r$ instead of $(\{a\})_r$. In particular, this defines right annihilators of subsets of $R$; the *left* annihilator of a subset $S$ of $R$ (or more generally, of a left $R$-module) is defined similarly. The ring $R$ is said to satisfy the *maximum condition on right annihilators* if every collection of right annihilators in $R$ (of subsets of $R$) has a maximal member; e.g. any right Noetherian ring satisfies the maximum condition on right annihilators.

PROPOSITION 4.3 *Let $R$ be a semiprime ring with maximum condition on right annihilators. Then every nil left or right ideal is* 0.

*Proof.* It is clearly enough to prove the result for principal ideals, and we need only consider left ideals, for $Ra$ is nil iff $(xa)^n = 0$ for all $x$ and suitable $n = n(x)$. Now $(xa)^n = 0$ implies $(ax)^{n+1} = a(xa)^n x = 0$, hence if $Ra$ is nil, then so is $aR$.

Thus let $Ra$ be a non-zero nil left ideal and choose a maximal annihilator not equal to $R$ of the form $(xa)_r$. Writing $b = xa$, we choose $y \in R$; if $yb \neq 0$, take $v \geqslant 2$ such that $(yb)^{v-1} \neq 0$, $(yb)^v = 0$. Then $(b)_r \subseteq ((yb)^{v-1})_r$ and by maximality we have equality here; since $yb \in ((yb)^{v-1})_r$, we conclude that $byb = 0$. This holds for all $y \in R$, even when $yb = 0$. Hence $bRb = 0$, $b \neq 0$, in contradiction to the fact that $R$ is semiprime. Hence every nil left (or right) ideal is 0. ■

We shall also need the notion of a singular submodule. For any ring $R$ and any right $R$-module $M$ consider the set

$$Z(M) = \{m \in M \mid (m)_r \text{ is a large right ideal of } R\}.$$

This set is a submodule of $M$, called the *singular submodule*. To verify the module

property, let $u, v \in Z(M)$; then $\mathfrak{a} = (u)_r$ and $\mathfrak{b} = (v)_r$ are right large, hence so is $\mathfrak{a} \cap \mathfrak{b}$ and $(u - v)(\mathfrak{a} \cap \mathfrak{b}) = 0$, therefore $(u - v)_r$ is right large. Further, if $a \in R$, then $(ua)_r$ is right large, by **L.2**, for $x \in (ua)_r$ iff $ax \in (u)_r$ and the latter is right large by hypothesis. Thus $Z(M)$ is indeed a submodule. In particular, taking $M = R$, we obtain the *right singular ideal* $Z(R)$ of $R$. By what has been shown, it is a right ideal; in fact it is two-sided, for if $(a)_r$ is right large, then so is $(ba)_r \supseteq (a)_r$.

Although Goldie's theorem is concerned with Noetherian rings, it applies to a somewhat wider class, defined as follows. A ring which is of finite right rank and satisfies the maximum condition on right annihilators is called a *right Goldie ring*. In particular, every right Noetherian ring is right Goldie; of course the converse is false, as the example of commutative integral domains shows. In a right Goldie ring the right singular ideal is nilpotent; we shall only need the special case where the ring is semiprime, when it is a consequence of the next result (cf. Ex. (8) of **10.6** for the general case).

PROPOSITION 4.4 *In a right Goldie ring $R$, for each $a \in R$ there exists $n \geqslant 0$ such that $a^n R + (a^n)_r$ is right large. Moreover, $(a^v)_r = (a^n)_r$ for all $v \geqslant n$ and the sum $a^n R + (a^n)_r$ is direct.*

*Proof.* The sequence $(a)_r \subseteq (a^2)_r \subseteq \cdots$ becomes stationary, say $(a^v)_r = (a^n)_r$ for $v \geqslant n$. It follows that $(a^n)_r \cap a^n R = 0$, for if $a^n \cdot a^n x = 0$, then $x \in (a^{2n})_r = (a^n)_r$, hence $a^n x = 0$. If $\mathfrak{c}$ is any right ideal such that $\mathfrak{c} \cap (a^n R + (a^n)_r) = 0$, then the sum $\mathfrak{c} + a^n \mathfrak{c} + a^{2n} \mathfrak{c} + \cdots$ is direct, for if $a^{tn} c_t + a^{(t+1)n} c_{t+1} + \cdots + a^{sn} c_s = 0$, where $c_i \in \mathfrak{c}$, $c_t \neq 0$, then $c_t \in (a^n)_r + a^n R$, which is a contradiction. Since $R$ has finite rank, $a^{tn} \mathfrak{c} = 0$ for some $t \geqslant 1$, i.e. $\mathfrak{c} \subseteq (a^{tn})_r = (a^n)_r$ and it follows that $\mathfrak{c} = \mathfrak{c} \cap (a^n R + (a^n)_r) \neq 0$. This proves that $a^n R + (a^n)_r$ is right large, and we have seen that this sum is direct. ■

We note two extreme special cases of this result:

COROLLARY 4.5 *In a right Goldie ring $R$, if $a$ is left regular, then $aR$ is right large.*

For in this case $(a^n)_r = 0$ for all $n$, and if $a^n R$ is right large, then so is $aR$. ■

COROLLARY 4.6 *In a semiprime right Goldie ring $R$, $Z(R) = 0$.*

For by Prop. 4.4, if $(a)_r$ is right large, then $a$ is nilpotent. Hence $Z(R)$ is a nil ideal, and so must be zero, by Prop. 4.3. ■

The next lemma, essentially the converse of Cor. 4.5, will be needed for Goldie's theorem, but is also useful elsewhere.

LEMMA 4.7 *In a semiprime right Goldie ring any large right ideal contains a regular element.*

*Proof.* By Prop. 4.3, any nil right ideal of $R$ is 0, so a non-zero right ideal $\mathfrak{a}$ will contain a non-nilpotent element $a$. By Prop. 4.4 we can find a power $a_1$ of $a$ such that $(a_1)_r = (a_1^2)_r$ and so $a_1R + (a_1)_r$ is direct. If $(a_1)_r \cap \mathfrak{a} \neq 0$, we choose $a_2 \in (a_1)_r \cap \mathfrak{a}$ such that $a_2 \neq 0$ and $(a_2)_r = (a_2^2)_r$. Then $a_2R + ((a_1)_r \cap (a_2)_r \cap \mathfrak{a})$ is a direct sum contained in $(a_1)_r \cap \mathfrak{a}$, hence the sum $(a_1R + a_2R + ((a_1)_r \cap (a_2)_r \cap \mathfrak{a})$ is direct. If $(a_1)_r \cap (a_2)_r \cap \mathfrak{a} \neq 0$, we can continue the process; at the $n$th stage we have a direct sum

$$a_1R + \cdots + a_nR + ((a_1)_r \cap \cdots \cap (a_n)_r \cap \mathfrak{a}), \tag{4}$$

where $a_i \in (a_1)_r \cap \cdots \cap (a_{i-1})_r \cap \mathfrak{a}$ and $(a_i)_r = (a_i^2)_r \neq R$.

Since $R$ has finite right rank, the process must stop; if this happens at the $n$th stage, we have

$$(a_1)_r \cap \cdots \cap (a_n)_r \cap \mathfrak{a} = 0. \tag{5}$$

So far $\mathfrak{a}$ was any right ideal $\neq 0$; if we take $\mathfrak{a}$ to be right large, then by (5) we have

$$(a_1)_r \cap \cdots \cap (a_n)_r = 0. \tag{6}$$

Put $c = \sum a_i$; by construction, $c \in \mathfrak{a}$ and we claim that $c$ is regular. If $cx = \sum a_i x = 0$, then by the directness of (4), $a_i x = 0$ for all $i$, hence $x = 0$ by (6), i.e. $(c)_r = 0$. By Cor. 4.5, $cR$ itself is right large, hence $(c)_l \subseteq Z(R)$, but $Z(R) = 0$ by Cor. 4.6, so $(c)_l = 0$ and $c$ is regular. ■

In the presence of maximum conditions the definition of a right Ore set can be simplified a little; this is sometimes useful, although it is not actually needed here.

PROPOSITION 4.8 *Let $R$ be a ring with maximum condition on right annihilators and let $S$ be a multiplicative subset of $R$ such that*

(i) *for any $a \in R$, $s \in S$, $aS \cap sR \neq \varnothing$,*

(ii) *for any $a \in R$, $s \in S$, $as = 0 \Rightarrow a = 0$.*

*Then $S$ is a right Ore set consisting of regular elements.*

*Proof.* We need only show that for any $a \in R$, $s \in S$, $sa = 0 \Rightarrow a = 0$. By the maximum condition the sequence

$$(s)_r \subseteq (s^2)_r \subseteq \cdots$$

becomes stationary, say $(s^\nu)_r = (s^{\nu+1})_r$. If $sa = 0$, then by (i) there exist $s', a'$ such that $as' = s^\nu a'$; hence $s^{\nu+1}a' = sas' = 0$, so $a' \in (s^{\nu+1})_r = (s^\nu)_r$, and so $0 = s^\nu a' = as'$, hence $a = 0$ by (ii). ■

We now come to the main result of this section.

THEOREM 4.9 (Goldie's theorem) *A ring $R$ has a total quotient ring $Q$ which is semisimple if and only if $R$ is a semiprime right Goldie ring. Moreover, $Q$ is simple Artinian if and only if $R$ is prime Goldie.*

*Proof.* Assume that $R$ is semiprime right Goldie and let $S$ denote the set of all regular elements of $R$. We shall show that $S$ is a right Ore set. Given $a \in R$, $s \in S$, define

$$\mathfrak{c} = \{x \in R \mid ax \in sR\}.$$

By Cor. 4.5, $sR$ is right large, hence by **L.2**, so is $\mathfrak{c}$, therefore it contains a regular element (Lemma 4.7). This shows that $sR \cap aS \neq \emptyset$, and so $S$ is a right Ore set.

Let $Q = R_S$ be the total quotient ring. If $\mathfrak{A}$ is a large right ideal in $Q$, then $\mathfrak{A} \cap R$ is right large in $R$ and so contains a regular element. This must be a unit in $Q$, hence $\mathfrak{A} = Q$, i.e. $Q$ has no proper large right ideals, and so is semisimple, by **L.4**.

Conversely, let $R$ be a ring with semisimple quotient ring $Q$. We shall show that for any right ideal $\mathfrak{c}$ of $R$ the following conditions are equivalent:

(a) $\mathfrak{c}$ *is right large in* $R$,
(b) $\mathfrak{c}Q = Q$,
(c) $\mathfrak{c}$ *contains a regular element of* $R$.

(a)$\Leftrightarrow$(b). Assume that $\mathfrak{c}$ is right large and let $\mathfrak{A}$ be any non-zero right ideal of $Q$. Then $\mathfrak{A} \cap R \neq 0$, hence $\mathfrak{A} \cap R \cap \mathfrak{c} \neq 0$, and so $\mathfrak{A} \cap \mathfrak{c}Q \neq 0$. This shows that $\mathfrak{c}Q$ is right large in $Q$, hence $\mathfrak{c}Q = Q$ by **L.4**. Conversely, if $\mathfrak{c}Q = Q$ and $\mathfrak{a}$ is a non-zero right ideal in $R$, then $\mathfrak{a}Q \cap \mathfrak{c}Q \neq 0$, hence $\mathfrak{a} \cap \mathfrak{c} \neq 0$, so $\mathfrak{c}$ is right large in $R$.

(b)$\Leftrightarrow$(c). If $\mathfrak{c}Q = Q$, then $1 = as^{-1}$, where $a \in \mathfrak{c}$ and $s$ is a regular element in $R$, hence $a = s$ is regular. Conversely, if $\mathfrak{c}$ contains a regular element, then clearly $\mathfrak{c}Q = Q$.

We can now complete the proof of the theorem by verifying that $R$ is semiprime right Goldie.

Let $\mathfrak{n}$ be a nilpotent ideal of $R$. Then $(\mathfrak{n})_l$ is a right large ideal in $R$, by **L.3**, so it contains a regular element, and it follows that $\mathfrak{n} = 0$. Thus $R$ is semiprime.

Next let $\mathfrak{a} = \sum \mathfrak{a}_\lambda$ be a direct sum of non-zero right ideals in $R$ which is right large; then $\mathfrak{a}$ contains a regular element $c$, say:

$$c = x_1 + \cdots + x_n, \qquad \text{where } x_i \in \mathfrak{a}_{\lambda_i}.$$

Now $cR$ is right large and is contained in $\mathfrak{a}_{\lambda_1} + \cdots + \mathfrak{a}_{\lambda_n}$, hence the latter sum is right large, so the sum $\sum \mathfrak{a}_\lambda$ was finite.

For any subset $I$ of $R$, its right annihilator in $R$ is obtained by intersecting its annihilator in $Q$ with $R$. In an obvious notation we have

$$(I)_r^R = (I)_r^Q \cap R.$$

Now the maximum condition for right annihilators follows for $R$, because it holds in $Q$.

Finally, if $R$ is prime, then so is $Q$, by Prop. 1.10. It follows that $Q$ is simple. Conversely, when $Q$ is simple, $R$ must be prime, for if $\mathfrak{a}$, $\mathfrak{b}$ are ideals in $R$ such that $\mathfrak{ab} = 0$, then $\mathfrak{b}Q\mathfrak{a} \cap R$ is an ideal of $R$ such that $(\mathfrak{b}Q\mathfrak{a} \cap R)^2 = 0$. Since $R$ is semiprime, we have $\mathfrak{b}Q\mathfrak{a} \cap R = 0$, hence $\mathfrak{b}Q\mathfrak{a} = 0$. But $Q$ is simple and it follows that $\mathfrak{a} = 0$ or $\mathfrak{b} = 0$, and this shows that $R$ is prime. ■

For prime rings Th. 4.9 was proved by Goldie (1958) and extended by him to semiprime rings in 1960. The result raises the question of localizing relative to a prime ideal. If $R$ is a Noetherian ring with a prime ideal $\mathfrak{p}$, and $C_{\mathfrak{p}}$ denotes the set of all elements of $R$ that are *regular* mod $\mathfrak{p}$, i.e. whose image in $R/\mathfrak{p}$ is regular, then it is not necessarily the case that $C_{\mathfrak{p}}$ is a right Ore set. Here it is necessary to consider more than one prime (a so-called 'clique' of prime ideals), and to localize at the set of elements that are regular modulo all the prime ideals considered. For a detailed account of this method see Jategaonkar [1986], McConnell and Robson [1987] or Goodearl and Warfield [1989].

## Exercises

(1) Show that the socle of a module is the intersection of all its large submodules. (Hint: show first that every submodule is a direct summand of a large submodule.)

(2) Let $A$ be the direct sum of an infinite and a finite (non-zero) cyclic group. Show that for the dependence relation defined for $A$ as $\mathbf{Z}$-module, **D.**1 does not hold. Show also that **D.**1 does hold on any torsion-free module.

(3) Show that a non-zero submodule of a direct sum of uniform modules has a uniform submodule. (Hint: first find a non-zero submodule in a finite direct sum.)

(4) Find the rank of $\mathbf{Z}/m$, for a positive integer $m$.

(5) Show that for any ring $R$, $\operatorname{rk} \mathfrak{M}_n(R) = n \cdot \operatorname{rk} R$.

(6) Show that the injective hull of a uniform module is indecomposable. Use this fact and the Krull–Schmidt theorem to give another proof of Prop. 4.1.

(7) Let $R$ be a ring with a total quotient ring $Q$. Show that $Q$ is semisimple whenever every right $Q$-module is injective as right $R$-module. (Hint: take a right ideal in $Q$, form its complement $C$ as right $R$-module and verify that $C$ is a $Q$-module.)

(8) Show that every left and right self-injective ring is its own quotient ring ($R$ is *right self-injective* if $R_R$ is injective).

(9) Let $F$ be the ring of all real continuous functions on the unit interval with pointwise operations: $(f+g)(x) = f(x) + g(x)$, $(fg)(x) = f(x)g(x)$. Show that $F$ has no uniform ideals.

(10) Let $R$ be a ring with elements $a, b$ such that $ab = 1 \neq ba$. Put $e_{11} = 1 - ba$, $e_{ij} = b^{i-1}e_{11}a^{j-1}$ and show that $e_{ij}e_{kl} = \delta_{jk}e_{il}$. Verify that the $e_{ii}$ form a family of pairwise orthogonal idempotents. Deduce that $R$ has infinite (left or right) rank and does not satisfy the maximum condition on left or right annihilators.

(11) Show that the maximum condition on left annihilators is equivalent to the minimum condition on right annihilators.

(12) Show that a reduced ring ($x^2 = 0 \Rightarrow x = 0$) is non-singular.

(13) Show that over an Ore domain every finitely generated flat module is projective. (Hint: use Ex. (16), Ch. 6.)

(14) Show that a module of finite rank contains a large submodule which is a direct sum of uniform modules.

(15) Show that an Artinian semiprime ring is semisimple.

## 9.5 PI-algebras

Let $k$ be a field and $F = k\langle x_1, \ldots, x_d \rangle$ the free $k$-algebra on $x_1, \ldots, x_d$. The elements of $F$ are called *polynomials* and a $k$-algebra $A$ is said to satisfy the *polynomial identity*

$$p(x_1, \ldots, x_d) = 0, \tag{1}$$

if $p$ is an element of $F$ which vanishes for all values of the $x$'s in $A$. If $A$ satisfies a *non-trivial* polynomial identity, i.e. an identity (1) where $p$ is not the zero polynomial, then $A$ is called a *PI-algebra*. Many of the results proved for Noetherian rings have their counterparts for PI-algebras. It will simplify matters to assume a field of coefficients, but it is possible to consider more general coefficient rings, cf. Procesi [1973] or Rowen [1980].

*Examples*

1. Every commutative $k$-algebra satisfies the identity $xy - yx = 0$ and so is a PI-algebra.
2. Every Boolean ring satisfies the identity $x^2 - x = 0$.
3. Every finite-dimensional algebra satisfies an identity. If $\dim A < n$, then $A$ satisfies the identity

$$S_n(x_1, \ldots, x_n) = \sum_\sigma \operatorname{sgn}(\sigma) x_{1\sigma} x_{2\sigma} \cdots x_{n\sigma} = 0, \tag{2}$$

where $\sigma$ runs over all permutations of $1, \ldots, n$ and $\operatorname{sgn}(\sigma)$ is 1 or $-1$ according as $\sigma$ is even or odd. $S_n$ is called the *standard polynomial* and (2) *standard identity* of degree $n$. For any $a_1, \ldots, a_n \in A$ it is clear that $S_n(a_1, \ldots, a_n) = 0$ if two of the $a$'s coincide. Now take a $k$-basis $e_1, \ldots, e_r$ ($r = \dim A < n$) of $A$; then for any

$a_1, \ldots, a_n \in A$, $S_n(a_1, \ldots, a_n)$ can by linearity be written as a linear combination of terms $S_n(e_{i_1}, \ldots, e_{i_n})$. Since $r < n$, at least two of the $e$'s must coincide, so all terms vanish. This shows that $A$ satisfies (2).

4. If $A$ is a commutative $k$-algebra, then $\mathfrak{M}_n(A)$ satisfies the standard identity of degree $n^2 + 1$, for we have an $A$-basis of $n^2$ elements, so the result follows as in 3. In Th. 5.8 below we shall see that $\mathfrak{M}_n(A)$ actually satisfies $S_{2n} = 0$.

5. If every element of $A$ is algebraic over $k$, of degree at most $n$, each element of $A$ satisfies an equation

$$x^n + a_1 x^{n-1} + \cdots + a_n = 0, \qquad \text{where } a_i \in k. \tag{3}$$

If the equation for some element of $A$ has degree less than $n$, we can bring it to the form (3) by multiplying by a power of $x$. Writing $[x, y] = xy - yx$, we obtain from (3),

$$[x^n, y] + a_1[x^{n-1}, y] + \cdots + a_{n-1}[x, y] = 0.$$

Thus the commutators in this expression are linearly dependent and so $A$ satisfies the identity

$$S_n([x, y], [x^2, y], \ldots, [x^n, y]) = 0.$$

We collect some elementary facts about PI-algebras.

PROPOSITION 5.1 *Any subalgebra or homomorphic image of a PI-algebra is again a PI-algebra.*

The proof is immediate. ■

Since the free algebra is not a PI-algebra, we deduce from Prop. 1.8,

COROLLARY 5.2 *Every PI-algebra which is also an integral domain is a left and right Ore domain.* ■

A polynomial $p$ and the corresponding identity $p = 0$ is said to be *multilinear* if it is homogeneous of degree 1 in each variable. In a PI-algebra we can always find multilinear identities:

PROPOSITION 5.3 *Any algebra $A$ satisfying a polynomial identity of degree $n$ also satisfies a polynomial identity of degree $n$ which is multilinear.*

*Proof.* Let $p = p(x_1, \ldots, x_d)$ be a polynomial of degree $n$ which vanishes identically on $A$, and let $r$ be the highest degree to which any variable occurs, say $p$ has degree $r$ in $x_1$. If $r > 1$, replace $p$ by $p(x_1 + x_{d+1}, x_2, \ldots) - p(x_1, x_2, \ldots) - p(x_{d+1}, x_2, \ldots)$; since $(x_1 + x_{d+1})^r - x_1^r - x_{d+1}^r \neq 0$ (even in finite characteristic, because $x_1 x_{d+1} \neq x_{d+1} x_1$), we get a polynomial in which $x_1, x_{d+1}$ occur, but of

degree less than $r$ and the degree of the other variables is not raised. By a double induction, on the highest degree $r$ and on the number of variables occurring to this degree, we reduce $p$ to a polynomial of degree 1 in each variable. The process preserves the total degree, so we get a polynomial $q$ with a term $ax_1x_2\cdots x_n$ say. Now replace $q$ by $q(\ldots,x_i,\ldots)-q(\ldots,0,\ldots)$ $(1\leqslant i\leqslant n)$ to get rid of terms not involving $x_i$. ■

A multilinear identity has the advantage that to verify it we need only check the elements of a basis. Moreover, such identities are preserved under extensions:

COROLLARY 5.4 *Let $R$ be a $k$-algebra with centre $C$. If $R$ contains a subalgebra $A$ which is a PI-algebra such that $R=AC$, then $R$ is a PI-algebra.*

For by Prop. 5.3, $A$ satisfies a multilinear identity $p(x_1,\ldots,x_n)=0$. Let $\{u_\lambda\}$ be a $k$-basis for $R$ and put $a_i=\sum\alpha_{i\lambda}u_\lambda$, where $\alpha_{i\lambda}\in C$. Then

$$p(a_1,\ldots,a_n)=\sum\alpha_{1\lambda_1}\cdots\alpha_{n\lambda_n}p(u_{\lambda_1},\ldots,u_{\lambda_n})=0,$$

by multilinearity, thus $p$ vanishes on $R$. ■

To illustrate the proposition, suppose we have an algebra $A$ satisfying the identity

$$x^2=0. \tag{4}$$

Here we do not restrict our algebra to have a unit element (for otherwise it would have to be trivial, by (4)). Then $A$ also satisfies $(x+y)^2-x^2-y^2=0$, i.e.

$$xy+yx=0, \tag{5}$$

and this is multilinear. In characteristic not equal to 2 the identities (4) and (5) are equivalent, for we can get back from (5) to (4) by putting $y=x$, which gives $2x^2=0$.

In special cases we can make the assertion of the Cor. more precise.

PROPOSITION 5.5 *Let $A$ be a finite-dimensional algebra over an infinite field $k$. Then any polynomial identity $p(x_1,\ldots,x_d)=0$ for $A$ also holds in $A_E$, for any extension field $E$ of $k$.*

*Proof.* Let $A$ have the $k$-basis $e_1,\ldots,e_n$, take a field $F$ obtained by adjoining $dn$ independent indeterminates $t_{ij}(i=1,\ldots,d,j=1,\ldots,n)$ to $k$ and put $x_i=\sum t_{ij}e_j$. Then in $A_F$ we have

$$p(x_1,\ldots,x_d)=\sum f_v(t_{ij})e_v, \tag{6}$$

where the $f_v$ are polynomials in the $t_{ij}$ with coefficients in $k$. By hypothesis, $f_v$ vanishes identically on $k$: if $a_i=\sum\alpha_{ij}e_j(\alpha_{ij}\in k)$, then $\sum f_v(\alpha_{ij})e_v=p(a_1,\ldots,a_d)=0$;

hence $f_v(\alpha_{ij}) = 0$ for all $\alpha_{ij}$ in $k$. Since $k$ is infinite, $f_v$ vanishes identically, and by (6), $p = 0$ in $A_E$ for any extension field $E$ of $k$. ■

In the opposite direction from Prop. 5.3 one can show that every PI-algebra satisfies an identity in two variables. To prove this fact we shall need the result of Ex. (9) of Ch. 2, which asserts that in the free algebra $k\langle x, y\rangle$ the elements $z_n = xy^n$ are free. This follows because in any equation $f(z_1, \ldots, z_n) = 0$ we can equate homogeneous components, and if $\sum u_i z_i = 0$, then $\sum u_i x y^i = 0$, and it follows that $u_i = 0$ so we get the assertion by an induction on the degree.

PROPOSITION 5.6 *Every PI-algebra satisfies an identity in two variables.*

For if $p(x_1, \ldots, x_d) = 0$ is a non-trivial identity in $A$, then $p(xy, xy^2, \ldots, xy^d) = 0$ and as we have just seen, this is non-trivial. ■

One of the main results in PI-theory, Kaplansky's theorem, asserts that a primitive PI-algebra is finite-dimensional; this will be proved in Ch. 10, where primitive rings are discussed. For the moment we shall show that the $n \times n$ matrix ring satisfies the standard identity $S_{2n}$ (Amitsur–Levitzki theorem); the proof requires an elementary property of traces:

LEMMA 5.7 *Let $K$ be a commutative* **Q***-algebra and $A \in K_n$. If $\operatorname{tr}(A^n) = 0$ for $r = 1, \ldots, n$, then $A^n = 0$.*

*Proof.* Suppose first that $K$ is an algebraically closed field (necessarily of characteristic 0) and let $A$ be a matrix over $K$, with eigenvalues $\lambda_1, \ldots, \lambda_n$. The characteristic polynomial of $A$ is

$$\det(\lambda I - A) = x^n + c_1 x^{n-1} + \cdots + c_n, \tag{7}$$

where the $c_i$ are (except for sign) the elementary symmetric functions of the $\lambda$'s. Since we are in characteristic 0, we can express $c_1, \ldots, c_n$ as polynomials in the power sums of $\lambda$'s, $s_r = \sum \lambda_v^r = \operatorname{tr}(A^r)$, $r = 1, \ldots, n$ (Newton's formulae, cf. A.1, **6.9**):

$$c_i = f_i(s_1, \ldots, s_i), \qquad \text{where } s_r = \operatorname{tr}(A^r), \tag{8}$$

where $f_i$ is of weight $i$ and with rational coefficients.

Now let $K$ be as stated in the lemma, and $A$ the given matrix. Its characteristic polynomial is given by (7), where the $c_i$ are still given by (8) in terms of the $s_i$ (since these equations are identities in the entries of $A$). By hypothesis, $\operatorname{tr}(A^r) = 0$ for $r = 1, \ldots, n$, hence (7) reduces to $x^n$, and the Cayley–Hamilton theorem (which holds for any commutative ring, cf. Th. 3 of **11.1** of A.1) shows that $A^n = 0$. ■

THEOREM 5.8 (Amitsur–Levitzki 1950) *Let $K$ be any commutative ring and*

$A_1, \ldots, A_{2n} \in K_n$. *Then*

$$S_{2n}(A_1, \ldots, A_{2n}) = 0. \tag{9}$$

*Thus* $\mathfrak{M}_n(K)$ *satisfies a polynomial identity of degree* $2n$.

*Proof* (Rosset) Suppose first that $K$ is a $\mathbf{Q}$-algebra and let $E$ be the exterior algebra on the free $K$-module of rank $2n$, with basis $u_1, \ldots, u_{2n}$ say. In $\mathfrak{M}_n(E)$ consider the matrix

$$A = \sum_1^{2n} A_i u_i.$$

For any $v = 1, 2, \ldots$ we have

$$A^v = \sum S_v(A_{i_1}, \ldots, A_{i_v}) u_{i_1} \wedge \cdots \wedge u_{i_v}. \tag{10}$$

In particular, $A^{2n} = S_{2n}(A_1, \ldots, A_{2n}) u_1 \wedge \cdots \wedge u_{2n}$, so we have to prove that $A^{2n} = 0$. Let $E_0$ be the subalgebra of $E$ of terms of even degree; then $E_0$ is commutative and by (10), $A^2 \in \mathfrak{M}_n(E_0)$, so we need only show that $\operatorname{tr}(A^{2r}) = 0$, for $r = 1, \ldots, n$, by Lemma 5.7. By (10) this will follow if we show

$$\operatorname{tr}(S_m(A_1, \ldots, A_m)) = 0, \qquad \text{where } m \text{ is even.} \tag{11}$$

Let $S$ be the symmetric group on $1, \ldots, m$ and $T$ the stabilizer of 1 in $S$; then a transversal of $T$ in $S$ is $1, \tau, \ldots, \tau^{m-1}$, where $\tau = (1, 2 \cdots m)$. Every element of $S$ is uniquely expressible as $\tau^i \sigma$, where $0 \leqslant i < m$ and $\sigma \in T$; for $\tau^i$ brings 1 to the right place and $\sigma$ then permutes $2, \ldots, m$ as needed. Hence we can write

$$S_m(A_1, \ldots, A_m) = \sum \operatorname{sgn}(\tau^i \sigma) A_{1\tau^i\sigma} \cdots A_{m\tau^i\sigma}.$$

Now $\tau$ has the effect of permuting the factors cyclically, but the trace is unaffected by cyclic permutations, hence

$$\operatorname{tr}(S_m(A_1, \ldots, A_m)) = \sum_i \operatorname{sgn}(\tau^i) \left\{ \sum_\sigma \operatorname{sgn}(\sigma) \operatorname{tr}(A_{1\sigma} A_{2\sigma} \cdots A_{m\sigma}) \right\}.$$

Here the second sum is independent of $i$, and $\sum \operatorname{sgn}(\tau^i) = 0$, because $m$ is even, so (11) follows, and this proves (9) when $K$ is a $\mathbf{Q}$-algebra. Therefore it holds for a polynomial ring over $\mathbf{Z}$ (which can be embedded in a $\mathbf{Q}$-algebra), and hence for any commutative ring (which is a homomorphic image of a polynomial ring). ■

We note that the bound $2n$ in this result is best possible:

LEMMA 5.9 (Staircase lemma) *Let* $A$ *be a* $k$*-algebra with* $1 \neq 0$. *Then* $\mathfrak{M}_n(A)$ *satisfies no polynomial identity of degree less than* $2n$.

*Proof.* If $A_n$ satisfies an identity of degree $r < 2n$, then it also satisfies a

multilinear identity of degree $r$. Let this be $p = 0$, where each term of $p$ consists of $x_1, \ldots, x_r$ in some order. Thus $p$ has the form

$$p = \alpha x_1 x_2 \cdots x_r + p', \tag{12}$$

where $\alpha \in k^\times$ and $p'$ is the sum of products of the $x$'s in other orders than that shown. Now the matrix units in $A_n$ satisfy

$$e_{11}e_{12}e_{22}e_{23}\cdots e_{n-1n}e_{nn} = e_{1n} \neq 0,$$

while their product in any other order is 0, and this applies even if we only take the first $r$, where $r > 1$. Hence if we put $x_1 = e_{11}$, $x_2 = e_{12}$, $x_3 = e_{22}, \ldots$, then the first term in (12) is $\alpha e_{1i}$ for some $i$, while all other terms vanish, so $p$ does not vanish on $A_n$ and we have reached a contradiction. ■

## Exercises

(1) Show that a polynomial which vanishes identically on a non-trivial algebra must have zero constant term.

(2) Let $R$ be a prime PI-algebra, satisfying an identity of degree $d$. Show that the left (or right) uniform rank of $R$ is $< d$.

(3) Show that if a $\mathbf{Q}$-algebra with 1 satisfies the standard identity $S_{2n+1} = 0$, then it also satisfies $S_{2n} = 0$.

(4) Show that $S_n([x, y], [x^2, y], \ldots, [x^n, y]) = 0$ holds in $\mathfrak{M}_n(k)$, but not in $\mathfrak{M}_{n+1}(k)$, if $k$ is infinite.

(5) Show that $S_{n+1}(x_1, \ldots, x_{n+1}) = \sum(-1)^{i+1} x_i S_n(x_1, \ldots, \hat{x}_i, \ldots x_{n+1})$. Deduce that $S_n = 0$ implies that $S_m = 0$ for all $m > n$.

(6) Let $R$ be a central simple algebra of degree $n$ over an infinite field as centre. Show that $R$ satisfies $S_{2n} = 0$ but no identity of degree $< 2n$.

(7) Let $R, S$ be such that $\mathfrak{M}_n(R) \cong \mathfrak{M}_n(S)$ and $R$ is commutative. Show that $S$ is also commutative and deduce that $R \cong S$. (Hint: apply the standard identity with suitable arguments including $ae_{11}, be_{11}$.)

## 9.6 Varieties of PI-algebras and Regev's theorem

Let $I$ be any subset of the free algebra $F = k\langle X \rangle$ and denote by $V(I)$ the collection of all $k$-algebras on which all the members of $I$ vanish identically. It is clear that $V(I)$ is a variety; in fact the countably generated algebras in $V(I)$ are just the homomorphic images of $F/\mathfrak{t}$, where $\mathfrak{t}$ is the ideal of $F$ generated by all elements obtained by substituting elements of $F$ for the variables in members of $I$. To obtain another description of $F/\mathfrak{t}$ we need some definitions.

Let $A$ be a $k$-algebra and $Y$ a generating set of $A$. The algebra $A$ is said to be *relatively free* on $Y$ if every mapping $Y \to A$ extends to an endomorphism of $A$ (necessarily unique). For example, the free algebra $F$ is relatively free on $X$, since as we know, every mapping from $X$ to any algebra $A$ extends to a homomorphism $F \to A$, and we need only take the special case $A = F$.

An ideal t of the free algebra $F = k\langle X \rangle$ is said to be *fully invariant* or a *T-ideal* if it admits all endomorphisms of $F$. By the freeness of $F$ this means that if an element $f(x) \in F$ belongs to t, then $f(a) \in$ t for all possible ways of replacing $x_i \in X$ by $a_i \in F$. Given any subset $I$ of $F$, the ideal generated by all elements obtained from $I$ by replacing the $x_i \in X$ by $a_i \in F$ in all possible ways is the least $T$-ideal containing $I$; this is also called the $T$-ideal *generated* by $I$. Now the ideals defining relatively free algebras are precisely the $T$-ideals.

PROPOSITION 6.1 *For any k-algebra the following are equivalent:*

(a) *$A$ is relatively free on a set $Y$,*

(b) *$A \cong F/\mathfrak{t}$, where $F$ is the free k-algebra on a set $X$ equipotent with $Y$ and* t *is a T-ideal.*

*Proof.* Any $k$-algebra $A$ with generating set $Y$ can be written as a homomorphic image of a free algebra $F$ on a set $X$ equipotent with $Y$:

$$A \cong F/\mathfrak{t}, \tag{1}$$

where t is the ideal of relations in $A$. Suppose that (a) holds and that $\lambda: F \to A$ is a surjective homomorphism with kernel t, whose restriction to $X$ defines a bijection with $Y$. Given any mapping $\varphi: X \to F$, there is a unique endomorphism of $F$ agreeing with $\varphi$ on $X$, which we may again denote by $\varphi$. We have to show that $\varphi$ maps t into itself. Now by (a) the mapping $\lambda^{-1}\varphi\lambda: Y \to A$ extends to an endomorphism $g$ of $A$; thus $x\varphi\lambda = x\lambda g$ for all $x \in X$, hence $\varphi\lambda = \lambda g$ holds on all of $F$. If $p \in$ t, then $p\lambda = 0$, hence $p\varphi\lambda = p\lambda g = 0$ and it follows that $p\varphi \in \ker \lambda =$ t; thus (b) is proved.

Conversely, assume (b) and let a mapping $g: Y \to A$ be given. We define $\varphi: X \to F$ as follows: given $x \in X$, choose an element $u$ of $F$ such that $u\lambda = x\lambda g$. We put $u = x\varphi$; the mapping $\varphi$ extends to an endomorphism of $F$, which will again be denoted by $\varphi$. Since t is a $T$-ideal, if $p \in$ t, then $p\varphi \in$ t; thus $\varphi$ can be factored by $\lambda$ to give an endomorphism $h$ of $A$ such that $\varphi = \lambda h$. For any $x \in X$ we have $x\varphi\lambda = x\lambda h = x\lambda g$, hence $h$ is an endomorphism of $A$ which agrees with $g$ on $Y$, and this shows $A$ to be relatively free. ■

Our aim in what follows is to prove that the tensor product of two PI-algebras is again a PI-algebra. This is Regev's theorem; the proof given here is due to Latyshev. Some preparation is necessary.

Consider a permutation $\sigma$ of $1, 2, \ldots, n$. We use $\sigma$ to define a partial ordering on the set $\{1, \ldots, n\}$ by writing $i \prec j$ whenever $i < j$ and $i\sigma < j\sigma$. Let us recall

that an *antichain* in a partially ordered set is a subset of mutually incomparable elements, and the *width* of the set is the maximum number of elements in an antichain. For example, for our permutation $\sigma$ an antichain of $d$ elements is a set of numbers $i_1 < i_2 < \cdots < i_d$ such that $i_1\sigma > i_2\sigma > \cdots > i_d\sigma$. We also recall from A.2, Th. 1.4.1:

Dilworth's theorem. *In any finite partially ordered set $S$, the minimum number of disjoint chains into which $S$ can be decomposed is the width of $S$.* ■

Given a permutation $\sigma$, suppose that the corresponding partially ordered set can be decomposed into $d$ chains. Then to specify $\sigma$ we need only give the $d$ chains and their images under $\sigma$. For example, the permutation $\left(\begin{smallmatrix}1&2&3&4&5&6&7&8\\3&5&8&1&4&7&6&2\end{smallmatrix}\right)$ defines a partially ordered set of width 4, and it may be expressed as the union of the chains $\{1,2,3\}$, $\{4,5,6\}$, $\{7\}$, $\{8\}$ with images $\{3,5,8\}$, $\{1,4,7\}$, $\{6\}$, $\{2\}$. Dilworth's theorem allows us to estimate the number of permutations of a given width.

THEOREM 6.2 *The set of permutations of $1,2,\ldots,n$ for which any set of $d$ numbers $(2 \leqslant d \leqslant n)$ contains at least one pair in their natural order is at most $(d-1)^{2n}$.*

*Proof.* Let $\sigma$ be such a permutation and consider the partial ordering defined by $\sigma$ on $\{1,\ldots,n\}$. The hypothesis states that no antichain has $d$ elements, so the set has width less than $d$, and by Dilworth's theorem it can be written as a disjoint union of at most $d-1$ chains. Let us number these chains from 1 to $\delta$, where $\delta < d$. To specify $\sigma$ we have to give the distribution of $1,2,\ldots,n$ and their images under $\sigma$ over these $\delta$ chains. This can be done by defining two mappings from $\{1,\ldots,n\}$ to $\{1,\ldots,\delta\}$. For each mapping there are $\delta^n$ choices, hence in all there are $\delta^{2n}$ choices and so the number of permutations satisfying the given condition is at most $(d-1)^{2n}$. ■

Let $F = k\langle X\rangle$ be the free $k$-algebra on $X = \{x_1, x_2, \ldots\}$ and denote by $L_n$ the subspace of multilinear forms in $x_1,\ldots,x_n$; $L_n$ is spanned by the monomials

$$x_\sigma = x_{1\sigma}x_{2\sigma}\cdots x_{n\sigma}, \tag{2}$$

hence its dimension over $k$ is $n!$. We fix an integer $d$, $2 \leqslant d \leqslant n$, and call a monomial (2) *good* if the partial ordering defined by $\sigma$ has width $< d$; by Prop. 6.2 the number of good monomials (2) does not exceed $(d-1)^{2n}$. We shall use that result to bound the dimension of a relatively free algebra; here it will be convenient to restrict attention to multilinear elements.

PROPOSITION 6.3 *Let* t *be a T-ideal in the free algebra $F = k\langle X\rangle$, and put $\mathrm{t}_n = \mathrm{t} \cap L_n$, where $L_n$ is the space of multilinear forms of degree $n$ as above. If* t

*contains a polynomial of degree d, where* $2 \leqslant d \leqslant n$, *then*

$$[L_n/\mathrm{t}_n : k] \leqslant (d-1)^{2n}. \tag{3}$$

*Proof.* Let us take the monomial basis in $L_n$ with the lexicographic ordering. By linearization it follows that $F/\mathrm{t}$ satisfies a $d$-linear identity

$$y_1 y_2 \cdots y_d = \sum_{\sigma \neq 1} \alpha_\sigma y_\sigma \qquad \text{where } \alpha_\sigma \in k. \tag{4}$$

For the proof it will be enough to show that $L_n$ is spanned (mod $\mathrm{t}_n$) by the good monomials. Suppose that $u$ is a monomial which is not good. Then we have a factorization

$$u = p(x_{\alpha_1} \cdots x_{\beta_1})(x_{\alpha_2} \cdots x_{\beta_2}) \cdots (x_{\alpha_{d-1}} \cdots x_{\beta_{d-1}}) x_{\alpha_d} q,$$

where $\alpha_1 > \alpha_2 > \cdots > \alpha_d$. If in (4) we put $y_i = x_{\alpha_i} \cdots x_{\beta_i}, y_d = x_{\alpha_d}$, we obtain

$$u \equiv \sum \lambda_j v_j \qquad (\mathrm{mod}\, \mathrm{t}_n),$$

where $v_j$ comes before $u$ in the lexicographic ordering. Repeating the process if necessary we can after a finite number of steps reduce $u$ to a linear combination of good monomials. Now the conclusion follows by Prop. 6.2. ■

In Prop. 6.3 we have obtained in (3) a bound for $[L_n/\mathrm{t}_n : k]$ of the form $f(d)^n$, and so we have $f(d)^{2n} < n! = [L_n : k]$ for large enough $n$. This estimate now allows us to achieve our aim:

THEOREM 6.4 (Regev's theorem 1972) *The tensor product of two PI-algebras over a field k is again a PI-algebra.*

*Proof* (Latyshev). Let $A, B$ be PI-algebras over $k$; clearly there is a polynomial identity holding in both $A$ and $B$. Let $\mathrm{t}$ be the $T$-ideal of all identities holding in both $A$ and $B$ and suppose that $\mathrm{t}$ contains an identity of degree $d \geqslant 2$. From Prop. 6.3 we know that $[L_n/\mathrm{t}_n : k] \leqslant (d-1)^{2n}$ for all $n \geqslant d$. Let us fix $n \geqslant d$ and take a monomial basis $m_1(x), \ldots, m_v(x)$ of $L_n/\mathrm{t}_n$, where $v = [L_n/\mathrm{t}_n : k]$.

For any permutation $\sigma$ of $1, 2, \ldots, n$ we have

$$x_\sigma = x_{1\sigma} x_{2\sigma} \cdots x_{n\sigma} \equiv \sum_1^v \lambda_i(\sigma) m_i(x) \qquad (\mathrm{mod}\, \mathrm{t}_n). \tag{5}$$

In particular, this relation holds in $A$ and $B$. Now consider a multilinear polynomial

$$f(x) = \sum \gamma_\sigma x_{1\sigma} \cdots x_{n\sigma}, \tag{6}$$

with coefficients $\gamma_\sigma$ to be determined, in such a way that $f = 0$ is an identity for $A \otimes B$. By multilinearity we need only check $f$ on a spanning set. Given

$a_1, \ldots, a_n \in A$, $b_1, \ldots, b_n \in B$, we have

$$f(a_1 \otimes b_1, \ldots, a_n \otimes b_n) = \sum \gamma_\sigma a_{1\sigma} \otimes b_{1\sigma} \cdots a_{n\sigma} \otimes b_{n\sigma}$$

$$= \sum_{i,j=1}^{v} \left( \sum_\sigma \lambda_i(\sigma)\lambda_j(\sigma)\gamma_\sigma \right) m_i(a) \otimes m_j(b),$$

by (6) and (5). To ensure that $f = 0$ on $A \otimes B$, we have to solve the equations

$$\sum_\sigma \lambda_i(\sigma)\lambda_j(\sigma)\gamma_\sigma = 0. \tag{7}$$

These are $v^2$ homogeneous linear equations in the $n!$ indeterminates $\gamma_\sigma$ and they have a non-trivial solution when $n! > v^2$. By Prop. 6.3, $v \leqslant (d-1)^{2n}$; now $d$ is given and for large enough $n$ we have

$$v^2 \leqslant (d-1)^{4n} < n!.$$

For such $n$ the equations (7) have a non-trivial solution $\gamma_\sigma^0$ and then $f = \sum \gamma_\sigma^0 x_\sigma$ is the required polynomial identity for $A \otimes B$. ■

### Exercises

(1) Show that $A$ is universal for homomorphisms into a given family $\mathscr{C}$ of algebras iff $A$ is relatively free.

(2) Show that if $A, B$ satisfy the same polynomial identity of degree $d$, then $A \otimes B$ satisfies an identity of degree at most $(d-1)e$. (Hint: use Stirling's formula $(n-1)! \sim n^n e^{-n}\sqrt{2\pi/n}$.)

(3) Show that an algebra satisfying $x^2 = 0$ also satisfies $xyz = 0$.

(4) Show that if $A, B$ satisfy $x^2 = 0$, then $A \otimes B$ satisfies $x^3 = 0$. Find an identity for $A \otimes B$ when $A$ satisfies $x^2 = 0$ and $B$ satisfies $x^3 = 0$.

## 9.7 Generic matrix rings and central polynomials

As we have seen in **9.6**, a relatively free algebra may be characterized by the fact that it is universal for homomorphisms to a given class of algebras. We now define the *generic matrix ring* of degree $n$ in $x_1, \ldots, x_d$,

$$F_{(n)} = k\langle x_1, \ldots, x_d \rangle_{(n)}$$

as the $k$-algebra on $x_1, \ldots, x_d$ which is universal for homomorphisms into $n \times n$ matrix rings over commutative rings. The elements of $F_{(n)}$ may themselves be thought of as matrices, so that $F_{(n)}$ is a ring of $n \times n$ matrices, generated by $x_1, \ldots, x_d$ and every mapping $x_i \mapsto a_i \in \mathfrak{M}_n(A)$, where $A$ is commutative, can be extended to a unique $k$-algebra homomorphism.

An explicit construction of $F_{(n)}$ is obtained as follows: we adjoin $dn^2$ commuting indeterminates $t_{ij}^{\mu}(\mu = 1,\ldots,d, i,j = 1,\ldots,n)$ to $k$ and in the $n \times n$ matrix ring $\mathfrak{M}_n(k[t_{ij}^{\mu}])$ consider the subalgebra generated by the matrices $x_\mu = (t_{ij}^{\mu})$. Since the $t_{ij}^{\mu}$ are commuting indeterminates, the universal property is easily verified. When $d = 1, k\langle x_1\rangle_{(n)}$ is just the polynomial ring in one variable, but for $d > 1$ (and $n > 1$) $k\langle x_1,\ldots,x_d\rangle_{(n)}$ is non-commutative, since $n \times n$ matrices over a non-trivial ring do not all commute. On the other hand, $F_{(n)}$ satisfies certain identities; since it is a subring of $\mathfrak{M}_n(A)$, where $A$ is commutative, it satisfies e.g. the standard identity $S_{2n} = 0$. Thus we may think of $F_{(n)}$ as the $k$-algebra generated by $x_1,\ldots,x_d$ subject to all the identities holding between $n \times n$ matrices. This is expressed more precisely in

PROPOSITION 7.1 *Let $F = k\langle t_1,\ldots,t_d\rangle$ be the free $k$-algebra, $F_{(n)} = k\langle x_1,\ldots,x_d\rangle_{(n)}$ the generic matrix ring of degree $n$ and $v: F \to F_{(n)}$ the $k$-algebra homomorphism in which $t_i \mapsto x_i$. Then $p \in F$ vanishes identically on every $n \times n$ matrix ring over a commutative $k$-algebra if and only if $pv = 0$.*

*Proof.* If $p$ vanishes on every $n \times n$ matrix ring, then in particular, $pv = 0$. Conversely, if $pv = 0$, then since every homomorphism $\varphi: F \to \mathfrak{M}_n(A)$ ($A$ commutative) can be factored by $v$, say $\varphi = v\varphi'$, we have $p\varphi = pv\varphi' = 0$. ■

Thus if $p \in F$ and we want to check whether $p = 0$ holds in all matrix rings, we need only find its image in $F_{(n)}$. Here it is often convenient to embed the coefficient ring $k[t_{ij}^{\mu}]$ in an algebraically closed field $K$. Then we can transform any matrix over $K$ with distinct eigenvalues to diagonal form. Now the generic matrix $x_\mu = (t_{ij}^{\mu})$ certainly has distinct eigenvalues, since we can specialize it to any other matrix. Hence we can always transform $x_1$ to diagonal form; of course the same applies to $x_2,\ldots,x_d$, but we cannot transform more than one of the $x$'s simultaneously to diagonal form, because they do not commute.

It turns out that $F_{(n)}$ can be embedded in a skew field; this follows from the next result.

PROPOSITION 7.2 *The generic matrix ring $k\langle x_1,\ldots,x_d\rangle_{(n)}$ is a left and right Ore domain.*

*Proof.* We have already seen that $F_{(n)}$ is a PI-algebra; if we can show that it is an integral domain, the desired result will follow by Cor. 5.2.

It remains then to show that $F_{(n)}$ is an integral domain; the essence of the proof will be to show that any polynomial identity holding in $F_{(n)}$ also holds in a certain division algebra. Let $K$ be the field of fractions of the coefficient ring $k[t_{ij}^{\mu}]$ and let $E$ be an extension field of $K$ with a $K$-automorphism $\alpha$ of order $n$, e.g. we may take $E = K(\xi_1,\ldots,\xi_n)$ to be a rational function field and $\alpha$ a cyclic permutation of the $\xi$'s. Let $D = E(z;\alpha)$ be the skew field of fractions

of the skew polynomial ring. As we have seen in Example 6 of **9.3**, this is a central division algebra of degree $n$ over its centre $C$, say. Let $L$ be a splitting field; then

$$D \otimes_C L \cong L_n \cong K_n \otimes_K L. \tag{1}$$

Here the left-hand side contains $D$, while the right-hand side contains $K_n$. Now let $f, g \in F_{(n)}$ and suppose that $fg = 0$. Then $fg$ vanishes identically on $L_n$ and hence (by (1)) on $D$. Since $D$ is a skew field, it follows that for each choice of arguments either $f$ or $g$ vanishes, so if $y$ is a new indeterminate, then

$$f(x_1, \ldots, x_d) y g(x_1, \ldots, x_d) = 0, \tag{2}$$

identically in $D$. By Prop. 5.5, this also holds in $D_L \cong L_n$ and hence in $K_n$, but $K_n$ is simple, hence prime, so either $f = 0$ or $g = 0$, as elements of $F_{(n)}$. This shows $F_{(n)}$ to be an integral domain. ■

From this result it follows that $F_{(n)}$ has a skew field of fractions, called the *generic division algebra* of degree $n$ over $k$. We shall see below (in Cor. 7.5) that its degree over its centre is $n$.

These concepts have been used by Amitsur (1972) to prove that not every division algebra is a crossed product. It can be shown that if the generic division algebra of degree $n$ were a crossed product, with Galois group $\Gamma$, then every division algebra of degree $n$ would be a crossed product with group $\Gamma$. Now Amitsur constructs two division algebras of degree $n$ (for certain $n$, and in characteristic 0) which cannot be expressed as crossed products with the same Galois group; it follows that the generic division algebra cannot be a crossed product (cf. Rowen [1988]).

In studying the centre of a PI-algebra it would be useful to have 'centre-valued' or 'central' polynomials, i.e. polynomials $p$ which when evaluated, always yield elements of the centre. By a *central polynomial* for $n \times n$ matrices one understands a polynomial $p \in k\langle x_1, \ldots, x_d \rangle$ which when evaluated in $\mathfrak{M}_n(A)$, where $A$ is a commutative $k$-algebra, takes values in the centre of $\mathfrak{M}_n(A)$. Of course we are interested only in non-constant polynomials, i.e. polynomials taking at least two values.

As an example consider $2 \times 2$ matrices over $\mathbf{R}$. A commutator of $2 \times 2$ matrices has the form $\alpha = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$, for its trace must be 0, and hence

$$\alpha^2 = \begin{pmatrix} a^2 + bc & 0 \\ 0 & a^2 + bc \end{pmatrix}.$$

This is a scalar, hence $(xy - yx)^2$ is a central polynomial for $2 \times 2$ matrices. This was essentially the only non-constant central polynmial known for many years, but a large family of central polynomials for all values of $n$ was discovered in

1972 by E. Formanek. In 1974, Razmyslov discovered multilinear central polynomials, for $n \times n$ matrices, and we shall now describe his construction. We preface the main theorem by two remarks.

Let $k$ be a field; the matrix ring $k_n$ may be considered as an $n^2$-dimensional vector space with the bilinear form $\mathrm{tr}(xy)$. The usual matrix basis $\{e_{ij}\}$ has the dual basis $\{e_{ji}\}$, for we clearly have

$$\mathrm{tr}(e_{ij}e_{kl}) = \delta_{jk}\delta_{il}.$$

For simplicity we shall index the $e_{ij}$ by a single suffix $\lambda: e_\lambda$ $(\lambda = 1, \ldots, n^2)$ and denote the dual basis by $e_\lambda^*$. We observe that for any matrix $a = \sum a_{ij}e_{ij}$ we have

$$\sum e_\lambda a e_\lambda^* = \sum e_{ij}a_{rs}e_{rs}e_{ji} = \sum a_{jj}e_{ii} = \mathrm{tr}(a)\cdot 1.$$

If $\{f_\lambda\}, \{f_\lambda^*\}$ is another pair of dual bases for $k_n$, say $f_\lambda = \sum c_{\lambda\mu}e_\mu$, $e_\lambda^* = \sum f_\mu^* c_{\mu\lambda}$, then $\sum f_\lambda a f_\lambda^* = \sum c_{\lambda\mu}e_\mu a f_\lambda^* = \sum e_\mu a e_\mu^*$, and this proves the formula

$$\sum e_\lambda a e_\lambda^* = \mathrm{tr}(a)\cdot 1 \qquad \text{for any dual bases } \{e_\lambda\}, \{e_\lambda^*\} \text{ of } k_n. \tag{3}$$

Secondly, let $F = k\langle x_1, \ldots, x_d\rangle$ and consider the subspace $\Phi$ of elements that are homogeneous of degree 1 in $x_1$. These elements can be written in the form $\sum a_i x_1 b_i$, where $a_i, b_i \in k\langle x_2, \ldots, x_d\rangle$, and the space $\Phi$ admits the linear mapping

$$\sum a_i x_1 b_i \mapsto \sum b_i x_1 a_i,$$

because $\sum a_i \otimes b_i = 0 \Leftrightarrow \sum b_i \otimes a_i = 0$. In general rings there is no reason for this to hold, but it does hold when $x$ is a generic matrix:

LEMMA 7.3 *Let $R$ be a $k$-algebra, let $x_{ij}$ be $n^2$ commuting indeterminates over $k$ and write $x = (x_{ij})$. Then the $R$-bimodule generated by $x$ in $\mathfrak{M}_n(R \otimes k[x_{ij}])$ admits the $k$-linear mapping*

$$A_x: \sum a_i x b_i \mapsto \sum b_i x a_i. \tag{4}$$

*Proof.* The subspace of matrices linear in the $x_{ij}$ has the $R$-basis $x_{ij}e_{rs}$. Let us define $(x_{ij}e_{rs})A_x = x_{sr}e_{ji}$; then (4) holds for $a_1 = e_{ri}$, $b_1 = e_{js}$, for then $a_1 x b_1 = x_{ij}e_{rs}$, $b_1 x a_1 = x_{sr}e_{ji}$. Hence it holds generally, by linearity. ■

This transformation $A_x$ is called the *Razmyslov transposition*. We observe that if $u$ is linear homogeneous in $x$, then

$$\mathrm{tr}(y(uA_x|_{x\to 1})) = \mathrm{tr}(u|_{x\to y}), \tag{5}$$

where $x \to y$ indicates that $x$ is to be replaced by $y$. For $u = pxq$ the equation reduces to $\mathrm{tr}(yqp) = \mathrm{tr}(pyq)$, which holds by the cyclic symmetry of the trace; hence it holds generally by linearity.

We can now state the main result; instead of $A_{x_i}$ we write $A_i$.

THEOREM 7.4 (Razmyslov) *Let*

$$C = C(x_1, \ldots, x_{n^2}; y_0, \ldots, y_{n^2}) = \sum_\sigma \operatorname{sgn}(\sigma) y_0 x_{1\sigma} y_1 \cdots y_{n^2-1} x_{n^2\sigma} y_{n^2},$$

*where $\sigma$ ranges over all permutations of $1, \ldots, n^2$. Then*

$$D = D(x_1, \ldots, x_{n^2}; y_0, \ldots, y_{n^2}; z) = \sum x_i z(CA_i|_{x_i \to 1})$$

*is a non-constant central polynomial on $\mathfrak{M}_n(k)$; more precisely,*

$$D(x; y; z) = \operatorname{tr}(z) \operatorname{tr}(C) \cdot 1. \tag{6}$$

*Proof.* Let $\{e_\lambda\}$ be a basis for $k_n$ and denote by $E$ the field of fractions of the polynomial ring in the commuting indeterminates $x_i^{(\lambda)}, y_i^{(\lambda)} (\lambda = 1, \ldots, n^2, i = 0, 1, \ldots, n^2)$. We shall prove the theorem by taking $x_i = \sum x_i^{(\lambda)} e_\lambda$, $y_i = \sum y_i^{(\lambda)} e_\lambda$ and establishing (6) in $E_n$.

In the first place we note that $\operatorname{tr}(C) \neq 0$. For if we take the $x_i$ to be the $e_{rs}$ in some order, we can choose the $y_i$ so that

$$y_0 x_1 y_1 x_2 \cdots x_{n^2} y_{n^2} = e_{11},$$

while all the other terms in $C$ are 0, and with these values $\operatorname{tr}(C) = 1$.

It is clear from the definition that $C$ vanishes when we put $x_i = x_j$, where $i \neq j$. Hence by (5)

$$\operatorname{tr}(x_j(CA_i|_{x_i \to 1})) = \operatorname{tr}(C|_{x_i \to x_j}) = \begin{cases} 0 & \text{if } j \neq i, \\ \operatorname{tr}(C) & \text{if } j = i. \end{cases}$$

It follows that up to a scalar (non-zero!) factor $\{(CA_i|_{x_i \to 1})\}$ is a dual basis for $\{x_i\}$, so by (3),

$$D = \sum x_i z(CA_i|_{x_i \to 1}) = \operatorname{tr}(z) \operatorname{tr}(C) \cdot 1,$$

as we wished to show. ∎

The polynomial $C = C_n$ is called the *Capelli polynomial* and $D = D_n$ is the *Razmyslov polynomial* for $n \times n$ matrices. By Th. 7.4 it is a non-constant central polynomial on $\mathfrak{M}_n(k)$ whose value is relatively easy to calculate, using the formula (6).

As Procesi has observed, any central polynomial $\varphi$ for $n \times n$ matrices, with zero constant term, vanishes identically on $\mathfrak{M}_m(k)$ for $m < n$, for we can regard any $m \times m$ matrix as an $n \times n$ matrix whose last $n - m$ rows and columns are 0. Now the value of $\varphi$ must be central, i.e. a scalar, and this scalar is 0, as we see by looking at the $(n, n)$-entry. More generally, let $A$ be a simple algebra of degree $m$ with centre $k$. If $E$ is a splitting field, then $A_E = A \otimes E \cong E_m$ and since $D_n$ is multilinear, it vanishes on $A$ whenever $m < n$. This proves

COROLLARY 7.5 *The Razmyslov polynomial $D_n$ is central and non-vanishing for any central simple algebra of degree n, while it vanishes identically on central simple algebras of degree less than n.* ■

It is clear that $D_n$ does not vanish on the generic division algebra $F_{(n)}$, whereas $D_{n+1}$ does; this shows $F_{(n)}$ to be of degree $n$. We shall return to this point in **10.5**, when we come to discuss prime PI-algebras.

### Exercises

(1) Show that the exterior algebra $\Lambda(V)$ on any vector space $V$ is a PI-algebra, but if $V$ is infinite-dimensional, $\Lambda(V)$ does not satisfy a standard identity and so cannot be embedded in a matrix ring over a commutative ring.

(2) Show that in a prime PI-algebra every non-zero ideal contains a central regular element.

(3) Show that Cor. 7.5 holds for any central polynomial with zero constant term. (Hint: use Prop. 5.5 and treat the case of a finite ground field separately.)

(4) Show that $[\mathrm{tr}(ax)b]A_x = a \cdot \mathrm{tr}(xb)$, where $A_x$ is the Razmyslov transposition.

(5) (Razmyslov) Show that the commutators $[a,b] = ab - ba$ span an $(n^2-1)$-dimensional subspace of $k_n$ and use this fact to prove that (in the notation of Th. 7.4) $B = C(x, [u_2, v_2], \ldots, [u_{n^2}, v_{n^2}]; y_0, \ldots, y_{n^2}) = \mathrm{tr}(x)p$, where $p$ is a matrix depending on $u, v, y$. Deduce that $BA_x$ is again a central polynomial.

(6) Left $f$ be a polynomial in the entries of a square matrix $A$ over a field. Show that if $f$ is unchanged when $A$ is replaced by $P^{-1}AP$, then $f$ is a symmetric function of the eigenvalues of $A$.

## 9.8 Generalized polynomial identities

In all the polynomial identities considered so far the variables commute with all the elements of the ground field; in other words, our identities were obtained by equating elements of $k\langle X\rangle$ to zero. However, one may wish to consider situations where the variables do not centralize the ground ring; this leads to identities obtained by equating elements of the tensor ring $A_k\langle X\rangle$ to zero.

Let $A$ be any $k$-algebra. By a *generalized polynomial identity* (GPI) one understands a non-zero element of the tensor $A$-ring $A_k\langle X\rangle$ which vanishes under all mappings $X \to A$. Amitsur (1965) has shown that a primitive ring $R$ satisfies a GPI iff $R$ has a non-zero socle and the endomorphism ring $D$ of a simple $R$-module is finite-dimensional over its centre. The main difference from Kaplansky's theorem (as regards the conclusion) is that this time the degree

of the identity does not provide a bound on the dimension. This can be illustrated by the example of $n \times n$ matrix rings over a commutative ring, which satisfies a GPI of degree 2:

$$e_{11}xe_{11}ye_{11} - e_{11}ye_{11}xe_{11} = 0,$$

whereas an ordinary identity has degree at least $2n$, as we saw in the staircase lemma (Lemma 5.9). We shall confine our attention to the special case of Amitsur's theorem where $R$ is a skew field; thus we shall essentially prove that a skew field satisfying a GPI is finite-dimensional over its centre. This then will also provide an independent proof of Kaplansky's theorem for the case of skew fields. We begin with two lemmas.

LEMMA 8.1 *Let $D$ be a skew field, $\Sigma$ a multiplicative subset of $D$ and $K$ its centralizer in $D$. If $a_i, b_i \in D^\times$ $(i = 1, \ldots, n)$ are such that*

$$\sum a_i x b_i = 0 \qquad \text{for all } x \in \Sigma, \tag{1}$$

*then $a_1, \ldots, a_n$ are right linearly dependent over $K$.*

*Proof.* Since $a_1, b_1 \neq 0$, we have $a_1 x b_1 \neq 0$ and it follows that $n > 1$ in any relation (1). Taking $n$ minimal, we may assume that no such relation exists for a proper subfamily of $a_1, \ldots, a_n$. Now any element of $D$ can be written in the form $\sum u_j b_1 v_j$ for $u_j \in \Sigma$, $v_j \in D, j = 1, \ldots, r$, because $\Sigma b_1 D = D$. For $i = 1, \ldots, n$ we define a mapping $\gamma_i : D \to D$ by the rule

$$\gamma_i : \sum u_j b_1 v_j \mapsto \sum u_j b_i v_j, \qquad \text{where } u_j \in \Sigma, v_j \in D. \tag{2}$$

To show that $\gamma_j$ is well-defined we have to verify that

$$\sum u_j b_1 v_j = 0 \Rightarrow \sum u_j b_i v_j = 0 \qquad (i = 2, \ldots, n).$$

Suppose then that $\sum u_j b_1 v_j = 0$. For any $x \in \Sigma$ we have $\sum a_i x u_j b_i = 0$ by (1), hence

$$0 = -\sum_j a_1 x u_j b_1 v_j = \sum_{i=2}^{n} a_i x \left( \sum_j u_j b_i v_j \right).$$

This holds for all $x \in \Sigma$ and is a shorter relation than (1), hence the coefficient of each $a_i$ must vanish, i.e. $\sum u_j b_i v_j = 0$ for $i = 2, \ldots, n$, as we wished to show.

Hence the mapping (2) is well-defined. It is right $D$-linear, i.e. $\gamma_i(zc) = \gamma_i(z)c$ for all $c \in D$. Taking $z = 1$, we find that $\gamma_i(c) = \gamma_i \cdot c$ is left multiplication by an element $\gamma_i$. Moreover, $\gamma_i$ is also left $\Sigma$-linear: $\gamma_i w = w\gamma_i$ for all $w \in \Sigma$, hence $\gamma_i \in K$. By definition, $\gamma_i b_1 = b_i$, hence

$$0 = \sum a_i x b_i = \sum a_i x \gamma_i b_1 = (\sum a_i \gamma_i) x b_1.$$

Since $b_1 \neq 0$, we have $\sum a_i \gamma_i = 0$, and here $\gamma_1 = 1$, so the $a$'s are linearly dependent over $K$. ■

LEMMA 8.2 *Let $D$, $\Sigma$, $K$ be as in Lemma 8.1. Suppose that $a_1,\dots,a_n\in D$ are right linearly independent over $K$ and $b_1,\dots,b_n\in D^\times$ are such that the set $E=\{\sum a_i x b_i \mid x\in\Sigma\}$ is contained in a finite-dimensional left $K$-space. Then there exists $c\in D^\times$ such that $Kc\Sigma$ is finite-dimensional as left $K$-space.*

*Proof.* Let $u_1,\dots,u_r$ be a left $K$-basis for a space containing $E$, so that

$$\sum_1^n a_i x b_i = \sum_1^r \lambda_j(x)u_j \qquad \text{for all } x\in\Sigma \text{ and some } \lambda_j(x)\in K.$$

Here we may take $b_1=1$ by multiplying by $b_1^{-1}$ on the right. We shall use induction on $n$; for $n=1$ we have $a_1x=\sum\lambda_j(x)u_j$, so $u_1,\dots,u_r$ is a $K$-basis for a space containing $Ka_1\Sigma$ and we may take $c=a_1$.

Assume now that $n>1$. For any $y\in\Sigma$ we have

$$\sum_2^n a_i x(b_i y - y b_i) = \sum\lambda_j(x)u_j y - \sum\lambda_j(xy)u_j.$$

If $b_iy\neq yb_i$ for some $y\in\Sigma$ and some $i$, we can apply induction on $n$ to reach the conclusion. Otherwise $b_iy=yb_i$ for all $y\in\Sigma$ and so $b_i\in K$; hence $\sum a_i x b_i=(\sum a_i b_i)x$ and we are reduced to the case $n=1$. ■

We now come to the main result of this section.

THEOREM 8.3 (GPI-theorem) *Let $D$ be a skew field, $\Sigma$ a multiplicative subset of $D$ and $K$ its centralizer in $D$. If for all $c\in D^\times$, $Kc\Sigma$ is infinite-dimensional as left $K$-space, then any non-zero multilinear element of $D_K\langle X\rangle$ has a non-zero value for some choice of values of $X$ in $\Sigma$.*

*Proof.* Let $f$ be a non-zero multilinear polynomial in $D_K\langle X\rangle$ of degree $n$, say, which vanishes on $\Sigma$. We single out the terms in which $x_1$ occurs last and write

$$f=\sum_1^r g_i x_1 b_i + \sum_1^s p_j x_1 q_j, \tag{3}$$

where $b_i\in D$ and $q_j$ has positive degree in the $x$'s. We may again take $b_1=1$ and suppose $f$ chosen so that $r,s$ are minimal. Then for any $y\in\Sigma$,

$$f(x_1,\dots,x_n)y - f(x_1y,x_2,\dots,x_n)=\sum_2^r g_i x_1(b_iy-yb_i)+\sum_1^s p_j x_1(q_jy-yq_j). \tag{4}$$

Since $r$ was chosen minimal in (3), $1,b_2,\dots,b_n$ are linearly independent over $K$, in particular, $b_i\notin K$ for $i>1$, so none of the terms $b_iy-yb_i$ in the first sum can vanish identically for $y\in\Sigma$. Choosing $y=y_0\in\Sigma$, such that $b_2y_0\neq y_0b_2$ we obtain a GPI for $D$ with a smaller value of $r$, unless $r=1$, when the first sum on the

right of (4) is absent. In the latter case consider $q_j y - y q_j$; if this vanishes identically for all $y \in \Sigma$, then $q_j \in K$ for all values of the $x$'s in $\Sigma$. Write $q_j = \sum c_i x_2 d_i$, where the values of $x_3, \ldots, x_n$ in $\Sigma$ are chosen so that $q_j \neq 0$ (which is possible, by induction on $n$). Then the set $\{\sum c_i y d_i | y \in \Sigma\}$ is one-dimensional over $K$, hence $Kc\Sigma$ is finite-dimensional for some $c \in D^\times$, by Lemma 8.2, in contradiction to the hypothesis.

There remains the case where the $q_j y - y q_j$ do not all vanish identically; we show that this leads to a contradiction. In this case the left-hand side of (4), for suitable $y = y_0 \in D$ is a non-zero polynomial $f_1$, again multilinear in $x_1, \ldots, x_n$, with no terms in which $x_1$ is last. Moreover, each term in $f_1$ has the $x$'s in the same order as some term in $f$, so if $x_\nu$ does not come last in any term of $f$, then the same is true of $f_1$. We apply the same reduction to $x_2, \ldots, x_n$ in turn and finally obtain a polynomial $f^*$ in which no $x_\nu$ comes last. This is impossible, so this case cannot occur. ■

If in this theorem we choose $\Sigma = D$, then $K$ is the centre of $D$ and we obtain the skew field case of Amitsur's GPI-theorem:

THEOREM 8.4 (Amitsur) *Let $D$ be a skew field with centre $C$ such that $[D:C] = \infty$. If $f \in D_C\langle X \rangle$ is non-zero, then $f(a) \neq 0$ for some $a \in D^X$.*

For the proof we need only observe that a non-zero polynomial $f$ leads to a non-zero multilinear polynomial by the linearization process of Prop. 5.3, which clearly still applies to generalized polynomials. ■

The above proof uses simplifications by Martindale and Herstein, cf. Herstein [1976].

## Exercises

(1) Let $R$ be a simple ring with centre $C$. Show that Lemma 8.1 holds for $D = R$, $\Sigma = R$, $K = C$. Deduce that if $a, b \in R$ satisfy $axb = bxa$ for all $x \in R$, and $a \neq 0$, then $b = \lambda a$ for some $\lambda \in C$.

(2) Let $k$ be a field of characteristic 0. Show that in the Weyl algebra $A_1(k)$ with generators $u, v$ every non-trivial multilinear polynomial $f$ is non-zero when the variables in $f$ are replaced by suitable power products of $u$ and $v$.

(3) Let $G$ be a group and $\Delta(G)$ the set of elements of $G$ which have only finitely many conjugates; show that $\Delta(G)$ is a characteristic subgroup of $G$. Show that if the group algebra $kG$ has a GPI holding for all arguments in $G$, then $\Delta(G)$ is of finite index in $G$.

## Further exercises on Chapter 9

(1) Let $f$ be a homomorphism from a ring $R$ to a skew field $K$. Show that $f$ is an epimorphism iff $K$ is the skew subfield generated by im $f$. Show further that if $f$ is injective, then $K$ is left $R$-flat iff $R$ is a right Ore domain.

(2) Show that the skew field of fractions of a right Ore domain $R$ is unique up to an isomorphism (leaving $R$ fixed). (Note that this does not extend to general rings, e.g. a free algebra has many non-isomorphic skew fields of fractions, cf. Ex. (2) of **9.3** and Cohn [1977, 1985].)

(3) Let $k$ be an algebraically closed field and let $A$ be the translation ring over $k$, generated by $x, y$ with $xy = y(x+1)$ (cf. **9.3**, Example 4). Verify that $yA$ is a prime ideal and its complement is an Ore set. Show that any prime ideal other than 0 or $yA$ has the form $\mathfrak{c}_\alpha = yA + (x-\alpha)A$, where $\alpha \in k$. Verify that $y\mathfrak{c}_\alpha = \mathfrak{c}_{\alpha+1}y$ and deduce that the complement of $\mathfrak{c}_\alpha$ is not an Ore set, but the complement of $\bigcap_n \mathfrak{c}_{\alpha+n}$ is an Ore set, for each $\alpha \in k$.

(4) In the $k$-algebra with generators $x, y$ and defining relation $xy = \lambda yx$, where $\lambda \in k^\times$, find the maximal prime ideals and the complements of intersections of prime ideals that are Ore sets. (Hint: treat the case where $\lambda$ is a root of 1 separately.)

(5) (Ore) Let $k$ be a field containing a finite subfield $\mathbf{F}_q$. Show that the elements of $k[x]$ which as functions on $k$ are linear over $\mathbf{F}_q$ are the *q-polynomials* $\sum a_i x^{q^i}$, and that they form a ring under substitution as multiplication: $fg(x) = f(g(x))$. Verify that this ring is isomorphic to $k[z; \varphi]$, where $\varphi: a \mapsto a^q$.

(6) (P. Fatou) A power series over $\mathbf{Z}$ is called *primitive* if no prime divides all its coefficients. Show that the product of primitive power series is primitive. Deduce that if $P, Q \in \mathbf{Z}[x]$ are coprime polynomials such that the power series $P/Q$ has integer coefficients, then $Q(0) = \pm 1$. (Hint: find polynomials $f, g \in \mathbf{Z}[x]$ such that $fP + gQ = m$ is a positive integer and express $m$ as a product of $Q$ and another series in $\mathbf{Z}[\![x]\!]$.

(7) Let $R$ be a right Bezout domain. Show that any finitely generated torsion free left $R$-module is free. Deduce that over a 2-sided Bezout domain every finitely generated module splits over its torsion submodule.

(8) Let $R$ be a right Ore domain and $K$ its field of fractions. Show that any ring between $R$ and $K$ is again right Ore.

(9) Show that a semiprime right Goldie ring satisfies the minimum condition on right annihilator ideals. (Hint: in any chain the rank becomes stationary; now use (a)$\Rightarrow$(c) in the proof of Th. 4.9 and **L**.2 to show that essential extensions are trivial.)

(10) A module is called *semi-Artinian* if every non-zero quotient contains a simple submodule. Show that a semi-Artinian module is non-singular iff its socle is projective.

(11) (P. J. Higgins) Write $P(x,y)=\sum_0^{n-1} x^i y x^{n-1-i}$. Show that if an algebra $A$ over a field of characteristic prime to $n!$ satisfies $x^n=0$, then it also satisfies $\sum x_{1\sigma}\cdots x_{n\sigma}=0$, where $\sigma$ runs over all permutations of $1,\ldots,n$, and deduce that $P(x,y)=0$ in $A$. By evaluating the expression $\sum x^i z y^j x^{n-1-i} y^{n-1-j}$ in two ways, show that $A$ satisfies the identity $x^{n-1}zy^{n-1}=0$.

(12) (P. J. Higgins) Use Ex. (11) to prove the Nagata–Higman theorem: if an algebra $A$ of characteristic prime to $n!$ satisfies the identity $x^n=0$, then $A^{2^n-1}=0$. (Hint: let $I$ be the ideal generated by all elements $a^{n-1}$, where $a\in A$. Show that $n!IAI=0$ by Ex. (11) and apply induction on $n$ to $A/I$.)

# 10
# Rings without finiteness assumption

For general rings there is naturally not as much structure theory as in the Artinian or Noetherian case. It is true that some of the same methods can be used, e.g. a radical can be defined, semiprimitive rings can be expressed as subdirect products of primitive rings etc., but these methods are less precise and they do not lead to a complete classification. For primitive rings a structure theorem can be proved using a general version of the density theorem; in the case of simple rings with minimal right ideals this leads to a more explicit representation. We examine the connexion between semiprime and prime rings and see the effect of a polynomial identity in this case. We also take a brief look at a natural generalization of principal ideal domains to the non-commutative case: free ideal rings.

## 10.1 The endomorphism ring of a vector space

Let $K$ be a skew field and $V$ a left $K$-module; we know from Th. 1.4.5 that $V$ is free as $K$-module and that any two bases of $V$ have the same cardinal, called the *rank* or also the *dimension* of $V$ over $K$ and written $[V:K]$. Let $E = \mathrm{End}_K(V)$ be its endomorphism ring; when $[V:K]$ is finite, equal to $n$ say, then $E \cong \mathrm{End}_K(K^n) \cong \mathfrak{M}_n(K)$, the $n \times n$ matrix ring over $K$, and we know from **5.3** of A.2 that this is a simple ring. We now ask: what can we say about $E$ when $V$ is infinite-dimensional? In this case $E$ is no longer simple; its ideal structure is described in Th. 1.3 below.

The first step is to find a matrix representation for $E$. Here it is not necessary for $K$ to be a skew field; we may take any ring $R$ and consider a free $R$-module of infinite rank $\nu$. Thus we take an index-set $I$ of cardinal $\nu$ and take a free left $R$-module $V$ with basis $\{v_\alpha\}(\alpha \in I)$. In terms of this basis any endomorphism $a$ of $V$ is described by the equations expressing the image of each $v_\alpha$ in terms of the $v$'s:

$$v_\alpha a = \sum_\beta a_{\alpha\beta} v_\beta, \qquad \text{where } a_{\alpha\beta} \in R. \tag{1}$$

Here $(a_{\alpha\beta})$ is a $\nu \times \nu$ matrix, i.e. a square array of elements of $R$, whose rows and columns are indexed by a set $I$ of cardinal $\nu$. Moreover, for each $\alpha \in I$, there are only finitely many non-zero coefficients $a_{\alpha\beta}$ in (1), hence each row of the matrix $(a_{\alpha\beta})$ contains only finitely many non-zero entries, i.e. it is row-finite. Conversely, every row-finite $\nu \times \nu$ matrix over $R$ defines an endomorphism of $V$ relative to the basis $\{v_\alpha\}$. For we can define $v_\alpha a$ by (1), and for a general element $x = \Sigma \xi_\alpha v_\alpha$ of $V$ put

$$xa = \Sigma \xi_\alpha a_{\alpha\beta} v_\beta.$$

It is easily checked that the mapping $a$ so defined is an endomorphism of $V$, so that we have a bijection between $\mathrm{End}_R(V)$ and the set $\mathfrak{M}_\nu(R)$ of all row-finite $\nu \times \nu$ matrices over $R$. If we define addition and multiplication of row-finite matrices, as in the finite case, by the formulae

$$(a_{\alpha\beta}) + (b_{\alpha\beta}) = (a_{\alpha\beta} + b_{\alpha\beta}), \qquad (a_{\alpha\beta})(b_{\alpha\beta}) = (\sum a_{\alpha\lambda} b_{\lambda\beta}),$$

we find that $\mathfrak{M}_\nu(R)$ is a ring isomorphic to $E$. We observe that some restriction on the matrices, such as row-finiteness, is essential for the product to be defined. Our conclusion may be stated as

THEOREM 1.1 *Let $R$ be any ring and $V$ a free left $R$-module of infinite rank $\nu$. Then $\mathrm{End}_R(V)$ is isomorphic to the ring of all row-finite $\nu \times \nu$ matrices over $R$.* ∎

Let us return to the case of a skew field $K$. When $V$ is a finite dimensional $K$-space, say $[V:K] = n$, then $\mathrm{End}_K(V) \cong \mathfrak{M}_n(K)$ is a simple ring, and it can be written as a direct sum of $n$ pairwise isomorphic simple left ideals (A.2, Th. 5.3.2), corresponding to the columns of the matrix. In the infinite case the situation is rather different. We still have the minimal left ideals, corresponding to the columns of the matrix, but we can no longer express the general matrix as the sum of a finite number of such columns, and $\mathrm{End}_K(V)$ is no longer simple.

For any $a \in \mathrm{End}_K(V)$ let us define the *rank* of $a$, $\rho(a)$, as the $K$-dimension of the image space:

$$\rho(a) = [\mathrm{im}\, a : K].$$

This agrees with the usual definition in the finite-dimensional case, and we have the following rules:

**R.1** *$\rho(a)$ is a cardinal satisfying $0 \leqslant \rho(a) \leqslant [V:K]$,*
**R.2** $\rho(a) = 0 \Leftrightarrow a = 0$,
**R.3** $\rho(a - b) \leqslant \rho(a) + \rho(b)$,
**R.4** $\rho(ab) \leqslant \min\{\rho(a), \rho(b)\}$.

Of these, **R.**1, 2 are clear and **R.**3 follows because $V(a-b) \subseteq Va + Vb$, and so $[V(a-b):K] \leqslant [Va:K] + [Vb:K]$. Turning to **R.**4, we clearly have $Vab \subseteq Vb$,

hence $\rho(ab) \leqslant \rho(b)$, but also $[Vb:K] \leqslant [V:K]$; hence on replacing $V$ by $Va$ we find that $[Vab:K] \leqslant [Va:K]$, and so $\rho(ab) \leqslant \rho(a)$. ■

To elucidate the ideal structure of $\mathrm{End}_K(V)$, we note the following relation between ranks of endomorphisms:

LEMMA 1.2 *Let $K$ be a skew field and $V$ a left $K$-space. If $a, b \in \mathrm{End}_K(V)$ and $\rho(a) \geqslant \rho(b)$, then there exist $p, q \in \mathrm{End}_K(V)$ such that*

$$b = paq, \tag{2}$$

*and $\rho(p) = \rho(q) = \rho(b)$.*

*Proof.* Choose complements $N_a$, $N_b$ of $\ker a$, $\ker b$ in $V$ respectively, so that $V = \ker a \oplus N_a = \ker b \oplus N_b$. Then $N_a \cong \mathrm{im}\, a$, $N_b \cong \mathrm{im}\, b$, so by hypothesis, $[N_a:K] \geqslant [N_b:K]$, hence there exists $p \in \mathrm{End}_K(V)$ mapping $\ker b$ to 0 and embedding $N_b$ in $N_a$; clearly $\rho(p) = \rho(b)$. If $\{u_\alpha\}$ is a basis of $N_b$, then the $u_\alpha p$ are linearly independent in $N_a$ and the $u_\alpha pa$ are linearly independent, because the restriction $a|N_a$ is injective. Likewise the $u_\alpha b$ are linearly independent in $Vb$. Now choose a complement $L$ in $V$ for the subspace spanned by the $u_\alpha pa$, and define $q$ as the endomorphism mapping $L$ to 0 and $u_\alpha pa$ to $u_\alpha b$. Then $\rho(q) = \rho(b)$ and (2) holds, for both sides map $u_\alpha$ to $u_\alpha b$ and $\ker b$ to 0. ■

With this preparation we can describe the ideal structure of $\mathrm{End}_K(V)$:

THEOREM 1.3 *Let $K$ be a skew field and $V$ a $K$-space of infinite dimension $v$. For any infinite cardinal $\mu$ denote by $E_\mu$ the set of all endomorphisms of $V$ of rank less than $\mu$. Then the $E_\mu$ (for $\mu \leqslant v$) are distinct, each $E_\mu$ is an ideal in $E = \mathrm{End}_K(V)$, and these are the only ideals apart from 0 and $E$.*

*Proof.* Let $a, b \in E_\mu$, where $\mu$ is an infinite cardinal. Then

$$\rho(a - b) \leqslant \rho(a) + \rho(b) < 2\mu = \mu,$$

(by A.2, Prop. 1.2.6), hence $a - b \in E_\mu$. Next, if $a \in E_\mu$ and $c \in E$, then $\rho(ac) < \mu$, $\rho(ca) < \mu$, by **R.**4, and this shows $E_\mu$ to be an ideal in $E$. There are endomorphisms of any rank $\leqslant v$, e.g. projections on subspaces, hence all the $E_\mu$ are distinct. It remains to show that there are no other ideals.

Let $\mathfrak{a}$ be any non-zero ideal in $E$ and denote by $\mu$ the least cardinal $> \rho(a)$ for all $a \in \mathfrak{a}$. Then $\mu$ is infinite; for $\mathfrak{a}$ contains endomorphisms of positive rank, hence (by the lemma) $\mathfrak{a}$ contains all endomorphisms of rank 1. But every endomorphism of finite rank can be written as a sum of endomorphisms of rank 1, so $\mathfrak{a}$ contains all endomorphisms of finite rank and $\mu$ must be infinite. By definition, each $a \in \mathfrak{a}$ has rank $< \mu$, hence $\mathfrak{a} \subseteq E_\mu$. Conversely, if $b \in E_\mu$ then $\rho(b) < \mu$, hence $\rho(b) \leqslant \rho(a)$ for some $a \in \mathfrak{a}$. By Lemma 1.2, $b = paq$ for some $p, q \in E$,

so $b \in \mathfrak{a}$, and this shows that $\mathfrak{a} = E_\mu$. Thus every non-zero ideal in $E$ is of the form $E_\mu$ for some infinite cardinal $\mu$, and clearly if $\mu > \nu$, then $E_\mu = E$. ■

### Exercises

(1) In $\mathrm{End}_K(V)$ show that the endomorphisms with image in a given one-dimensional subspace form a minimal left ideal and that all such left ideals are isomorphic. Show further that the sum of all these left ideals is the unique minimal ideal in $E$ (it is the *socle* of $E$).

(2) Let $E = \mathrm{End}_K(V)$ be as in the text, where $[V:K] = \nu$ is infinite. Verify that $E/E_\nu$ is simple but not semisimple. Show that for $a, b \in E$, there exists $p$ such that $b = pa$ iff $\mathrm{im}\, b \subseteq \mathrm{im}\, a$ and there exists $q$ such that $b = aq$ iff $\ker b \supseteq \ker a$. Deduce that $E$ is neither Artinian nor Noetherian; prove the same for $E/E_\nu$. (Hint: consider the set of endomorphisms with image in a given finite-dimensional subspace, or with kernel containing a given finite-dimensional subspace.)

(3) In the notation of Th. 1.3 show that $E_\mu$, as algebra without 1, has as ideals precisely the $E_\lambda (\lambda \leqslant \mu)$ and 0.

(4) Show that for any vector space $V$ over a field $k$, $\mathrm{End}_k(V)$ is a regular ring.

(5) Let $k$ be any field of characteristic not 2 and $A = (a_{ij})$ a square matrix with rows and columns indexed by $\mathbf{Z}$, such that $a_{ii} = 1$, $a_{ii+1} = p$, $a_{ii-1} = q$, where $p, q \in k$, $p + q = 1$, and $a_{ij} = 0$ for $|i-j| > 1$. Define $v_n$ recursively by $v_n = \delta_{n0} - pv_{n-1} - qv_{n+1} (n \in \mathbf{Z})$, $v_0, v_1$ arbitrary. Show that (*i*) $A$ has a two-sided inverse $B = (b_{ij})$, where $b_{ij} = v_{j-i}$. (ii) $A$ is a two-sided zero-divisor: $AC = CA = 0$, where $C = (c_{ij})$, $c_{ij} = (-1)^{i+j}$. Explain this paradox.

## 10.2 The density theorem revisited

One of the basic results of ring theory is the Wedderburn–Artin theorem. A simple Artinian ring is a matrix ring over a skew field (A.2, Th. 5.3.2). A related result is the density theorem (Th. 7.1.1) which for a simple ring $A$ finite-dimensional over its centre $k$ tells us that $A^0 \otimes A$ is a full matrix ring over $k$. In 1945 Jacobson (and independently, Chevalley) proved a far-reaching generalization, as part of his theory of rings without finiteness conditions. Our presentation follows Bourbaki.

Let $R$ be any ring and $M$ a right $R$-module; the action of $R$ on $M$ may be described by saying that we have a homomorphism of $R$ into $\mathrm{End}(M)$, each $a \in R$ corresponding to an endomorphism $a'$ of $M$, qua additive group. The centralizer $S$ of this set $R' = \{a' | a \in R\}$ in $\mathrm{End}(M)$ is the ring of all $R$-endomorphisms, $S = \mathrm{End}_R(M)$, and we shall regard $M$ as left $S$-module. Since $(\alpha x)a = \alpha(xa)$ for all $x \in M$, $a \in R$, $\alpha \in S$, by definition of $S$, we see that $M$ becomes an $(S, R)$-bimodule in this way. Let $T$ be the centralizer of $S$; this is again a subring of $\mathrm{End}(M)$,

called the *bicentralizer* of $R$. Clearly for any $a \in R$, $a'$ centralizes $S$ and so lies in $T$, i.e. $R' \subseteq T$. If equality holds, we say that $R$ acts *bicentrally* on $M$, or also that $M_R$ is *bicentral.*

As an example consider the ring $R$ itself. As we saw in Th. 5.1.3 of A.2, the centralizer of ${}_RR$ is the set of all right multiplications; by symmetry the centralizer of $R_R$ is the set of all left multiplications, hence ${}_RR$ as well as $R_R$ is bicentral. For this property the presence of a unit element is of course material.

The definition of 'bicentral' may be restated as follows: given a right $R$-module $M$, $R$ acts *bicentrally* if for each $\theta$ in the bicentralizer $T$ there exists $a \in R$ such that

$$x\theta = xa \qquad \text{for all } x \in M. \tag{1}$$

This is a very strong requirement, because $a$ in (1) is independent of $x$. To obtain a weaker condition, let us say that $R$ acts *fully* on $M$ if for each $\theta$ in the bicentralizer $T$ and each $x \in M$ there exists $a \in R$ such that $x\theta = xa$, where $a$ may depend on the choice of $x \in M$. But the most useful condition is intermediate between these two. We shall say that $R$ acts *densely* on $M$ if for each $\theta \in T$ and each finite family $x_1, \ldots, x_n \in M$ there exists $a \in R$ such that

$$x_i\theta = x_i a \qquad \text{for } i = 1, \ldots, n. \tag{2}$$

It is clear that bicentral $\Rightarrow$ dense $\Rightarrow$ full. The next result gives a useful condition under which dense $\Leftrightarrow$ bicentral.

PROPOSITION 2.1 *Let $M$ be a right $R$-module with centralizer $S$. If $M$ is finitely generated as $S$-module, then $R$ acts densely on $M$ if and only if it acts bicentrally.*

*Proof.* Let $M = \sum_1^n Su_i$; given $\theta$ in the bicentralizer, choose $a \in R$ such that $u_i\theta = u_i a$ $(i = 1, \ldots, n)$. Now consider the set $N = \{x \in M \mid x\theta = xa\}$. This is clearly an $S$-submodule of $M$ and it contains the generating set $u_1, \ldots, u_n$ of $M$, hence $N = M$, i.e. $x\theta = xa$ for all $x \in M$, so $R$ acts bicentrally. The converse is clear. ■

To motivate the terminology we remark that $\mathrm{End}(M)$ may be regarded as a set of mappings from $M$ to $M$, i.e. a subset of $M^M$, and this set can be topologized as a product, taking the discrete topology on $M$. The topology so defined on $\mathrm{End}(M)$ is called the *topology of pointwise convergence.* Given $f \in M^M$, we obtain a typical neighbourhood of $f$ by taking a finite set $x_1, \ldots, x_n \in M$ and considering all $f' \in M^M$ such that $x_i f' = x_i f$ $(i = 1, \ldots, n)$. Now (2) may be restated by saying that $R$ is *dense* in the bicentralizer $T$ (i.e. every non-empty open subset of $T$ contains an element of $R$). In this connexion we note that any centralizer and hence any bicentralizer is closed in the topology defined here on $\mathrm{End}(M)$.

Let $M$ be a right $R$-module, denote the centralizer (acting on the left) by $S$ and the bicentralizer by $T$. By Cor. 4.3.2 of A.2, the centralizer of ${}^nM_R$ is $S_n$,

so that we may regard ${}^nM$ as left $S_n$-module, and the centralizer of this module is $T$, again by A.2, Th. 4.3.6. Thus we have

PROPOSITION 2.2 *For any $R$-module $M$ and any $n \geqslant 1$, $M$ and ${}^nM$ have isomorphic bicentralizers.* ■

To say that $R$ acts densely on $M$ means that for all $n$ and all $x \in {}^nM$, $\theta \in T$ there exists $a \in R$ such that $x\theta = xa$. This just states that $R$ acts fully on ${}^nM$, for all $n$; thus we have proved

THEOREM 2.3 (Density theorem) *Let $M_R$ be any module. Then $R$ acts densely on $M$ if and only if $R$ acts fully on ${}^nM$, for all $n \geqslant 1$.* ■

We give several applications, which show the power of this result. In the first place, we clearly have

COROLLARY 2.4 *If $R$ acts densely on $M$, then it acts densely on ${}^nM$, for any $n \geqslant 1$.* ■

An important case of dense action is provided by semisimple modules.

THEOREM 2.5 *Let $M_R$ be a semisimple module. Then $R$ acts densely on $M$. Moreover, if $R$ is right Artinian, then $M$ is finitely generated over the centralizer of $R$ and $R$ acts bicentrally on $M$.*

*Proof.* Denote by $S$ the centralizer and by $T$ the bicentralizer of $R$ on $M$. We first show that every $R$-submodule of $M$ is a $T$-submodule. Let $M_1$ be an $R$-submodule; since $M$ is semisimple, we have $M = M_1 \oplus M_2$ for a submodule $M_2$. Let $\theta_1$ be the projection on $M_1$; then $\theta_1 \in S$, hence for any $t \in T$ and $x \in M_1$ $xt = (\theta_1 x)t = \theta_1(xt)$, so $xt \in M_1$ as claimed.

Next we show that $R$ acts fully on $M$, i.e. for each $\theta \in T$ and $x \in M$ there exists $a \in R$ such that $x\theta = xa$. This states that for each $x \in M$, $xT \subseteq xR$. But $xR$ is an $R$-submodule containing $x$, hence it admits $T$, by what has been shown, and so $xT \subseteq xR$, as required.

We apply this result to ${}^nM$. This is again semisimple, hence $R$ acts fully on ${}^nM$ and so, by the density theorem, $R$ acts densely on $M$.

Finally assume that $R$ is right Artinian. By Th. 4.2.6 of A.2, $M$ is semisimple as $S$-module, say $M = \oplus Su_i$, where each $Su_i$ is simple. We claim that this sum is finite; for if not, then there is a countable subset: $u_1, u_2, \ldots$; If we put $\mathfrak{a}_n = \{a \in R \mid u_1 a = \cdots = u_n a = 0\}$, then $\mathfrak{a}_n$ is a right ideal of $R$ and

$$\mathfrak{a}_1 \supseteq \mathfrak{a}_2 \supseteq \cdots. \tag{3}$$

The projection of $M$ on the complement of $Su_1 \oplus \cdots \oplus Su_n$ is an

$S$-endomorphism, i.e. there exists $\theta \in T$ such that $u_i\theta = 0$ $(i = 1, \ldots, n)$, $u_i\theta = u_i$ $(i > n)$. By density there exists $a \in R$ such that $u_i a = 0 (i = 1, \ldots, n)$, $u_{n+1}a = u_{n+1}$. Hence $a \in \mathfrak{a}_n \setminus \mathfrak{a}_{n+1}$ and this shows the inclusions in (3) to be strict. But this contradicts the fact that $R$ is right Artinian; therefore $M = Su_1 \oplus \cdots \oplus Su_n$ for some $n$, and hence by Prop. 2.1, $R$ acts bicentrally. ■

For a simple module we can say rather more. By combining Cor. 2.4 with Schur's lemma (Lemma 5.3.1 of A.2), we obtain a generalization of Wedderburn's first structure theorem:

COROLLARY 2.6 *Let $M_R$ be a simple module. Then the centralizer $K$ of $R$ is a skew field and $R$ acts as a dense ring of linear transformations on $M$. In particular, if $R$ is right Artinian, then the image $\tilde{R}$ of $R$ in* End$(M)$ *acts bicentrally, and so is a full matrix ring over $K$.*

Only the last part still needs proof. If $R$ is right Artinian, $M$ is finitely generated over $K$ and so is finite-dimensional, say $M = K^n$; hence $\tilde{R}$ as the centralizer of $K$ is $K_n$. ■

## Exercises

(1) Show that in any monoid the centralizer of any subset is its own bicentralizer. Deduce that for any $R$-module $M$, $\mathrm{End}_R(M)$ acts bicentrally.

(2) Let $A$ be a finitely generated abelian group, as $\mathbf{Z}$-module. Show that $\mathbf{Z}$ acts bicentrally on $A$.

(3) Show that any simple ring with a minimal right ideal is right Artinian and so of the form $\mathfrak{M}_n(K)$, $K$ a skew field (here the presence of a unit element is essential, cf. **10.8**).

(4) Show that for any $R$-module $M$, $R$ acts bicentrally on $R \oplus M$. (Hint: for any $u \in R \oplus M$ find $\alpha \in \mathrm{End}_R(R \oplus M)$ such that $1\alpha = u$ and deduce that for any $\lambda$ in the bicentralizer, $u\lambda = ua$, where $a = 1 \cdot \lambda \in R$.)

(5) Show that $R$ acts bicentrally on any generator of the category $\mathscr{M}_R$. (Hint: recall that $M$ is a generator iff $R$ is a direct summand of ${}^nM$ and use Ex. (4).)

(6) Show that a simple ring acts bicentrally on any right ideal.

## 10.3 Primitive rings

Let $R$ be any ring and $M$ a right $R$-module. We say that $R$ acts *faithfully* on $M$ or that $M$ is a *faithful* $R$-module if for any $a \in R$, $a \neq 0$, we have $Ma \neq 0$. This means that the standard homomorphism $R \to \mathrm{End}(M)$ is injective. A ring $R$ is

called *primitive* if it has a simple faithful right $R$-module. Strictly speaking, this notion should be called 'right primitive' and a corresponding notion 'left primitive' defined. In fact these concepts are distinct, as examples show, but we shall only be dealing with right primitive rings and therefore omit the qualifying adjective.

To obtain an internal characterization of primitive rings, we define for any right ideal $\mathfrak{a}$ of $R$ its *core* as the set

$$(\mathfrak{a}:R) = \{x \in R \mid Rx \subseteq \mathfrak{a}\}. \tag{1}$$

If we regard $M = R/\mathfrak{a}$ as right $R$-module, $(\mathfrak{a}:R)$ is the annihilator of $M$ in $R$. This shows the core to be an ideal in $R$; moreover, $(\mathfrak{a}:R) \subseteq \mathfrak{a}$ and any (two-sided) ideal of $R$ contained in $\mathfrak{a}$ is contained in $(\mathfrak{a}:R)$, by the definition; thus the core of $\mathfrak{a}$ is the largest ideal of $R$ contained in $\mathfrak{a}$. With its help primitive rings can be characterized as follows:

PROPOSITION 3.1 *A ring $R$ is (right) primitive if and only if $R$ contains a maximal right ideal whose core is zero.*

For if $R$ is primitive, there is a faithful simple right $R$-module $M$. We have $M = R/\mathfrak{a}$ where $\mathfrak{a}$ is a maximal right ideal (by the simplicity of $M$), and since $M$ is faithful, $(\mathfrak{a}:R) = 0$. Conversely, if $\mathfrak{a}$ is a maximal right ideal with zero core, then $R/\mathfrak{a}$ is faithful simple right $R$-module. ■

COROLLARY 3.2 *A commutative ring is primitive if and only if it is a field.*

For in this case the core of $\mathfrak{a}$ is $\mathfrak{a}$ itself, so 0 must be a maximal ideal of $R$ and this means that $R$ is a field. ■

To give an example, any simple ring (Artinian or not) is primitive, for $R$ has a maximal right ideal, by Krull's theorem (A.2, Th. 2.2.11) and its core is a proper ideal, which must be 0. The converse is false: if $V$ is an infinite-dimensional vector space over a field and $E$ is its endomorphism ring, then $E$ acts faithfully on $V$ and $V$ is clearly simple as $E$-module, so $E$ is primitive, but as we saw in **10.1**, $E$ is not simple. The next result describes primitive rings more precisely:

THEOREM 3.3 *Any primitive ring is isomorphic to a dense ring of linear transformations in a vector space over a skew field $K$. Conversely, any dense subring of $\mathrm{End}_K(V)$ is primitive.*

*Proof.* Given a primitive ring $R$, let $V$ be a simple $R$-module on which $R$ acts faithfully. Its centralizer is a skew field $K$, by Schur's lemma, and $R$ is naturally embedded as a dense subring in $\mathrm{End}_K(V)$, by Cor. 2.6. For the converse we need only observe that any dense subring $R$ of $\mathrm{End}_K(V)$ acts simply: given

$u, v \in V$, $u \neq 0$, there exists $a \in R$ such that $ua = v$. Hence the submodule generated by any $u \neq 0$ is $V$, i.e. $V$ is simple. ■

As we saw, a primitive ring need not be simple, but if $R$ is right Artinian as well as primitive, say it is a dense subring of $\mathrm{End}_K(V)$, then $V$ is finitely generated over $K$ by Th. 2.5, hence $V \cong K^n$ and $R$ acts bicentrally: $R \cong \mathrm{End}_K(K^n) \cong K_n$. This expression is unique up to an isomorphism of $K$, as we saw in Th. 5.3.2 of A.2. In the general case there is no such uniqueness, but we have the following consequence which is sometimes useful:

PROPOSITION 3.4 *Let $R$ be a primitive ring which is not right Artinian. Then for every $n \geqslant 1$, $R$ has a subring with a homomorphism onto an $n \times n$ matrix ring over a skew field.*

*Proof.* We may take $R$ to be a dense subring of $\mathrm{End}_K(V)$, where $V$ is an infinite-dimensional vector space over $K$. Given $n \geqslant 1$, take an $n$-dimensional subspace $U$ of $V$, with a basis $u_1, \ldots, u_n$. Let $R_1$ be the subring of $R$ mapping $U$ into itself; every element of $R_1$ defines by restriction an endomorphism of $U$, thus we have a homomorphism $R_1 \to \mathrm{End}_K(U) \cong K_n$, and this is surjective, by density, so it is the required homomorphism. ■

Let $R$ be any ring with a minimal right ideal and define the *socle* $\mathfrak{s}$ of $R$ as the sum of all minimal right ideals. This socle is an ideal, for, given any minimal right ideal $\mathfrak{a}$ of $R$ and $x \in R$, $x\mathfrak{a}$ is a minimal right ideal or 0 and so is contained in $\mathfrak{s}$. Primitive rings with non-zero socle have a more precise description:

THEOREM 3.5 *A primitive ring has a non-zero socle if and only if in its representation as a dense ring of linear transformations of a $K$-space, $R$ contains transformations of finite (non-zero) rank. When this is so, all faithful simple right ideals of $R$ are isomorphic and the skew field $K$ is determined up to isomorphism as the centralizer of a faithful simple $R$-module.*

*Proof.* If an element of $R$ defines a linear transformation of finite rank, take $c \in R$ such that the rank $\rho(c)$ is the least positive number possible. Then $\ker cx \supseteq \ker c$ for any $x \in R$, and if $cx \neq 0$, then $\rho(cx) = \rho(c)$, hence $\ker cx = \ker c$, and the complement of $\ker c$ is finite-dimensional. By density we can find $y \in R$ such that $cxy = c$; this shows $cR$ to be minimal, hence $R$ has a non-zero socle.

Now assume that $R$ has a non-zero socle, and hence a minimal right ideal $\mathfrak{a}$. If $M$ is any faithful simple right $R$-module, then $M\mathfrak{a} \neq 0$, so there exists $u \in M$ such that $u\mathfrak{a} \neq 0$. But $u\mathfrak{a}$ is a submodule of $M$, hence $u\mathfrak{a} = M$ and so the mapping $x \mapsto ux$ $(x \in \mathfrak{a})$ is a surjective homomorphism $\mathfrak{a} \to M$. By the minimality of $\mathfrak{a}$, $M \cong \mathfrak{a}$ as right $R$-module, so $\mathfrak{a}$ is also faithful; this shows that every simple faithful right $R$-module is isomorphic to $M$. It follows that the isomorphism

type of $K$ is uniquely determined as the endomorphism ring of $\mathfrak{a}_R$. Finally, since $\mathfrak{a}$ is faithful, $\mathfrak{a}^2 \neq 0$, hence $\mathfrak{a}^2 = \mathfrak{a}$, and so $c\mathfrak{a} = \mathfrak{a}$ for some $c \in \mathfrak{a}$. We claim that $\rho(c) = 1$; for if not, then there exist $x, y \in V$, where $V$ is the $K$-space on which $R$ acts, such that $xc$, $yc$ are linearly independent, and by density there exists $b \in R$ such that $xcb \neq 0$, $ycb = 0$. Then $\mathfrak{a} = cR$ meets the annihilator $\mathfrak{n}$ of $y$ in $R$, and $\mathfrak{n}$ is a right ideal, so $\mathfrak{a} \subseteq \mathfrak{n}$, by the minimality of $\mathfrak{a}$. But this means that $yc = 0$, which is a contradiction, and it shows that $R$ contains elements of rank 1. ■

### Exercises

(1) For any ring $R$ and any $n \geqslant 1$, show that $R_n$ is primitive iff $R$ is. More generally, show that being primitive is a Morita invariant.

(2) Show that every minimal right ideal in a primitive ring has an idempotent generator. (Hint: use the primitivity to show that the right ideal is not nilpotent.)

(3) Show that if $e$ is a non-zero idempotent in a primitive ring $R$, then $eRe$ is primitive.

(4) Show that for any idempotent $e$ in a ring $R$, $eR$ is a minimal right ideal iff $Re$ is a minimal left ideal. Deduce that in a primitive ring with non-zero socle, the socle coincides with the left socle (defined correspondingly).

(5) Show that the socle of a primitive ring is a minimal two-sided ideal. Deduce that a simple ring with non-zero socle is Artinian.

(6) Show that a left Artinian primitive ring is simple. What can be said about a primitive ring with minimal left ideals?

(7) Show that the centre of a primitive ring is an integral domain. If $A$ is any commutative integral domain with field of fractions $K$, then the set of infinite matrices over $K$ which are equal to a scalar in $A$ outside a finite square is a primitive ring with centre $A$.

## 10.4 Semiprimitive rings and the Jacobson radical

In the last section we defined primitive rings as rings with a faithful simple module. Often one needs to consider a wider class of rings, and we define a ring to be *semiprimitive* if it has a faithful semisimple module; clearly it is equivalent to require that for each $a \in R^\times$ there exists a simple module $M$ such that $Ma \neq 0$; for if such a module $M_a$ is chosen for each $a$, then $\oplus M_a$ is faithful and semisimple, and conversely, if $M$ is faithful and semisimple, then each $a \neq 0$ acts non-zero on at least one simple summand.

We recall from A.2 (Lemma 5.4.2) that $\mathbf{J}(R)$, the Jacobson radical of $R$, was defined in the following equivalent ways:

$\mathbf{J}(R)$ is the set of all $a \in R$ such that

(a) $Ma = 0$ *for each simple right* $R$*-module* $M$,

(b) *a belongs to each maximal right ideal,*
(c) *1-ay has a right inverse for each* $y \in R$,
(d) *1-xay has an inverse for all* $x, y \in R$,
($a^0$)–($d^0$) *the left-hand analogues of* (a)–(d).

It is clear from (a) of this definition that a ring $R$ is semiprimitive precisely when $\mathbf{J}(R) = 0$. Moreover, the symmetry of the definition shows that the notion 'semiprimitive' is left–right symmetric. This is in contrast to the situation for primitive rings, where one-sided examples exist, first found by Bergman 1962 and Jategaonkar 1968 (cf. e.g. Cohn [1985]).

We also recall that a *quasi-inverse* of an element $c$ is an element $c'$ such that

$$c + c' = cc' = c'c. \tag{1}$$

The quasi-inverse, when it exists, is unique, because $1 - c'$ is the inverse of $1 - c$. An element which has a quasi-inverse is sometimes called *quasi-regular*; for example, any nilpotent element is quasi-regular. Now $\mathbf{J}(R)$ may also be defined as the largest ideal consisting entirely of quasi-regular elements; this is an easy consequence of (d) above.

Semiprimitive rings admit a subdirect product representation which is sometimes useful; however, it should be borne in mind that a given product may contain many different subdirect products, and the relation to the direct product is not very close.

THEOREM 4.1 *Every semiprimitive ring $R$ is a subdirect product of primitive rings which are homomorphic images of $R$. Conversely, every subdirect product of primitive rings is semiprimitive.*

*Proof.* Let $\{\mathfrak{p}_\lambda\}$ be the family of all maximal right ideals of the semiprimitive ring $R$, and denote the core of $\mathfrak{p}_\lambda$ by $\mathfrak{c}_\lambda$. Since $R$ is semiprimitive, we have $\bigcap \mathfrak{p}_\lambda = 0$, and since $\mathfrak{c}_\lambda \subseteq \mathfrak{p}_\lambda$, it follows that $\bigcap \mathfrak{c}_\lambda = 0$. If we put $R_\lambda = R/\mathfrak{c}_\lambda$, then $R_\lambda$ is primitive, for it is represented faithfully on the simple module $R/\mathfrak{p}_\lambda$. Now the natural maps $f_\lambda : R \to R_\lambda$ can be combined to a homomorphism into the direct product

$$f : R \to P = \prod R_\lambda, \tag{2}$$

and $\ker f = \bigcap \ker f_\lambda = \bigcap \mathfrak{c}_\lambda = 0$. Thus $f$ is injective and if $\varepsilon_\lambda : P \to R_\lambda$ denotes the canonical projection, then $f\varepsilon_\lambda = f_\lambda$ is surjective, by the definition of $R_\lambda$, so (2) is the required subdirect product representation of $R$. Conversely, if $\{R_\lambda\}$ is a family of primitive rings and $M_\lambda$ is a faithful simple $R_\lambda$-module, then for any subdirect product $R$ of the $R_\lambda$ and any $a \in R^\times$ we have $a\varepsilon_\lambda \neq 0$ for some $\lambda$, hence $a$ has a non-zero action on $M_\lambda$ and this shows $R$ to be semiprimitive. ■

In particular, when $R$ is commutative, we have by Cor. 3.2,

COROLLARY 4.2 *Any commutative semiprimitive ring is a subdirect product of fields, and conversely, such a subdirect product is semiprimitive.* ■

For example, $\mathbf{Z}$ is semiprimitive because $0, 2$ are the only quasi-regular elements, and so $\mathbf{J}(\mathbf{Z}) = 0$. Hence $\mathbf{Z}$ is a subdirect product of fields; in fact $\mathbf{Z}$ is a subdirect product of the fields $\mathbf{F}_p$, where $p$ ranges over all primes.

The definition of $\mathbf{J}(R)$ shows that it measures how far $R$ is from being semiprimitive. It is a pleasant (and by no means self-evident) property that $R/\mathbf{J}(R)$ is semiprimitive. This follows from the next result, itself more general.

PROPOSITION 4.3 *Let $R$ be a ring and $\mathfrak{a}$ an ideal such that $\mathfrak{a} \subseteq \mathbf{J}(R)$. Then*

$$\mathbf{J}(R/\mathfrak{a}) = \mathbf{J}(R)/\mathfrak{a}. \tag{3}$$

*Proof.* The natural homomorphism $R \to R/\mathfrak{a}$ induces a lattice-isomorphism between the lattice of all right ideals of $R/\mathfrak{a}$ and that of all right ideals of $R$ which contain $\mathfrak{a}$. Each maximal right ideal of $R/\mathfrak{a}$ corresponds to a maximal right ideal of $R$; the converse also holds because $\mathfrak{a} \subseteq \mathbf{J}(R) = \bigcap \{\text{max.right ideals}\}$. Taking intersections of these sets of maximal right ideals we obtain (3). ■

If in (3) we put $\mathfrak{a} = \mathbf{J}(R)$, the right-hand side reduces to 0 and we deduce

COROLLARY 4.4 *For any ring $R$, $R/\mathbf{J}(R)$ is semiprimitive.* ■

We note that (3) may not hold without the restriction on $\mathfrak{a}$, for the mapping $\mathbf{J}(R) \to \mathbf{J}(R/\mathfrak{a})$ is not generally surjective; e.g. if $R = \mathbf{Z}$, $\mathfrak{a} = (4)$, then $\mathbf{J}(\mathbf{Z}) = 0$, but $\mathbf{J}(\mathbf{Z}/4) \neq 0$.

From **5.4** of A.2 we know that the Jacobson radical contains all nil ideals and for an Artinian ring $R$, $\mathbf{J}(R)$ is nilpotent, although in general $\mathbf{J}(R)$ need not even be nil, e.g. in the power series ring $k[\![x]\!]$ the Jacobson radical is $(x)$. However, in the absence of nil ideals we obtain a semiprimitive ring by adjoining an indeterminate:

PROPOSITION 4.5 (Amitsur) *If $R$ is a ring with no non-zero nil ideals, then $R[t]$ is semiprimitive, where $t$ is an indeterminate.*

*Proof.* We have to show that the radical $J$ of $R[t]$ is zero. If $J \neq 0$, let $\mathfrak{a}$ be the set consisting of 0 and all leading coefficients of elements in $J$. It is clear that $\mathfrak{a}$ is an ideal in $R$ and the conclusion will follow if we prove that $\mathfrak{a} = 0$. Let $a_1 \in \mathfrak{a}$, say $f = a_1 t^n + \cdots \in J$. Then $ft \in J$ and so there exists $g \in R[t]$ such that $(1+g)(1-ft) = 1$, i.e.

$$g = ft + gft = ft + f^2t^2 + gf^2t^2 = \cdots = ft + f^2t^2 + \cdots + f^rt^r + gf^rt^r,$$

for all $r \geqslant 1$. Hence we obtain

$$(1+g)(1-f^r t^r) = 1 + ft + \cdots + f^{r-1}t^{r-1}.$$

Let us take $r > \deg g$ and equate the coefficients of terms of degree $\geqslant r(n+1)$. On the right there is no contribution while on the left we have $(1+g)a_1^r$, therefore $a_1^r = 0$. Thus $\mathfrak{a}$ is a nil ideal, hence $\mathfrak{a} = 0$ and so $J = 0$, as we had to show. ■

We can now also prove the basic theorem on PI-algebras:

THEOREM 4.6 (Kaplansky's theorem 1948) *Let $R$ be a primitive PI-algebra, with a polynomial identity of degree $d$. Then $R$ is a simple algebra of finite dimension $n^2$ over its centre, where $n \leqslant d/2$. More precisely, if $V$ is a simple faithful $R$-module and $D = \mathrm{End}_R(V)$, then $R \cong \mathfrak{M}_m(D)$, where $m = [V:D]$.*

*Proof.* Let $p$ be a polynomial of degree $d$ which vanishes on $R$; by Prop. 9.5.3 we may take $p$ to be multilinear. Since $R$ is primitive, there is a simple faithful $R$-module $V$; we identify $R$ with its image in $\mathrm{End}(V)$ and put $D = \mathrm{End}_R(V)$. By Schur's lemma $D$ is a skew field, and by the density theorem (Th. 2.3) $R$ is dense in $\mathrm{End}_D(V)$. If $[V:D]$ is finite, we have $R = \mathrm{End}_D(V)$; otherwise, by Prop. 3.4 we can for any $m$, find a subring of $R$ mapping onto $D_m$. So $D_m$ again satisfies $p = 0$, therefore $d \geqslant 2m$, by the staircase lemma (Lemma 9.5.9) and we have a bound on $m$. It follows that $[V:D] = m \leqslant d/2$ and $R = \mathrm{End}_D(V) = D_m$.

In particular, this shows that $R$ is simple and its centre is the centre of $D$, a field $C$. Let $K$ be a maximal commutative subfield of $D$ containing $C$; we have a natural homomorphism $R \otimes_C K \to RK$, but the left-hand side is simple, hence $R \otimes_C K \cong RK$. Now $RK$ is a $K$-algebra, its centre is $K$ and it acts densely on $V$, therefore $RK \cong K_n$, where again $n \leqslant d/2$. Moreover, $[R:C] = [RK:K] = n^2$. ■

We record separately the special case of a skew field:

COROLLARY 4.7 *A skew field which satisfies a polynomial identity of degree $d$ is of finite dimension $\leqslant [d/2]^2$ over its centre.* ■

Now the main result can be extended to semiprimitive rings as follows. We note that any ring $R$ can be embedded in the full matrix ring $R_d$ (e.g. as scalar matrices), hence $R_m$ can be embedded in $R_n$ whenever $m \mid n$.

COROLLARY 4.8 *Any semiprimitive PI-algebra satisfying an identity of degree $d$ can be embedded in a matrix algebra $\mathfrak{M}_r(E)$ over a commutative ring $E$, where $r \leqslant [d/2]!$.*

*Proof.* By Th. 4.1, $R$ is a subdirect product of primitive rings $R_\lambda$, where $R_\lambda$

is a homomorphic image of $R$ and hence again satisfies an identity of degree $d$. By Th. 4.6, $R_\lambda$ is a simple algebra of degree $n_\lambda$ over its centre, where $n_\lambda \leqslant d/2$. By taking a splitting field $E_\lambda$ we can thus embed $R_\lambda$ in a matrix algebra $\mathfrak{M}_{n_\lambda}(E_\lambda)$. The least common multiple of the degrees $n_\lambda$ which occur is $r \leqslant [d/2]!$ and $R_\lambda$ can also be embedded in $\mathfrak{M}_r(E_\lambda)$. Now $\prod \mathfrak{M}_r(E_\lambda) \cong \mathfrak{M}_r(E)$, where $E = \prod E_\lambda$ and so we have an embedding of $R$ in $\mathfrak{M}_r(E)$. ■

If $R$ is a PI-algebra without non-zero nil ideals, then $R[t]$ is semiprimitive, by Prop. 4.5 and is again a PI-algebra, by Cor. 9.5.4, hence we obtain

THEOREM 4.9 *Let $R$ be a PI-algebra without non-zero nil ideals. Then $R$ can be embedded in $\mathfrak{M}_n(E)$, where $E$ is a commutative ring, and if $R$ satisfies an identity of degree $d$, then $n \leqslant [d/2]!$.* ■

For Artinian rings 'semiprimitive' reduces to 'semisimple'; this follows from results proved in A.2, Th. 5.4.5, that for an Artinian ring $R$, $R/\mathbf{J}(R)$ is semisimple, and it will be derived in a more general context below in **10.5**. In the Artinian case the radical may also be described as the intersection of all maximal two-sided ideals. This does not hold generally: we clearly have, for any ring $R$,

$$\bigcap\{\text{max.left ideals}\} = \bigcap\{\text{max.right ideals}\} \subseteq \bigcap\{\text{max.ideals}\}, \tag{4}$$

where the first equality holds by the characterization of $\mathbf{J}(R)$. For an example where the inclusion is strict, take $R = \mathrm{End}_K(V)$, where $V$ is an infinite-dimensional vector space over a field $K$. We have seen in **10.3** that $R$ is primitive, hence $\mathbf{J}(R) = 0$, but the ideals of $R$ form a chain, so the intersection on the right of (4) is the unique maximal ideal of $R$. Let us denote by $\mathfrak{m}_0$ the socle of $R$; in the representation on $V$ this corresponds to the set of elements of finite rank. Assuming $[V:K]$ to be countable, with basis $\{e_i\}$, let us write $\mathfrak{a}_i$ for the set of elements of $R$ mapping $e_i$ to 0. Then $\mathfrak{a}_i$ is a maximal right ideal and $\mathfrak{a}_i \not\supseteq \mathfrak{m}_0$, $\bigcap \mathfrak{a}_i = 0$. Similarly, if $\mathfrak{b}_i$ is the set of elements of $R$ mapping $V$ into $\sum_{i \neq j} Ke_j$, then $\mathfrak{b}_i$ is a maximal left ideal such that $\mathfrak{b}_i \not\supseteq \mathfrak{m}_0$, $\bigcap \mathfrak{b}_i = 0$. By Krull's theorem $R$ also has maximal right ideals (and maximal left ideals) containing $\mathfrak{m}_0$, but the proof is non-constructive and there is no obvious procedure for finding them.

## Exercises

(1) Show that a subdirect product of semiprimitive rings is semiprimitive.

(2) Show that the radical contains no non-zero idempotent. Deduce that every regular ring is semiprimitive.

(3) Verify that in an Artinian ring the intersection of all maximal two-sided ideals is just the radical.

(4) Let $R$ be a ring and $\mathfrak{a}$ an ideal in $R$. Show that if $\mathbf{J}(R/\mathfrak{a}) = 0$, then $\mathbf{J}(R) \subseteq \mathfrak{a}$.

(5) Show that a subdirect product of a finite number of simple rings is a direct product of simple rings.

## 10.5 Algebras without a unit-element

So far we have taken the existence of a unit-element or 'one' as part of the definition of a ring, but there are some occasions when a 'ring without a one' arises naturally. For example, the algebra $C(X)$ of all continuous functions on a topological space $X$ has a one precisely when $X$ is compact. We shall maintain the convention that a ring necessarily has a one, and allow for the case where a one is lacking by speaking of an *algebra*. The coefficient ring $K$ (with 1) may be any commutative ring; this is no restriction since every ring may be regarded as a $\mathbf{Z}$-algebra.

Let $A$ be a $K$-algebra; by a right $A$-module $M$ we understand a $(K, A)$-bimodule with the rule $\alpha(xa) = x(\alpha a)$ for all $x \in M$, $\alpha \in K$, $a \in A$. Even if $A$ has a one, $e$ say, this need not define the identity mapping on $M$. If it does, i.e. if

$$xe = x \qquad \text{for all } x \in M,$$

the module $M$ is said to be *unital*.

Our first observation is that a $K$-algebra may always be embedded in a $K$-algebra with one. This may be done in many ways; we shall single out one by recalling the following definition from **5.1** of A.2. A $K$-algebra $A$ with a one is said to be *augmented* if there exists is a $K$-algebra homomorphism, called the *augmentation mapping*:

$$\varepsilon : A \to K.$$

This means that $\varepsilon$ is a ring homomorphism such that $(\alpha a)\varepsilon = \alpha(a\varepsilon)$ for all $\alpha \in K$, $a \in A$. In particular, $(\alpha e)\varepsilon = \alpha$, hence $\alpha e \mapsto \alpha$ is a bijection between $K \cdot e$ (where $e$ is the one of $A$) and $K$, and we may embed $K$ in $A$ by identifying $\alpha$ with $\alpha e$. The kernel of $\varepsilon$ is the augmentation ideal of $A$ and we have the direct sum decomposition

$$A = \ker \varepsilon \oplus K. \tag{1}$$

For any $x \in A$ may be written in the form

$$x = (x - x\varepsilon) + x\varepsilon.$$

We recall from A.2, Th. 5.1.1, that for any $K$-algebra $A$ we can form the augmented algebra $A^1 = A \oplus K$ with multiplication

$$(a, \alpha)(b, \beta) = (ab + a\beta + \alpha b, \alpha\beta),$$

and this algebra has an augmentation ideal isomorphic to $A$. Conversely, if $C$

is an augmented $K$-algebra with augmentation ideal $A$, then $A^1 \cong C$, as augmented $K$-algebras. Moreover, the category of $A$-modules is equivalent to the category of unital $A^1$-modules.

For $K$-algebras the notion of simple module has to be modified. A right module $M$ over a $K$-algebra $A$ is said to be *simple* if $MA \neq 0$ and $M$ has no submodules other than 0 and $M$. When $A$ has a one and acts unitally, this reduces to the previous definition. Our object will be to study simple $A$-modules in terms of $A^1$ (where the results of **10.4** can be used).

Any simple $A$-module $M$ can again be represented as $A/I$ for a maximal right ideal $I$, but we must also have $A^2 \not\subseteq I$, to ensure that $MA \neq 0$. Further, we can no longer use Krull's theorem to find maximal right ideals because $A$ may not be finitely generated as right $A$-module. To overcome these difficulties, let us look more closely at the correspondence between right ideals in $A$ and in $A^1$.

Let $I$ be a right ideal of $A$ and $I'$ a right ideal of $A^1$ such that $I' \supseteq I$. Then $I \subseteq I' \cap A$, so there is a natural homomorphism of $A$-modules:

$$A/I \to A/(I' \cap A) \cong (A + I')/I' \subseteq A^1/I'.$$

This is an isomorphism iff $I' \cap A = I$ and $I' + A = A^1$. Given any right ideal $I'$ of $A^1$ such that $I' + A = A^1$, we can put $I = I' \cap A$; then by what has just been said, $A/I \cong A^1/I'$. In particular, this holds for any maximal right ideal $I'$ of $A'$ which does not contain $A$. If we start from $I$ in $A$, the next lemma shows under what conditions we can find a suitable $I'$.

LEMMA 5.1 *Let $A$ be a $K$-algebra with a right ideal $I$. Then there is a right ideal $I'$ of $A^1$ such that*

$$I' + A = A^1, \qquad I' \cap A = I, \tag{2}$$

*if and only if $A$ contains an element $e$ such that*

$$(1 - e)A \subseteq I. \tag{3}$$

In (3), 1 is the one of $A^1$, but we can express (3) entirely within $A$ by writing $a - ea \in I$ for all $a \in A$. A right ideal $I$ satisfying (3) for some $e \in A$ is called a *modular* right ideal.

*Proof.* If (2) holds, we can write $1 = u + e$, where $u \in I'$, $e \in A$. Then $(1 - e)A = uA \subseteq I' \cap A = I$, and (3) follows. Conversely, given (3), we put $I' = I + (1 - e)K$. Then $I'$ is a right ideal in $A^1$, by (3), and $I' + A$ is a right ideal containing $(1 - e) + e = 1$, hence $I' + A = A^1$. Moreover, if $x = a + (1 - e)\alpha \in A$, where $\alpha \in K$, $a \in I$, then $\alpha = 0$, hence $x \in I$, so $I' \cap A = I$. ■

This lemma shows in particular that for any maximal right ideal $I'$ of $A^1$ which does not contain $A$, $I' \cap A$ is a modular right ideal of $A$.

By a *maximal modular right ideal* of $A$ we shall understand a maximal member of the set of all proper modular right ideals. Any right ideal containing a modular right ideal is again modular (by the definition), hence a maximal modular right ideal is also maximal in the set of all proper right ideals. Moreover, any proper modular right ideal is contained in a maximal modular right ideal, for given $I \supseteq (1-e)A$, we can by Krull's theorem find a right ideal containing $I$ but not $e$, and maximal with these properties, and this is easily seen to be maximal modular. In fact the notion of modular right ideal may be regarded as a device for producing maximal right ideals in algebras without a one; the corresponding quotients are simple modules, as we shall see below.

In any $K$-algebra $A$ let us define the *Jacobson radical* $\mathbf{J}(A)$ as the set of elements of $A$ represented by 0 in any simple $A$-module. If $\mathbf{J}(A) = 0$, $A$ is said to be *semiprimitive*. For algebras with a one this reduces to the previous definition, by the characterization quoted in **10.4**. Now $\mathbf{J}(A)$ can be described as follows in terms of the maximal modular right ideals of $A$:

THEOREM 5.2 *Let $A$ be a $K$-algebra, where $K$ is a semiprimitive coefficient ring, and denote by $A^1$ the corresponding augmented $K$-algebra. Then $\mathbf{J}(A) = \mathbf{J}(A^1)$, and $\mathbf{J}(A)$ is the intersection of all the maximal modular right ideals of $A$.*

*Proof.* It is clear that a right $A$-module $M$ such that $MA \neq 0$ is simple iff it is simple as $A^1$-module. Thus for any $a \in A$,

$$\begin{aligned} a \in \mathbf{J}(A) &\Leftrightarrow a \text{ is represented by 0 in every simple } A\text{-module,} \\ &\Leftrightarrow a \text{ is represented by 0 in every simple } A^1\text{-module,} \\ &\Leftrightarrow a \in \mathbf{J}(A^1) \cap A. \end{aligned}$$

It follows that $\mathbf{J}(A) = \mathbf{J}(A^1) \cap A$. Now assume that $x \in \mathbf{J}(A^1) \setminus A$; then

$$x = a + \alpha \cdot 1, \qquad \text{where } a \in A,\ \alpha \in K,\ \alpha \neq 0. \tag{4}$$

Since $K$ is semiprimitive, there is a maximal ideal $\mathfrak{m}$ of $K$ not containing $\alpha$. The residue class field $K/\mathfrak{m}$ has a natural $K$-module structure (by multiplication) and we can define an $A^1$-module structure by the rule $u \cdot y = u(y\varepsilon)$, for $u \in K/\mathfrak{m}$, $y \in A^1$. Then $K/\mathfrak{m}$ is a simple $A^1$-module, and if $\bar{1}$ is the residue class of 1, then by applying the element (4), we get $\bar{1} \cdot x = \bar{1} \cdot \alpha \neq 0$. This contradicts the fact that $x \in \mathbf{J}(A^1)$; hence $\mathbf{J}(A^1) \subseteq A$, and it follows that $\mathbf{J}(A^1) = \mathbf{J}(A)$. Now

$$\begin{aligned} a \in \mathbf{J}(A^1) &\Leftrightarrow a \in \text{each maximal right ideal of } A^1, \\ &\Leftrightarrow a \in \text{each maximal right ideal of } A^1 \text{ which does not contain } A, \\ &\Leftrightarrow a \in \text{each maximal modular right ideal of } A, \end{aligned}$$

by the remarks preceding the theorem. ■

From the definition of the Jacobson radical quoted in **10.4** we now obtain

COROLLARY 5.3 *Let $A$ be a $K$-algebra, where $K$ is semiprimitive. Then $\mathbf{J}(A)$ consists of all elements $a \in A$ such that $ay$ has a quasi-inverse, for all $y \in A$.* ■

We note that the intersection of all the maximal right ideals of $A$ is in general different from $\mathbf{J}(A)$. For example, let $k$ be a field and $A$ the algebra of all formal power series in $x$ with zero constant term. Then $\mathbf{J}(A) = A$, but the intersection of all maximal ideals is $xA$. More generally, simple rings have been constructed which coincide with their Jacobson radical, by Sąsiada 1961 (cf. Sąsiada–Cohn (1968)).

We can now prove the Wedderburn structure theorem for semisimple rings under weaker hypotheses. First an auxiliary result on modular right ideals; we recall that two right ideals $\mathfrak{a}, \mathfrak{b}$ of $A$ are called *comaximal* if $\mathfrak{a} + \mathfrak{b} = A$.

LEMMA 5.4 *Let $A$ be any $K$-algebra and $\mathfrak{a}, \mathfrak{b}$ modular right ideals which are comaximal. Then $\mathfrak{a} \cap \mathfrak{b}$ is again modular.*

*Proof.* By hypothesis there exist $e, f \in A$ such that $(1-e)A \subseteq \mathfrak{a}$, $(1-f)A \subseteq \mathfrak{b}$. Since $\mathfrak{a} + \mathfrak{b} = A$, by comaximality, we can write

$$e = a_1 + b_1, \qquad \text{where } a_1 \in \mathfrak{a},\ b_1 \in \mathfrak{b},$$
$$f = a_2 + b_2, \qquad \text{where } a_2 \in \mathfrak{a},\ b_2 \in \mathfrak{b}.$$

Hence for any $x \in A$,

$$(a_2 + b_1)x \equiv b_1 x \equiv ex \equiv x \pmod{\mathfrak{a}},$$
$$(a_2 + b_1)x \equiv a_2 x \equiv fx \equiv x \pmod{\mathfrak{b}}.$$

Therefore $(1 - (a_2 + b_1))A \subseteq \mathfrak{a} \cap \mathfrak{b}$ and the conclusion follows. ■

THEOREM 5.5 (Wedderburn decomposition theorem) *Let $A$ be a semiprimitive $K$-algebra over a semiprimitive coefficient ring $K$, and assume that $A$ is right Artinian. Then $A$ has a one and is semisimple, as ring.*

*Proof.* Let $\{I_\lambda\}$ be the family of all maximal modular right ideals of $A$. Since $A$ is semiprimitive, $\bigcap I_\lambda = 0$, and since $A$ is right Artinian, the chain $I_1 \supseteq I_1 \cap I_2 \supseteq \cdots$ breaks off, hence for some $r$,

$$I_1 \cap \cdots \cap I_r = 0. \tag{5}$$

By omitting superfluous terms, we can choose this representation to be irredundant, i.e. such that $I_i \not\supseteq I_1 \cap \cdots \cap I_{i-1}$. Since $I_i$ is maximal modular, it is comaximal with $I_1 \cap \cdots \cap I_{i-1}$, so by Lemma 5.4, 0 is modular, i.e. $(1-e)A = 0$ for some $e \in A$. Hence $A(1-e)$ is a nilpotent left ideal, which must be 0, by semiprimitivity, so $x = xe = ex$ for all $x \in A$ and $e$ is the one of $A$.

Since each $I_i$ is a maximal right ideal, $P_i = A/I_i$ is a simple right $A$-module and $A$ is a submodule of the direct sum $\oplus P_i$. As submodule of a semisimple module, $A$ itself is semisimple as right $A$-module, hence it is semisimple as a ring. ■

The generalization of Wedderburn–Artin theory set forth in **10.3–5** was developed by Jacobson in 1945; the density theorem generalizes a theorem of Burnside to the effect that a monoid acting irreducibly on an $n$-dimensional vector space over an algebraically closed field contains $n^2$ linearly independent endomorphisms.

### Exercises

(1) Show that a non-zero ideal in a primitive ring is again primitive (as algebra without 1).

(2) Let $A$ be an algebra without nilpotent (two-sided) non-zero ideal. Show that $A$ has no nilpotent non-zero left or right ideals. Deduce that for any (non-zero) idempotent $e$ in $A$, $eA$ is a minimal right ideal iff $eAe$ is a skew field.

(3) Let $A$ be a nil algebra (i.e. every element is nilpotent). Show that every maximal right ideal is twosided and contains $A^2$.

(4) Let $A$ be an algebra over a field $k$; show that if $A$ has no zero-divisors, then $A^1$ is an integral domain iff 0 is modular, as right ideal. By considering End($A$), show that $A$ can be embedded in an integral domain.

(5) Let $k$ be a field and $A$ the $k$-subalgebra of the free algebra $k\langle x, y\rangle$ consisting of all polynomials with zero constant term. Find two modular right ideals of $A$ whose intersection is not modular.

## 10.6 Semiprime rings and nilradicals

We have already briefly met prime and semiprime rings and now come to examine the relation between them. We recall that a ring $R$ is said to be *prime* if $R \neq 0$ and for any two-sided ideals $\mathfrak{a}, \mathfrak{b}$ of $R$,

$$\mathfrak{ab} = 0 \Rightarrow \mathfrak{a} = 0 \quad \text{or} \quad \mathfrak{b} = 0.$$

Equivalently, for any $a, b \in R$, $aRb = 0$ implies $a = 0$ or $b = 0$. In the commutative case a prime ring is just an integral domain. In general every integral domain is prime, but not conversely, as is shown by simple rings. More generally we have

PROPOSITION 6.1 *Every primitive ring is prime.*

*Proof.* Let $R$ be primitive and take a simple faithful $R$-module $M$. If $\mathfrak{a}, \mathfrak{b}$ are

non-zero ideals in $R$, then $M\mathfrak{a} \neq 0$, hence $M\mathfrak{a} = M$ by the simplicity of $M$; likewise $M\mathfrak{b} = M$, so $M\mathfrak{ab} = (M\mathfrak{a})\mathfrak{b} = M\mathfrak{b} = M$; this shows that $\mathfrak{ab} \neq 0$. ■

Since the notion 'prime' is left-right symmetric, a similar result holds for left primitive rings. On the other hand, the converse is false, as we see already in the commutative case, when primitive rings are fields, whereas prime rings are integral domains.

By a *prime ideal* $\mathfrak{p}$ in a ring $R$ we understand a two-sided ideal $\mathfrak{p}$ such that $R/\mathfrak{p}$ is a prime ring; for commutative rings this reduces to the definition given in **9.2** of A.2. We observe that a prime ideal must always be proper.

There is a method of constructing prime ideals, rather as in the commutative case; the notion of multiplicative set has to be replaced here by that of an $m$-system. By this term one understands a subset $M$ of $R$ such that $1 \in M$ and if $a, b \in M$, then $axb \in M$ for some $x \in R$. We note that an ideal is prime iff its complement in $R$ is an $m$-system. Now there is an analogue of Th. 9.2.2 of A.2.

THEOREM 6.2 *Let $R$ be a ring, $M$ an $m$-system in $R$ and $\mathfrak{a}$ an ideal of $R$ disjoint from $M$. Then there exists an ideal $\mathfrak{p}$ in $R$ which contains $\mathfrak{a}$, is disjoint from $M$ and is maximal with respect to these properties. Any such ideal $\mathfrak{p}$ is prime.*

*Proof.* Let $\mathscr{A}$ be the set of all ideals $\mathfrak{a}'$ of $R$ such that $\mathfrak{a}' \supseteq \mathfrak{a}$, $\mathfrak{a}' \cap M = \varnothing$. Then $\mathfrak{a} \in \mathscr{A}$, so $\mathscr{A}$ is not empty. It is easily seen to be inductive and so by Zorn's lemma, it contains a maximal member $\mathfrak{p}$, which is an ideal with the required properties.

Now let $\mathfrak{p}$ be an ideal with the properties stated and assume that $\mathfrak{a}, \mathfrak{b} \not\subseteq \mathfrak{p}$, $\mathfrak{ab} \subseteq \mathfrak{p}$. Then $\mathfrak{a} + \mathfrak{p}$, $\mathfrak{b} + \mathfrak{p}$ are strictly larger than $\mathfrak{p}$ and so must meet $M$, say $s = p + a$, $t = q + b \in M$, where $p, q \in \mathfrak{p}$, $a \in \mathfrak{a}$, $b \in \mathfrak{b}$. By hypothesis there exists $x \in R$ such that $sxt \in M$; thus $M$ contains

$$(p + a)x(q + b) = px(q + b) + axq + axb \in \mathfrak{p} + \mathfrak{ab} = \mathfrak{p},$$

a contradiction. Hence $\mathfrak{a}, \mathfrak{b} \not\subseteq \mathfrak{p}$ implies $\mathfrak{ab} \not\subseteq \mathfrak{p}$, and clearly $1 \notin \mathfrak{p}$, so $\mathfrak{p}$ is indeed prime. ■

In the commutative case we found that the intersection of all prime ideals of $R$ is the set of all nilpotent elements of $R$. Th. 6.2 can be used to obtain a corresponding result for general rings; however, the situation is a little more complicated here because the ideal generated by a nilpotent element need not be nilpotent. Let us recall that an ideal $\mathfrak{a}$ is called *nilpotent* if $\mathfrak{a}^r = 0$ for some $r \geqslant 1$, i.e. $x_1 \cdots x_r = 0$ for any $x_i \in \mathfrak{a}$, and $\mathfrak{a}$ is *nil* if it consists of nilpotent elements. Clearly every nilpotent ideal is nil, but the converse does not hold generally.

We also recall that a ring $R$ is called *semiprime* if for any two-sided ideal $\mathfrak{a}$ of $R$,

$$\mathfrak{a}^2 = 0 \Rightarrow \mathfrak{a} = 0. \tag{1}$$

In terms of elements this states that $aRa = 0$ implies $a = 0$, for all $a \in R$. We observe that a ring is semiprime iff it has no non-zero nilpotent ideals. For when this holds, (1) is clearly satisfied. Conversely, assume (1) and let $\mathfrak{a}$ be a nilpotent ideal, say $\mathfrak{a}^{r-1} \neq 0$, $\mathfrak{a}^r = 0$, where $r \geqslant 2$. Then $2r - 2 \geqslant r$, hence $0 = \mathfrak{a}^{2r-2} = (\mathfrak{a}^{r-1})^2 \neq 0$, by (1), and this is a contradiction.

In the commutative case the semiprime rings are the *reduced* rings, i.e. rings without nilpotent elements not equal to 0, and in the general case every reduced ring is semiprime, but not conversely.

An ideal $\mathfrak{a}$ in a ring $R$ is said to be *semiprime* if $R/\mathfrak{a}$ is a semiprime ring. We note that $R$ itself, as an ideal in $R$, is semiprime but not prime.

We first elucidate the relation between prime and semiprime ideals.

PROPOSITION 6.3 *Let $R$ be a ring. Then every intersection of prime ideals is semiprime; conversely every semiprime ideal is an intersection of prime ideals.*

*Proof.* Let $\mathfrak{c} = \bigcap \mathfrak{p}_\lambda$, where the $\mathfrak{p}_\lambda$ are prime ideals, and suppose that $aRa \subseteq \mathfrak{c}$. Then $aRa \subseteq \mathfrak{p}_\lambda$, hence $a \in \mathfrak{p}_\lambda$ for all $\lambda$, and so $a \in \cap p_\lambda = \mathfrak{c}$, and this shows $\mathfrak{c}$ to be semiprime. Conversely, let $\mathfrak{c}$ be a semiprime ideal in $R$; by passing to the residue class ring $R/\mathfrak{c}$, we may take $\mathfrak{c} = 0$. We have to show that in a semiprime ring the intersection of all prime ideals is 0. Take any $a \in R^\times$; then $aRa \neq 0$, so there exists $b_0$ such that $a_1 = ab_0a \neq 0$. Generally, if we have constructed $a_1, \ldots, a_n$ such that $a_{i+1} \in a_iRa_i$ and $a_n \neq 0$, then there exists $b_n$ such that $a_{n+1} = a_nb_na_n \neq 0$. The set $M = \{1, a_0 = a, a_1, a_2, \ldots\}$ is an $m$-system, for given $a_r, a_s$, choose $n > r, s$; then $a_r, a_s$ are factors of $a_n$ and $a_nb_na_n \neq 0$, hence $a_rua_s \neq 0$ for some $u \in R$. Thus $M$ is an $m$-system; moreover, $a \in M$, $0 \notin M$, hence we can find a prime ideal $\mathfrak{p}$ disjoint from $M$, by Th. 6.2. It follows that $a \notin \mathfrak{p}$; since $a$ was any non-zero element of $R$, we see that the intersection of all prime ideals of $R$ is zero. ■

COROLLARY 6.4 *A ring is semiprime if and only if the intersection of all its prime ideals is* 0. *Hence any semiprime ring $R$ can be written as a subdirect product of prime rings, which are homomorphic images of $R$.*

*Proof.* The first part follows by applying Prop. 6.3 to the zero ideal, and the second part now follows as in the proof of Th. 4.1. ■

Just as semiprimitivity leads to the Jacobson radical, so there is a type of radical arising from semiprimeness, but it is not as clear cut and in general there is more than one such radical. Let us define a *nilradical* in a ring $R$ as an ideal $\mathfrak{N}$ which is nil and such that $R/\mathfrak{N}$ is semiprime. A ring may have more than one nilradical; in what follows we shall describe the greatest and least such radical. We begin with a lemma.

LEMMA 6.5 *The sum of any family of nil ideals in a ring is a nil ideal.*

*Proof.* Consider first two nil ideals $\mathfrak{a}_1$, $\mathfrak{a}_2$ and write $\mathfrak{a} = \mathfrak{a}_1 + \mathfrak{a}_2$. Then the ideal $\mathfrak{a}/\mathfrak{a}_2$ of the ring $R/\mathfrak{a}_2$ is nil, because $\mathfrak{a}/\mathfrak{a}_2 \cong \mathfrak{a}_1/(\mathfrak{a}_1 \cap \mathfrak{a}_2)$ and the latter is a homomorphic image of $\mathfrak{a}_1$ and hence nil. Thus any element of $\mathfrak{a}$ has a power in $\mathfrak{a}_2$ and a power of this is 0, hence $\mathfrak{a}$ is nil. Now an induction shows that the sum of any finite number of nil ideals is nil. In the general case let $\mathfrak{a} = \sum \mathfrak{a}_\lambda$, where each $\mathfrak{a}_\lambda$ is a nil ideal. Any element of $\mathfrak{a}$ lies in the sum of a finite number of the $\mathfrak{a}_\lambda$ and hence is nilpotent, therefore $\mathfrak{a}$ is nil. ■

By this lemma the sum of all nil ideals in a ring $R$ is a nil ideal; it is denoted by $U = U(R)$ and called the (*Baer*) *upper nilradical.* It is indeed a nilradical, for $R/U$ cannot have any non-zero nil ideals, by the maximality of $U$, and *a fortiori* must be semiprime. Since $U$ contains all nil ideals of $R$, it contains all nilradicals.

To obtain the least nilradical we need another definition. An element $c$ of $R$ is said to be *strongly nilpotent* if any sequence $c_1 = c$, $c_2$, $c_3, \ldots$ such that $c_{n+1} \in c_n R c_n$ is ultimately zero. It is clear that such an element is nilpotent and that any element of a nilpotent (left or right) ideal is strongly nilpotent. Moreover, in a semiprime ring $R$, the only strongly nilpotent element is 0. For if $c \neq 0$, then $cRc \neq 0$, say $c_2 = cac \neq 0$, now $c_3 = c_2 a' c_2 \neq 0$ for some $a' \in R$, and continuing in this way we obtain a sequence $c_1 = c$, $c_2, \ldots$ which is never zero and so $c$ is not strongly nilpotent.

Now the least nilradical may be described as follows:

THEOREM 6.6 *In any ring $R$ the set $L(R)$ of all strongly nilpotent elements is the least nilradical, and it is equal to the intersection of all prime ideals in $R$:*

$$L(R) = \bigcap \{\mathfrak{p} \mid \mathfrak{p} \text{ prime in } R\}. \tag{2}$$

*Proof.* If $x$ is strongly nilpotent in $R$, so is its residue class $\bar{x} (\mathrm{mod}\, \mathfrak{p})$ for any prime ideal $\mathfrak{p}$, so $\bar{x} = 0$, which means that $x \in \mathfrak{p}$. Thus $L(R)$ is contained in the right-hand side of (2). Now suppose that $c \notin L(R)$; then $c$ is not strongly nilpotent, so there exists a sequence $\{c_n\}$ such that $c_1 = c$, $0 \neq c_{n+1} \in c_n R c_n$. The set $S = \{1, c_1, c_2, \ldots\}$ is an $m$-system, for given $c_m$, $c_n$, where $m \leqslant n$ say, we have $c_{n+1} \in c_n R c_n \cap c_m R c_m$. By Th. 6.2 there is a maximal ideal $\mathfrak{p}$ disjoint from $S$, and $\mathfrak{p}$ is prime, thus $c \notin \mathfrak{p}$ and this shows that equality holds in (2), and incidentally, that $L(R)$ is indeed an ideal.

Now $L(R)$ is clearly a nil ideal and $R/L(R)$ is semiprime, hence $L(R)$ is a nilradical, and from (2) it is the least ideal with semiprime quotient, hence it is the least nilradical. ■

$L(r)$ is called the *prime radical* or also the (*Baer*) *lower nilradical.*

If a ring $R$ has a non-zero nilpotent right ideal $\mathfrak{a}$, then $R\mathfrak{a}$ is a two-sided ideal, again nilpotent, since $(R\mathfrak{a})^n \subseteq R\mathfrak{a}^n$. The following conjecture, raised over 60 years ago, is still unanswered:

Köthe's conjecture. *If a ring has a non-zero nil right ideal, then it has a non-zero nil ideal.*

Equivalently, the prime radical contains every nil right ideal. We remark that in Noetherian rings every nil right ideal is nilpotent, so the conjecture is valid in that case.

In general rings the upper and lower nilradical may well be distinct; to give an example it is clearly enough to construct a semiprime ring with a non-zero nil ideal. Let $A$ be the $k$-algebra (without 1) generated by $x_1, x_2, \ldots$ with the following defining relations: for any element $a$ involving only $x_1, x_2, \ldots, x_n$ we have $a^n = 0$. Then every element of $A$ is nilpotent, and in the augmented algebra $R = A^1$ we have the nil ideal $A$. However, $R$ is semiprime, for given $a \in R^\times$, let $a$ be of degree $m$ in the $x$'s and put $N = 2m + 2$. Any relation involving $x_N$ consists of terms of degree at least $N$, hence $ax_N a \neq 0$, because this expression has only terms of degree at most $2m + 1$. This shows $R$ to be semiprime.

For another example, this time finitely generated, take the finitely generated non-nilpotent nil algebra $A$ constructed by Golod (cf. Ex. (5) of **2.3**). It can be verified that $A^1$ is semiprime but it has the nil ideal $A$.

However, in a right Noetherian ring the upper and lower nilradical coincide. For as we saw in Prop. 9.4.3, in a semiprime ring with maximum condition on right annihilators every nil left or right ideal is 0. Hence if $R$ is right Noetherian, then $R/L(R)$ has no non-zero nil ideals and so $U(R) = L(R)$; by symmetry the same holds for left Noetherian rings:

PROPOSITION 6.7 *In a right (or left) Noetherian ring the upper and lower nilradical coincide. Thus in a Noetherian ring the prime radical is a nilpotent ideal containing all nilpotent ideals.* ■

Sometimes a radical intermediate between $U$ and $L$ is considered, which is defined as follows. An algebra $A$ is said to be *locally nilpotent* if the subalgebra generated by any finite subset is nilpotent. It is clear that for any ideal we have the implications: nilpotent $\Rightarrow$ locally nilpotent $\Rightarrow$ nil. Moreover, it is not hard to show that the sum of any number of locally nilpotent ideals is again locally nilpotent. Now the *Levitzki nilradical* $N$ of a ring $R$ is defined as the sum of all locally nilpotent ideals of $R$. By what has been said, $N$ is locally nilpotent, hence nil, and it can be shown that $R/N$ has zero Levitzki nilradical, so it is semiprime, and this shows $N$ to be indeed a nilradical. In any ring we have

$$U \supseteq N \supseteq L,$$

and in the first example given above $N \supset L$, at least in characteristic 0, by the Nagata–Higman theorem (cf. Ex. (12) of Ch. 9), while $U \supset N$ in Golod's example.

## Exercises

(1) Show that in a right Noetherian ring 0 can be expressed as a product of prime ideals.

(2) Show that a prime ring with a minimal right ideal is primitive.

(3) Show that in any semiprime ring the socle and the left socle coincide. (Hint: use Ex. (2) of **10.5**)

(4) Let $R$ be the subring of $\mathfrak{M}_2(\mathbf{Z})$ consisting of all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $a \equiv d$, $b \equiv c \pmod 2$. Show that $R$ is a prime ring, but not an integral domain; find all its idempotents.

(5) Show that the sum of any family of locally nilpotent ideals is locally nilpotent. (Hint: imitate the proof of Lemma 6.5.)

(6) In any ring $R$ with Levitzki nilradical $N$, verify that $R/N$ has zero Levitzki nilradical.

(7) In any ring $R$, define $N_1$ as the sum of all nilpotent ideals, and define $N_\alpha$ for any ordinal $\alpha$ inductively by $N_{\alpha+1}/N_\alpha = N_1(R/N_\alpha)$, while at a limit ordinal, $N_\alpha = \bigcup_{\beta<\alpha} N_\beta$. Show that the union of all $N_\alpha$ is the lower nilradical of $R$.

(8) Show that in a left or right Noetherian ring any nilpotent element is strongly nilpotent. Deduce that every nil (left or right) ideal is nilpotent.

(9) Show that a reduced prime ring is an integral domain.

(10) Show that in a reduced ring $R$, if a product of $n$ elements in a certain order is 0, then the product in any order is 0. (Hint: show that if $a_1 \cdots a_n = 0$, then $a_1 x_1 a_2 x_2 \cdots a_n x_n = 0$, for all $x_i \in R$.)

(11) Show that in any ring $R$, every prime ideal contains a minimal prime ideal. (Hint: take a maximal $m$-system containing the complement of the given prime ideal.)

(12) Let $R$ be a reduced ring and $\mathfrak{p}$ a minimal prime ideal. Show that the monoid $M$ generated by the complement of $\mathfrak{p}$ does not contain 0. Deduce that $M$ does not meet $\mathfrak{p}$ and hence that $R/\mathfrak{p}$ is an integral domain.

(13) Show that every reduced ring is a subdirect product of integral domains. (Hint: use Ex. (9)–(12). This is a theorem of Andrunakievich–Ryabukhin; the proof is due to Herstein.)

(14) Show that any ideal of a semiprime ring is semiprime, qua algebra without 1.

## 10.7 Prime PI-algebras

In some respects the existence of a polynomial identity has effects similar to the maximum condition and we shall in this section obtain Posner's theorem, which is an analogue of Goldie's theorem for PI-rings. The first step is to form fractions with respect to a central Ore set.

THEOREM 7.1 *Let $R$ be a semiprimitive PI-algebra with centre $C$. Then every non-zero ideal $\mathfrak{a}$ of $R$ meets $C$ non-trivially, $\mathfrak{a} \cap C \neq 0$.*

*Proof.* By hypothesis $R$ has a family of primitive ideals $\mathfrak{t}_\lambda$ whose intersection is 0. Let $f = 0$ be a polynomial identity for $R$; then $R_\lambda = R/\mathfrak{t}_\lambda$ is a primitive PI-algebra satisfying $f = 0$ and we have an embedding

$$R \to \prod R_\lambda,$$

by Th. 4.1. If $d = \deg f$, then by Th. 4.6, $R_\lambda$ is simple of degree at most $d/2$ over its centre. The canonical homomorphism $\varepsilon_\lambda : R \to R_\lambda$ is surjective and $R_\lambda$ contains $\mathfrak{a}\varepsilon_\lambda$ as ideal, which is therefore 0 or $R_\lambda$, by simplicity. Put

$$\Lambda_0 = \{\lambda \in \Lambda \mid \mathfrak{a}\varepsilon_\lambda \neq 0\},$$

and choose $\lambda_0 \in \Lambda_0$ such that the degree of $R_{\lambda_0}$, $n_0$ say, is maximal. Then the Razmyslov polynomial $D = D_{n_0}$ is central and non-zero on $R_{\lambda_0}$, and since $\mathfrak{a}\varepsilon_{\lambda_0} = R_{\lambda_0}$, there exist $a_1, \ldots, a_m \in \mathfrak{a}$ such that $x = D(a_1\varepsilon_{\lambda_0}, \ldots, a_m\varepsilon_{\lambda_0})$ is a non-zero element of the centre of $R_{\lambda_0}$. It follows that $y = D(a_1, \ldots, a_m) \neq 0$; moreover, $y\varepsilon_\lambda = 0$ for $\lambda \notin \Lambda_0$, while for $\lambda \in \Lambda_0$, $y\varepsilon_\lambda$ is in the centre of $R_\lambda$; therefore $y \in \mathfrak{a} \cap C$. ■

COROLLARY 7.2 *If $R$ is a semiprimitive PI-algebra whose centre is a field, then $R$ is simple.*

For if $\mathfrak{a} \neq 0$, then $\mathfrak{a} \cap C \neq 0$, so $\mathfrak{a}$ contains a unit and $\mathfrak{a} = R$. ■

Here the conditions cannot be omitted, for the polynomial ring $k[x]$ is a semiprimitive PI-algebra which is not simple, and for an infinite-dimensional $k$-space $V$, $\mathrm{End}_k(V)$ is a primitive $k$-algebra whose centre is a field, but which is not simple (and of course, not a PI-algebra).

Th. 7.1 has a useful generalization. To prove it we first note a PI-analogue of Prop. 9.4.3.

PROPOSITION 7.3 *Any prime PI-algebra $A$ satisfies the maximum condition on left (or right) annihilators; hence every nil left or right ideal of $A$ is zero.*

*Proof.* Let $0 \subset \mathfrak{l}_1 \subset \mathfrak{l}_2 \subset \cdots$ be a strictly ascending chain of left annihilators and put $\mathfrak{r}_n = (\mathfrak{l}_n)_r$, so that $\mathfrak{r}_1 \supset \mathfrak{r}_2 \supset \cdots$. By Prop. 9.5.3 we may take the polynomial

identity to be multilinear; denote by $d$ the least integer for which there is a homogeneous multilinear polynomial $f$ of degree $d$:

$$f = \sum a_\sigma x_{1\sigma} \cdots x_{d\sigma},$$

such that $f$ vanishes for all $x_i \in \mathfrak{l}_i$ $(i = 1, \ldots, d)$. Here $d > 1$, for if $a_1 \mathfrak{l}_1 = 0$, then $\mathfrak{l}_1 = 0$, because $A$ is prime. We thus have

$$\sum a_\sigma x_{1\sigma} \cdots x_{d\sigma} \mathfrak{r}_{d-1} = 0, \qquad \text{for all } x_i \in \mathfrak{l}_i. \tag{1}$$

Now when $d\sigma \neq d$, then $x_{d\sigma} \in \mathfrak{l}_{d-1}$ and the corresponding term in (1) is 0; the remaining terms have the form $a_\sigma x_{1\sigma} \cdots x_{(d-1)\sigma} x_d \mathfrak{r}_{d-1}$. Hence

$$\sum a_\sigma x_{1\sigma} \cdots x_{(d-1)\sigma} \mathfrak{l}_d \mathfrak{r}_{d-1} = 0, \tag{2}$$

where $\sigma$ ranges over all permutations of $1, \ldots, d-1$. But $\mathfrak{l}_d \mathfrak{r}_{d-1}$ is a non-zero two-sided ideal and $A$ is prime, so

$$\sum a_\sigma x_{1\sigma} \cdots x_{(d-1)\sigma} = 0 \qquad \text{for } x_i \in \mathfrak{l}_i,$$

and this contradicts the definition of $d$. Hence the chain of $\mathfrak{l}_i$ breaks off and the maximum condition holds. By symmetry the same is true for right annihilators and now the rest follows by Prop. 9.4.3. ■

COROLLARY 7.4 *In a semiprime PI-algebra every nil ideal is zero.*

For if $R$ is semiprime, then $R$ can be embedded in $\prod R_\lambda$, where $R_\lambda = R/\mathfrak{c}_\lambda$ is a prime PI-algebra and $\bigcap \mathfrak{c}_\lambda = 0$. If $\mathfrak{n}$ is a nil ideal in $R$, then $\mathfrak{n}\varepsilon_\lambda$ is a nil ideal in $R_\lambda$, hence equal to 0, by Prop. 7.3, and so $\mathfrak{n} \subseteq \bigcap \ker \varepsilon_\lambda = 0$. ■

We deduce

THEOREM 7.5 (Rowen) *In a semiprime PI-algebra, every non-zero ideal meets the centre non-trivially. In particular, a semiprime PI-algebra whose centre is a field is simple.*

*Proof.* Let $R$ be a semiprime PI-algebra with centre $C$. By Cor. 7.4, $R$ has no non-zero nil ideals, hence by Prop. 4.5, the polynomial ring $R[t]$ is semiprimitive, and its centre is clearly $C[t]$, so by Th. 7.1, for any non-zero ideal $\mathfrak{a}$ in $R$ we have $\mathfrak{a}[t] \cap C[t] \neq 0$. On comparing coefficients we find that $\mathfrak{a} \cap C \neq 0$. Now the rest follows as in Cor. 7.2. ■

It is now an easy matter to deduce Posner's theorem, in a rather strong form, following Rowen 1973 (cf. Rowen [1980]).

THEOREM 7.6 (Posner 1960) *Let $R$ be a prime PI-algebra with centre $C$. Then $C$ is an integral domain, and if $K$ is its field of fractions, then the natural*

*mapping $R \to Q = R \otimes_C K$ is an embedding and $Q$ is a finite-dimensional simple $K$-algebra.*

*Proof.* The first part follows by Cor. 9.1.11; now any multilinear identity in $R$ also holds in $Q$, so by Th. 7.5, $Q$ is simple. Hence by Kaplansky's theorem (Th. 4.6), $Q$ is finite-dimensional over $K$. ■

Let $R$ be a prime PI-algebra and $Q$ its quotient ring by fractions of the centre. By Th. 7.6, $Q$ is of finite dimension over its centre and this dimension is a square, $n^2$ say (Prop. 7.2.2); the number $n$ is called the PI-*degree* of $R$. We can now determine the PI-degree of the generic division algebra:

COROLLARY 7.7 *The generic division algebra of degree n has PI-degree n.*

*Proof.* The generic matrix ring $F_{(n)}$ is a prime PI-algebra, hence Th. 7.6 applies, and its skew field of fractions $D$ is a finite-dimensional central simple algebra. Moreover, $F_{(n)}$ satisfies the standard identity $S_{2n}$ but no identity of lower degree, hence the same is true of $D$. It follows that on passing to a splitting field we obtain an $n \times n$ matrix ring, hence $D$ is of degree $n$ over its centre. ■

The PI-degree has been used to give a simple characterization of separable algebras, by M. Artin and C. Procesi.

For any ring $R$ with centre $C$, the mapping

$$\varphi_{a,b}: x \mapsto axb \qquad (a, b \in R)$$

is an endomorphism of $R$ as $C$-module and it is easily checked that the correspondence $a \otimes b \mapsto \varphi_{a,b}$ defines a homomorphism

$$\lambda: R^0 \otimes_C R \to \mathrm{End}_C(R). \tag{3}$$

If $R$ is a finitely generated projective $C$-module and (3) is an isomorphism, $R$ is called an *Azumaya algebra* or also *central separable* (it can be shown that this is equivalent to $R$ being separable as $C$-algebra, cf. **6.7**). The proof of the Artin–Procesi theorem given below uses the Razmyslov polynomial $D$. We recall that $D$ is multilinear alternating in $n^2$ variables $x_1, \ldots, x_{n^2}$ (and others which we shall not name explicitly). If we put $D_v(x) = D(x, x_1, \ldots, \hat{x}_v, \ldots, x_{n^2})$, where $\hat{x}_v$ means that $x_v$ is omitted, then the expression $x_0 D - \sum x_v D_v(x_0)$ is an alternating multilinear function of the $n^2 + 1$ variables $x_0, \ldots, x_{n^2}$ and so vanishes on any prime ring of PI-degree $n$, because such a ring can be embedded in its quotient ring which is spanned by $n^2$ elements over $C$.

THEOREM 7.8 (M. Artin–C. Procesi) *Let $R$ be a prime PI-algebra of PI-degree n. Then $R$ is Azumaya provided that each simple homomorphic image of $R$ has PI-degree n.*

*Proof.* (W. Schelter) Consider the Razmyslov polynomial $D(x_v)$; if its arguments lie in $R$, its values lie in the centre $C$ of $R$. If they all lie in some maximal ideal of $R$, we get a homomorphic image of $R$ of lower PI-degree. Hence the left ideal generated by the values must be all of $R$:

$$1 = \sum b_i D(a_{vi}), \qquad \text{where } a_{vi}, b_i \in R, \tag{4}$$

and the other arguments of $D$ also lie in $R$. Denote by $D_{vi}$ the function obtained from $D_v$ (defined above) by replacing $x_\mu$ by $a_{\mu i}$ for $\mu \neq v$. By the remark preceding the theorem, we have $rD(a_{vi}) - \sum a_{vi} D_{vi}(r) = 0$ for any $r \in R$; hence by (4),

$$r = \sum b_i r D(a_{vi}) = \sum b_i a_{vi} D_{vi}(r).$$

Thus we have a (finite) projective coordinate system for $R$, so $R$ is finitely generated projective over $C$.

To show that (3) is an isomorphism we note that $\mathrm{End}_C(R)$ is generated by elements of the form $rD_{vi}$ and if $D_{vi}(x) = \sum_j u_{vij} x v_{vij}$, then $(\sum r u_{vij} \otimes v_{vij})\lambda = rD_{vi}$, so $\lambda$ is surjective. Now

$$\begin{aligned} \sum p_j \otimes q_j &= \sum_{jvi} p_j \otimes b_i a_{vi} D_{vi}(q_j) \\ &= \sum p_j D_{vi}(q_j) \otimes b_i a_{vi} \\ &= \sum_{vir} \left\{ \sum_j p_j u_{vir} q_j \right\} v_{vir} \otimes b_i a_{vi}. \end{aligned} \tag{5}$$

If $\sum p_j \otimes q_j \in \ker \lambda$, then $\sum p_j x q_j = 0$ for all $x \in R$, hence $\sum p_j \otimes q_j = 0$ by (5); this shows $\lambda$ to be injective, so it is an isomorphism, as claimed. ■

It can be shown that the sufficient condition of this theorem is also necessary (cf. Rowen [1980]).

## Exercises

(1) Let $D$ be a skew field with centre $k$. Show that for any (skew) subfield $E$ of $D$, $E$ and $Ek$ have the same PI-degree.

(2) Show that in any PI-algebra the upper and lower nilradical coincide.

(3) (Amitsur) Let $R$ be a PI-algebra and $\mathfrak{N}$ the sum of all its nilideals. Use Cor. 4.8 to show that $R/\mathfrak{N}$ can be embedded in a matrix ring $\mathfrak{M}_n(E)$, where $E$ is a commutative ring. Deduce that $R$ satisfies an identity $S_{2n}^m = 0$, where $S_{2n}$ is the standard polynomial and $m \geqslant 1$. (By Ex. (1) of **9.7**, $m$ cannot always be taken to be 1. Hint: express $R$ as homomorphic image of a relatively free algebra satisfying a given polynomial identity holding in $R$.)

## 10.8 Simple algebras with minimal right ideals

In **10.3** we gave a description of primitive rings with minimal right ideals. In a simple ring $R$ with a minimal right ideal the socle must be the whole of $R$. Since $R$, as right ideal, is finitely generated by 1, it is a sum of finitely many minimal right ideals, hence a direct sum and so is right Artinian, hence a full $n \times n$ matrix ring over a skew field (cf. Ex. (5) of **10.3**). It is of interest to consider the corresponding situation for algebras; an algebra $A$ is called *simple* if $A^2 \neq 0$ and $A$ has no ideals other than $0, A$. Now a simple algebra with minimal right ideals is again equal to its socle, but the latter need not be finitely generated. To describe such algebras more precisely we need the notion of a Rees matrix algebra.

All our algebras in this section will be bimodules over a skew field $K$ such that $x(yz) = (xy)z$ for any $x, y, z$ in $K$ or in the algebra. Thus they could be described as $K$-rings, were it not for the fact that they will in general lack 1. We shall regard them as algebras over the centre of $K$; an alternative would be to suspend the convention that rings have a 1, but we shall not take this course to avoid confusion. In practical terms this makes no difference, since we shall not have occasion to consider the augmented algebra.

Let $C$ be any algebra and $I$, $\Lambda$ any sets; we shall write ${}^{\Lambda}C^{I}$ for the set of all matrices over $C$ with rows indexed by $\Lambda$ and columns indexed by $I$, briefly, $\Lambda \times I$ matices. Fix a matrix $P$ in ${}^{\Lambda}C^{I}$ which is *regular*, i.e. whose rows are left linearly independent and whose columns are right linearly independent over $C$. By the *Rees matrix algebra* over $C$ with *sandwich matrix* $P$ one understands the set $M$ of all $I \times \Lambda$ matrices $A = (a_{i\lambda})$ over $C$, with almost all entries zero, with componentwise addition:

$$(a_{i\lambda}) + (b_{i\lambda}) = (a_{i\lambda} + b_{i\lambda}),$$

and with the multiplication

$$A * B = APB, \qquad \text{where } A = (a_{i\lambda}), \quad B = (b_{i\lambda}).$$

This is well defined, since $A, B$ are both $I \times \Lambda$ and zero almost everywhere, while $P$ is $\Lambda \times I$. With these definitions we have the following characterization of simple algebras with minimal right ideals.

THEOREM 8.1 *For any algebra $A$ the following conditions are equivalent:*

(a) *$A$ is a simple algebra with a minimal right ideal,*

(b) *$A$ is a prime algebra which coincides with its socle,*

(c) *$A$ is isomorphic to a Rees matrix algebra over a skew field,*

(d) *$A$ is isomorphic to a dense algebra of linear transformations of finite rank on a vector space over a skew field,*

*$(a^0)$–$(d^0)$ the left-right analogues of (a)–(d).*

*Proof*. The equivalence (a)⇔(d) follows as in **10.3**, so it only remains to prove (a), (b), (c) equivalent; the equivalence to $(a^0)$–$(d^0)$ then follows by the symmetry of (c).

(a)⇒(b). By hypothesis the socle of $A$ is not zero, so it equals $A$, and being simple, $A$ is prime.

(b)⇒(c). Let $\mathfrak{r}$ be a minimal right ideal. Since $A$ is prime, $\mathfrak{r}^2 \neq 0$, so $\mathfrak{r}^2 = \mathfrak{r}$ and hence $\mathfrak{r} = a\mathfrak{r}$ for some $a \in \mathfrak{r}$. Choose $e \in \mathfrak{r}$ such that $ae = a$; if $e^2 \neq e$, then $(e^2 - e)\mathfrak{r} = \mathfrak{r}$ and we have $\mathfrak{r} = a\mathfrak{r} = a(e^2 - e)\mathfrak{r} = 0$, a contradiction. Hence $e^2 = e$ and of course $e \neq 0$, so $\mathfrak{r} = e\mathfrak{r}$ and $e$ is an idempotent generator. By Lemma 6.3.8 and Schur's lemma, $\mathrm{End}_A(eA) = eAe$ is a skew field, $K$ say. Now $A$, being equal to its socle, is semisimple as right $A$-module. Let $\mathfrak{s}$ be the sum of all right ideals isomorphic to $\mathfrak{r}$. This is a two-sided ideal; if $\mathfrak{s} \neq A$, then we have $A = \mathfrak{s} \oplus \mathfrak{s}'$ for some non-zero right ideal $\mathfrak{s}'$, hence $\mathfrak{s}'\mathfrak{s} = 0$, which contradicts the fact that $A$ is prime. Therefore $A = \mathfrak{s}$ is a direct sum of right ideals isomorphic to $\mathfrak{r}$. Now $\mathfrak{r} = eA$ is a left $K$-space with basis $u_\lambda$ ($\lambda \in \Lambda$); likewise $Ae$ is a right $K$-space with basis $v_i$ ($i \in I$); further we note that $u_\lambda v_i \in eAe = K$. We define a $\Lambda \times I$ matrix $P = (p_{\lambda i})$ over $K$ by

$$p_{\lambda i} = u_\lambda v_i,$$

and claim that $P$ is regular. For if $(c_\lambda)$ is a family in $K$, almost all zero, such that $\sum c_\lambda p_{\lambda i} = 0$ for all $i$, then $\sum c_\lambda u_\lambda v_i = 0$, hence $\sum c_\lambda u_\lambda$ annihilates $Ae$ and so must be 0, and now we have $c_\lambda = 0$ by the linear independence of the $u_\lambda$. Similarly $\sum p_{\lambda i} d_i = 0$ for all $\lambda$ implies $d_i = 0$.

Let $M$ be the Rees matrix algebra over $K$ with sandwich matrix $P$ and define a mapping $f : M \to A$ by

$$(a_{i\lambda})f = \sum v_i a_{i\lambda} u_\lambda. \tag{1}$$

Since almost all the $a_{i\lambda}$ vanish, this is well defined. We shall establish (c) by proving that $f$ is an isomorphism. It is clearly additive, and we have

$$\begin{aligned}[(a_{i\lambda})*(b_{i\lambda})]f &= [\sum a_{i\lambda} u_\lambda v_j b_{j\mu}]f \\ &= \sum v_i a_{i\lambda} u_\lambda v_j b_{j\mu} u_\mu \\ &= (a_{i\lambda})f \cdot (b_{j\mu})f;\end{aligned}$$

thus $f$ is a homomorphism. It is surjective because $AeA$ is the socle and so equal to $A$, and $f$ is injective because $P$ is regular, for if $\sum v_i a_{i\lambda} u_\lambda = 0$, then $\sum p_{\mu i} a_{i\lambda} p_{\lambda j} = 0$, hence $a_{i\lambda} = 0$. Thus $f$ is an isomorphism and (c) follows.

(c)⇒(a). Given a Rees matrix algebra $M$ over $K$ with sandwich matrix $P$, take any non-zero matrix $C = (c_{i\lambda})$ in $M$. For any $(\mu, j) \in \Lambda \times I$ consider

$$b_{\mu j} = \sum p_{\mu i} c_{i\lambda} p_{\lambda j}.$$

If $b_{\mu j} = 0$ for all $\mu, j$, then by the regularity of $P$, $c_{i\lambda} = 0$ for all $i, \lambda$, a contradiction. Hence $b_{\mu j} \neq 0$ for some $(\mu, j) \in \Lambda \times I$. Let us write $E_{i\lambda}$ for the matrix unit with

$(i, \lambda)$-entry 1 and all others 0. For any $d \in K$ we have

$$dE_{i\mu} * C * E_{j\lambda} = dE_{i\mu} PCPE_{j\lambda} = (f_{k\rho}),$$

where

$$f_{k\rho} = \sum d\delta_{ki} p_{\mu l} c_{l\sigma} p_{\sigma j} \delta_{\lambda\rho} = \delta_{ki} \delta_{\lambda\rho} db_{\mu j}.$$

Since $d$ was arbitrary in $K$, $f_{k\rho}$ ranges over $K$, and every matrix of $M$ is a finite sum of terms $dE_{i\lambda}$, it follows that every matrix of $M$ lies in the ideal generated by $C$; hence $M$ is simple.

It remains to find a minimal right ideal. By the definition of $P$, $p_{\lambda i} \neq 0$ for some pair $(\lambda, i) \in \Lambda \times I$. Writing $p_{\lambda i} = p$, we have $p^{-1}E_{i\lambda} * p^{-1}E_{i\lambda} = p^{-1}E_{i\lambda} PE_{i\lambda} p^{-1} = p^{-1}E_{i\lambda} \neq 0$, hence $e = p^{-1}E_{i\lambda}$ is a non-zero idempotent in $M$. Now the mapping $\varphi: K \to M$ defined by $c \mapsto p^{-1}E_{i\lambda} c$ is an injective homomorphism with image $eMe$, as is easily verified. Hence $eMe$ is a skew field and so $eM$ is a minimal right ideal of $M$. ■

This result is closely analogous to the structure theorem on 0-simple semigroups by D. Rees 1940. The equivalence of (a) and (c) is due to Hotzel and Petrich 1974; the equivalence of (a) and (d) was proved by Jacobson 1945, building on work by Dieudonné 1942.

### Exercises

(1) Let $A$ be a Rees matrix algebra over a skew field $K$ with sandwich matrix $P$. Verify directly that if either the number of rows or the number of columns of $P$ is finite, then $P$ is square and $A$ is a full matrix ring over $K$.

(2) Let $A$ be a Rees matrix algebra over a skew field $K$ with $\Lambda \times I$ sandwich matrix $P$. Show that $|\Lambda| \leqslant 2^{|I|}$ and give an example where equality occurs. (Hint: if $V$ is a $K$-space of dimension $|I|$, then its dual $V^*$ has dimension $2^{|I|}$; interpret the rows of $P$ as vectors in $V^*$.)

(3) Show that if a $\Lambda \times I$ sandwich matrix is row-finite, then $|I| \leqslant |\Lambda|$.

## 10.9 Firs and semifirs

Principal ideal domains (PIDs) include several important types of commutative ring such as the ring of integers, polynomial rings in one variable over a field and certain rings of algebraic integers, more specifically, any Dedekind domain with unique factorization (cf. A.2, **9.5**). In the non-commutative case there are no such striking examples of PIDs, but there is a wider class of rings which reduce to PIDs when commutativity is imposed.

DEFINITION By a *free right ideal ring* or *right fir* for short, we understand

a ring $R$ with invariant basis number (IBN) whose right ideals are free, as right $R$-modules. *Left firs* are defined similarly and a left and right fir is called a *fir*. We have already briefly met firs in **11.5** of A.2, where it was shown that the free algebra $k\langle X\rangle$ on any set $X$ over a field $k$ is a fir; this may be regarded as a generalization of the fact that the polynomial ring $k[x]$ in a single variable $x$ is a PID (A.1, **10.5**). We also note that in Prop. 9.1.9 we had a situation where free ideals occur naturally.

It will be convenient to consider a somewhat wider class of rings (than firs): to describe it we shall need to look at general relations in rings. A relation

$$x_1y_1 + \cdots + x_ny_n = 0 \tag{1}$$

in a ring $R$ is said to be *trivial* if for each $i = 1, \ldots, n$, either $x_i = 0$ or $y_i = 0$. Every non-zero ring has non-trivial relations, for example, if $x = (1, 1)$, $y = (-1, 1)^T$, then

$$x \cdot y = (1, 1)\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 0. \tag{2}$$

However, this relation can be transformed into a trivial relation by replacing $x, y$ by $x'$, $y'$, given by

$$x' = (1, 1)\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = (2, 0), \qquad y' = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3}$$

Let us write the relation (1) as $x \cdot y = 0$, where $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)^T$. Then (1) is said to be *trivializable* if there is an invertible $n \times n$ matrix $P$ over $R$ such that the relation $xP^{-1} \cdot Py = 0$ is trivial; we also say that (1) is *trivialized* by P. More generally, a matrix relation $XY = 0$, where $X$ is $r \times n$ and $Y$ is $n \times s$, is *trivial* if for each $i = 1, \ldots, n$ either the $i$th column of $X$ or the $i$th row of $Y$ is 0, and $XY = 0$ is *trivializable* if $XP^{-1}$. $PY = 0$ is trivial for some $P \in \mathbf{GL}_n(R)$. In the above example, (3) shows the relation (2) to be trivializable.

We begin by describing the rings in which every relation is trivializable.

THEOREM 9.1 *Let $R$ be a non-zero ring. Then the following conditions are equivalent:*

(a) *Every relation in $R$ can be trivialized,*

(b) *every finitely generated right ideal in $R$ is free, as right $R$-module, of unique rank,*

(c) *$R$ has IBN and every finitely generated submodule of a free right $R$-nodule is free,*

(d) *every matrix relation in $R$ can be trivialized,*

*$(a^0)$–$(d^0)$ the left–right analogues of (a)–(d). Further, any such ring is an integral domain.*

*Proof.* (a)⇒(b). Let $\mathfrak{a}$ be a finitely generated right ideal of $R$ and let $n$ be the least integer such that $\mathfrak{a}$ has an $n$-element generating set, $u_1,\dots,u_n$ say. Then $\mathfrak{a}$ is free on $u_1,\dots,u_n$, for if not, take a non-trivial relation $u\cdot a=0$. By (a) this can be trivialized say $u'=uP^{-1}$, $a'=Pa$. Since $a\neq 0$, we have $a'\neq 0$, say $a'_n\neq 0$. But $u'\cdot a'=0$ is trivial, so $u'_n=0$ and hence $\mathfrak{a}$ is generated by $u'_1,\dots,u'_{n-1}$, thus contradicting the choice of $n$; hence $\mathfrak{a}$ is free. If $\mathfrak{a}$ has another basis $v_1,\dots,v_m$ and $m\neq n$, then $m>n$ and by mapping $v_i\mapsto u_i$ $(i=1,\dots,n)$, $v_j\mapsto 0$ $(j>n)$, we obtain an endomorphism $f$ of $\mathfrak{a}$ which is surjective but not injective $(fv_m=0)$. Thus $fv_1,\dots,fv_n$ generate $\mathfrak{a}$ but not freely; by the first part, $\mathfrak{a}$ can be generated by fewer than $n$ elements, which is again a contradiction, hence $m=n$ and $\mathfrak{a}$ has unique rank.

(b)⇒(c). Let $F$ be a free right $R$-module and $G$ a finitely generated submodule. The finite generating set of $G$ involves only finitely many generators of $F$; by ignoring the rest we may take $F$ to be finitely generated. Let $\lambda_1$ be the projection of $F$ on the first factor $R$ and denote by $F'$ the kernel of $\lambda_1$. Then we have the exact sequence

$$0\to F'\cap G\to G\to \mathfrak{a}\to 0, \tag{4}$$

where $\mathfrak{a}$ is the image of $G$ under $\lambda_1$, a finitely generated right ideal of $R$. By (b), $\mathfrak{a}$ is free, hence (2) splits and $G\cong(F'\cap G)\oplus\mathfrak{a}$. By induction $F'\cap G$ as finitely generated submodule of $F'$ is free and it follows that $G$ is free.

We next show that $R$ is an integral domain. Let $a\in R$ and denote its right annihilator by $\mathfrak{n}$. Then $aR\cong R/\mathfrak{n}$; since $aR$ is free, we have $R=\mathfrak{n}\oplus\mathfrak{a}$, where $\mathfrak{a}\cong aR$. Both $\mathfrak{n}$ and $\mathfrak{a}$ are free; by the uniqueness of the rank, either $\mathfrak{a}=0$ or $\mathfrak{n}=0$, so $a$ is either 0 or right regular. This holds for all $a\in R$, hence $R$ is an integral domain. Now if $R$ is right Ore, then it is embeddable in a skew field and so has IBN (Th. 4.4.7 of A.2); otherwise by Prop. 9.1.9 it has free right ideals of any finite rank and this rank is unique by (b); hence $R$ has IBN.

(c)⇒(d). Let $XY=0$ be a matrix relation, where $X\in {}^rR^n$, $Y\in {}^nR^s$. The matrix $X$ defines a linear map $\varphi:{}^nR\to {}^rR$ by left multiplication, and we have an exact sequence

$$0\to\ker\varphi\to {}^nR\to\operatorname{im}\varphi\to 0. \tag{5}$$

As a finitely generated submodule of ${}^rR$, $\operatorname{im}\varphi$ is free, so the exact sequence (5) splits, and by changing the basis in ${}^nR$ we obtain a basis adapted to the submodule $\ker\varphi:{}^nR=\ker\varphi\oplus F$, where $F\cong\operatorname{im}\varphi$, $\dim\ker\varphi=t$, say. If this change of basis is described by $P\in\mathbf{GL}_n(R)$, we put $X'=XP^{-1}$, $Y'=PY$. Since $X'\cdot Y'=X\cdot Y=0$, the columns of $Y'$ lie in $\ker\varphi$; thus the rows of $Y'$ after the first $t$ are 0, while the first $t$ columns of $X'$ are zero. Hence the matrix relation $X'\cdot Y'=0$ is indeed trivial.

Now (a) is a special case of (d) and $(a^0)$–$(d^0)$ follow by the evident symmetry of (a). ■

A non-zero ring satisfying the conditions of this theorem is called a *semifir*. By (b) every left or right fir is a semifir; however, there are right firs that are not left firs (cf. Ex. (5)). We also remark that a commutative semifir is just a Bezout domain, while a commutative fir is a PID.

As already remarked, the free algebra $k\langle X\rangle$ over a field is a fir; more generally this holds for the tensor $D$-ring $D_k\langle X\rangle$, where $D$ is any skew field and $k$ a subfield. This is proved by means of the weak algorithm, which we shall not define here (cf. Cohn [1985], Ch. 2); it is easier to prove the weaker statement that $D_k\langle X\rangle$ is a semifir, and we shall do so now:

THEOREM 9.2 *Let $D$ be a skew field and $F$ a subfield. Then the tensor $D$-ring $D_F\langle X\rangle$ on any set $X$ is a semifir.*

*Proof.* Let $\{u_v\}$ be a left $F$-basis of $D$ and $X=\{x_i\}$; then every element of $D_F\langle X\rangle$ can be uniquely written in the form $c+\sum u_v x_i f_{vi}$, where $c\in D$ and $f_{vi}\in D_F\langle X\rangle$. This follows because we can express any element of $D$ as $a=\sum u_v a_v$, where $a_v\in F$, and $ax_i=\sum u_v a_v x_i=\sum u_v x_i a_v$.

Suppose now that we have a relation in $D_F\langle X\rangle$:

$$\sum_1^n a_i b_i = 0. \tag{6}$$

In order to show that this relation can be trivialized it is enough to do this in a given degree; thus we may assume that the $a_i$, $b_i$ are homogeneous and that $\deg a_i + \deg b_i = r > 0$. We shall use double induction, on $n$ and $r$. If each $a_i$ has positive degree, we can write $a_i=\sum u_v x_t a_{vti}$: equating cofactors of $u_v x_t$ we find

$$\sum_i a_{vti} b_i = 0,$$

and now the result follows by induction on $r$, for we can make a transformation reducing one of the $b$'s to 0. There remains the case where some $a_i$, say $a_1$ has degree 0. Then we can replace $a_2$ by $a_2 - a_1 \cdot a_1^{-1} a_2 = 0$ and $b_1$ by $b_1' = b_1 + a_1^{-1} a_2 b_2$; we thus obtain

$$\sum a_i b_i = a_1 b_1' + a_3 b_3 + \cdots + a_n b_n = 0,$$

and we have now diminished $n$, and so can apply induction to complete the proof. ∎

In Ch. 11 we shall need the inertia lemma for semifirs (cf. Cohn [1985], Lemma 4.6.3). Below we state and prove it in the form needed here. We recall that from any ring $R$ we can form the polynomial ring $R[t]$ in a central indeterminate $t$; its completion by power series is the formal power series ring $R[\![t]\!]$, and if we localize now at the powers of $t$ we obtain the formal Laurent

series ring $R((t))$. Since $t$ is regular in $R[\![t]\!]$, the latter ring is embedded in the Laurent series ring.

Let $B$ be any ring and $A$ a subring. Then $A$ is said to be *finitely inert* in $B$ if for any matrix $Z \in {}^rA^s$, if $Z = XY$ over $B$, where $X$ is $r \times n$ and $Y$ is $n \times s$, there exists $P \in \mathbf{GL}_n(B)$ such that $XP^{-1}$, $PY$ have their entries in $A$.

LEMMA 9.3 (Inertia lemma) *Let $R$ be a semifir. Then for any central indeterminate $t$, the formal power series ring $R[\![t]\!]$ is finitely inert in $R((t))$.*

*Proof.* Let us put $S = R[\![t]\!]$, $T = R((t))$ and indicate the natural homomorphism $S \to R$ by a bar; it amounts to putting $t = 0$. We take $A \in {}^rS^s$ and suppose that over $T$:

$$A = PQ, \qquad \text{where } P \text{ is } r \times n \text{ and } Q \text{ is } n \times s. \tag{7}$$

If $P$ or $Q$ is 0, there is nothing to prove, so we may suppose that $P, Q \neq 0$. Over $T$ every non-zero matrix $C$ can be written in the form $t^vC'$, where $v \in \mathbf{Z}$, $C'$ has entries in $S$ and $\overline{C'} \neq 0$. Let $P = t^\mu P'$, $Q = t^\nu Q'$, where $\overline{P'}$, $\overline{Q'} \neq 0$. Dropping the dashes and writing $\mu + \nu = -\lambda$, we can rewrite (7) as

$$A = t^{-\lambda}PQ, \qquad \text{where } P \in {}^rS^n,\ Q \in {}^nS^s \text{ and } \bar{P},\ \bar{Q} \neq 0. \tag{8}$$

If $\lambda \leqslant 0$, there is nothing to prove, so assume that $\lambda > 0$. Then $\overline{PQ} = 0$; since $R$ is a semifir, we can find a matrix $U \in \mathbf{GL}_n(R)$ trivializing this relation, and on replacing $P, Q$ by $PU^{-1}$, $UQ$ we find that all the columns in $\bar{P}$ after the first $h$ are 0, while the first $h$ rows of $\bar{Q}$ are 0. If we multiply $P$ on the right by $V = tI_h \oplus I_{n-h}$ and $Q$ on the left by $V^{-1}$, then $P$ becomes divisible by $t$, while $Q$ still has entries in $S$. In this way we can, by cancelling a factor $t$, replace $\lambda$ by $\lambda - 1$ in (8) and after $\lambda$ steps we obtain the same equation with $\lambda = 0$. This proves the finite inertia. ■

We remark that there is a stronger notion, total inertia, and the inertia theorem asserts that an inversely filtered fir with inverse weak algorithm is totally inert in its completion (cf. Cohn [1985], Th. 2.9.15), but this will not be needed here (cf. Ex. (5) of Ch. 11).

## Exercises

(1) Show that a right Ore domain is a right fir iff it is right principal.

(2) Show that every semifir is weakly finite.

(3) Let $R$ be a semifir and $A, B$ finitely generated submodules of a free $R$-module. Show that $A \cap B$, $A + B$ are again free and that $\mathrm{rk}(A + B) + \mathrm{rk}(A \cap B) = \mathrm{rk}\, A + \mathrm{rk}\, B$.

(4) Let $R$ be a right fir. Show that any submodule of a free right $R$-module is free.

(5) In the group algebra over a field $k$ of the free group on $x, y$ let $R$ be the subalgebra generated by $x, y, x^{-1}y, x^{-2}y, \ldots$. Verify that $R$ is a semifir but not a left fir (it can be shown that $R$ is a right fir, cf. Cohn [1985], 2.10).

## Further exercises on Chapter 10

(1) Show that if $V$ is a vector space of infinite dimension $v$ over a skew field $K$ of cardinal $\alpha$, then $V$ has cardinal $v\alpha$ and $V^* = \mathrm{Hom}_K(V, K)$ has cardinal $\alpha^v$. Show that if $K$ is commutative and $\dim V^* = v^*$, then $v^* \geqslant \alpha$, and deduce that $v^* = \alpha^v$ (this is true even when $K$ is skew, cf. Jacobson [1953]).

(2) Let $K$ be a skew field and for $n \geqslant 1$ embed $\mathfrak{M}_{2^n}(K)$ in $\mathfrak{M}_{2^{n+1}}(K)$ by mapping $A$ to $A \oplus A$. Show that the union of the rings $\mathfrak{M}_{2^n}(K)$ is simple. What are its projective modules?

(3) Let $K$ be a skew field. Describe the ideal structure of the following infinite matrix rings over $K$: (i) the ring of all row-finite and column-finite matrices, (ii) the ring of all matrices which are equal to a scalar outside a finite square.

(4) Let $V$ be a $(K, R)$-bimodule, where $K$ is a skew field and $R$ acts fully on $V$, with centralizer $K$. Show that $V_R$ is simple; if moreover $R$ acts fully on ${}^2V$, then $R$ acts densely on $V$.

(5) (Jacobson) Let $k$ be a field and $R$ the $k$-algebra (with 1) generated by $u, v$ subject to $uv = 1$. Show that $R$ is primitive by representing it by the linear transformations $e_i u = e_{i+1}$, $e_i v = e_{i-1}$ if $i > 1$, and $e_1 v = 0$.

(6) (P. Samuel) Let $k$ be a field and $k\langle x, y\rangle$ the free $k$-algebra. Show that this is primitive by representing it as follows: $e_i x = e_{i^2+1}$, $e_i y = e_{i-1}$ if $i > 1$, $e_1 y = 0$.

(7) Show that in any ring $R$ the ideals with primitive quotient are just the cores of maximal right ideals. Show further that the intersection of these ideals is $\mathbf{J}(R)$.

(8) Show that in a prime ring $R$ every non-zero right ideal is a faithful $R$-module.

(9) Show that every maximal ideal in a ring is a prime ideal.

(10) Show that the least ideal $\mathfrak{N}$ in a ring $R$ such that $R/\mathfrak{N}$ is semiprime is a nil ideal (and hence $\mathfrak{N}$ is the lower nilradical).

(11) Find the socle of the ring of all upper triangular matrices over a field, and show that it differs from the left socle.

(12) Let $K$ be a commutative ring and $t$ an indeterminate. Show that if $a_0 + a_1 t + \cdots + a_n t^n \in K[t]$ is a unit, then $a_0$ is a unit and $a_1, \ldots, a_n$ are nilpotent. Deduce that the Jacobson radical of $K[t]$ is its nilradical.

(13) Show that the left (or right) ideal generated by a strongly nilpotent element is nilpotent.

(14) Let $R$ be a prime ring. Show that any non-zero central element of $R$ is a non-zero-divisor; deduce that the centre of $R$ is an integral domain.

(15) (Kaplansky) With any square matrix $A$ over a ring $K$ we can associate an 'infinite periodic' matrix by taking the diagonal sum of countably many copies of $A$. Let $R$ be the ring of all upper triangular infinite periodic matrices over a field $k$. Show that $R$ is prime, with Levitzki nilradical equal to the Jacobson radical.

(16) Let $R$ be any $K$-algebra and define the *centroid* of $R$ as $\mathrm{End}({}_RR_R)$. Show that the centroid $\Gamma$ is a commutative $K$-algebra and that $R$ has a natural $\Gamma$-algebra structure. If $R$ is primitive and $R^2 = R$, show that $\Gamma$ is an integral domain.

(17) Show that $k\mathbf{Q}_+$ is a Bezout domain. For real numbers $\alpha, \beta$ show that the monoid $M$ of all positive numbers of the form $m\alpha + n\beta$ is such that the monoid algebra $kM$ is Bezout iff $\alpha/\beta$ is rational.

(18) Over a semifir show that if a matrix product $AB$ has an $r \times s$ block of zeros, then for a suitable invertible $n \times n$ matrix $P$, $AP^{-1}$ has an $r \times t$ block of zeros and $PB$ has an $n - t \times s$ block of zeros (this is the partition lemma, cf. Cohn [1985] Lemma 1.1.4).

(19) Show that the group algebra of a free group is a semifir.

# 11
# Skew fields

Skew fields arise quite naturally in the application of Schur's lemma and elsewhere, but most of the known theory deals with the case of skew fields finite-dimensional over their centres (cf. Ch. 7). In the general case only isolated results are known, much of it depending on the coproduct construction (cf. Cohn[1977], Schofield [1985]). This lies outside our framework and we shall confine ourselves to presenting some of the highlights which do not require special prerequisites.

After some general remarks (including the Cartan–Brauer–Hua theorem) we shall give an account of determinants over skew fields. This is followed by a proof of the existence of free fields, based on the specialization lemma whose proof has been much simplified. Many of the concepts (though fewer of the actual results) of valuation theory carry over to skew fields and we shall examine the situation and pursue some of the results that continue to hold in the general case. The final section is concerned with pseudo-linear extensions, a more tractable class of finite-dimensional extensions, and we shall meet examples of skew field extensions whose left and right dimensions are different (E. Artin's problem).

Throughout this chapter we shall use the term 'field' to mean 'not necessarily commutative division ring'; the prefix 'skew' is sometimes added for emphasis.

## 11.1 Generalities

Let $K$ be a skew field. Its centre $C$ is a commutative field, and the characteristic of $C$ is also called the *characteristic* of $K$. We have already seen in Th. 7.1.13 that every finite field is commutative. However, an infinite field may well have a finite centre, in fact, for every commutative field $k$ there is a skew field with centre $k$. In characteristic 0 we can adjoin variables $u, v$ with the relation $uv - vu = 1$ to obtain such a field, but there is another construction which applies quite generally; a third construction, generally valid, is used in Th. 3.3 below.

PROPOSITION 1.1 *Let $k$ be any commutative field. The there is a skew field $D$ whose centre is $k$.*

*Proof.* Consider the group algebra of the additive group of rationals over $k$:

$k[x^\lambda | \lambda \in \mathbf{Q}]$. Clearly this is a commutative $k$-algebra; we take the subalgebra generated by $x^{2^{-n}}$, $n = 0, 1, \ldots$ and form its field of fractions:

$$E = k(x, x^{1/2}, x^{1/4}, \ldots).$$

This field has the automorphism $\alpha: f(x) \mapsto f(x^2)$, which is of infinite order. Moreover, $k$ is the precise subfield fixed by $\alpha$, for if $f$ involves $x$, let $2^n$ $(n \geqslant 0)$ be the largest denominator in any power of $x$ occurring in $f$. Then $x^{r/2^n}$ occurs in $f$ for some odd $r$, but it does not occur in $f^\alpha$, so $f$ is not fixed under $\alpha$. Now form the skew polynomial ring $E[y; \alpha]$ and let $D$ be its field of fractions. By Th. 9.3.5, $D$ has the precise centre $k$. ■

Much of linear algebra can be done over a skew field (cf. A.2, Ch. 4); this rests on the fact that every module over a field is free. As we saw in A.2, Th. 4.4.8, this property actually characterizes skew fields.

We go on to prove two general results, which although not used here, are of importance, with many applications. The first concerns normal subgroups of the multiplicative group of a field; the proof below is based on an idea of J. Treur.

THEOREM 1.2 (Cartan–Brauer–Hua theorem) *Let $K \subseteq L$ be skew fields, and assume that $K^\times$ is a normal subgroup of $L^\times$. Then either $K = L$ or $K$ is contained in the centre of $L$.*

*Proof.* For any $c \in L^\times$, the mapping $\alpha_c: x \mapsto c^{-1}xc$ satisfies

$$(x + y)\alpha_c = x\alpha_c + y\alpha_c.$$

Further we have $x\alpha_c = x \cdot (x, c)$, where $(x, c) = x^{-1}c^{-1}xc \in K$ whenever $c \in K$. If $K \neq L$, take $a \in L \setminus K$; for any $c \in K^\times$, since $(1, c) = 1$, we have

$$(a + 1)(a + 1, c) = a(a, c) + 1.$$

By the linear independence of 1, $a$ over $K$ we have $(a, c) = (a + 1, c) = 1$, thus $ac = ca$ for all $c \in K$, $a \notin K$. But if $b \in K$, then $a + b \notin K$, therefore $bc - cb = (a + b)c - c(a + b) - [ac - ca] = 0$, hence $c$ commutes with every element of $L$, i.e. $K$ is contained in the centre of $L$. ■

This result was obtained in the case of division algebras by H. Cartan 1947 as a consequence of his Galois theory of skew fields. R. Brauer and independently, L. K. Hua observed in 1949 that the general case could be proved directly.

Our second result concerns additive mappings preserving inversion:

THEOREM 1.3 (Hua's theorem) *Let $\sigma: K \to L$ be a mapping between two skew*

*fields such that*

$$(a+b)^\sigma = a^\sigma + b^\sigma, \qquad 1^\sigma = 1, \qquad (a^{-1})^\sigma = (a^\sigma)^{-1}. \tag{1}$$

*Then $\sigma$ is either a homomorphism or an antihomomorphism.*

*Proof*. (E. Artin) We must show that for all $a, b \in K$, $(ab)^\sigma = a^\sigma b^\sigma$ or for all $a, b \in K$, $(ab)^\sigma = b^\sigma a^\sigma$. We observe that $a^\sigma (a^{-1})^\sigma = 1$ by (1), so $a \neq 0 \Rightarrow a^\sigma \neq 0$ and it follows that $\sigma$ is injective. We start from the following identity (Hua's identity):

$$a - (a^{-1} + (b^{-1} - a)^{-1})^{-1} = aba, \tag{2}$$

valid whenever all inversions are defined, i.e. $ab \neq 0, 1$. To prove (2), we observe that for any $x \neq 0, 1$,

$$(x^{-1} - 1)^{-1} = (1 - x)^{-1} - 1, \tag{3}$$

as we see by multiplying out. Let $ab \neq 0, 1$; then $a^{-1}(b^{-1} - a) = (ba)^{-1} - 1$, hence on taking $x = ba$ in (3) we find

$$(b^{-1} - a)^{-1} a = ((ba)^{-1} - 1)^{-1} = (1 - ba)^{-1} - 1,$$
$$(b^{-1} - a)^{-1} = (1 - ba)^{-1} a^{-1} - a^{-1} = (a - aba)^{-1} - a^{-1}.$$

Hence

$$a - aba = (a^{-1} + (b^{-1} - a)^{-1})^{-1},$$

and (2) follows from this on rearranging the terms.

Now $(a^{-1})^\sigma = (a^\sigma)^{-1}$ and we may denote both sides by $a^{-\sigma}$. If we apply $\sigma$ to (2) and observe that $\sigma$ is compatible with all the operations on the left, we obtain

$$(aba)^\sigma = a^\sigma b^\sigma a^\sigma. \tag{4}$$

Clearly this still holds if $a$ or $b$ is 0; if $b = a^{-1}$, then $b^\sigma = a^{-\sigma}$ and both sides reduce to $a^\sigma$, so (4) holds in all cases. Put $b = 1$ in (4):

$$(a^2)^\sigma = (a^\sigma)^2. \tag{5}$$

Next replace $a$ by $a + b$ in (5) and simplify, using (5):

$$(ab + ba)^\sigma = a^\sigma b^\sigma + b^\sigma a^\sigma. \tag{6}$$

Now consider $(c^\sigma - a^\sigma b^\sigma) c^{-\sigma} (c^\sigma - b^\sigma a^\sigma)$ for any $c \neq 0$. This equals

$$c^\sigma - a^\sigma b^\sigma - b^\sigma a^\sigma + a^\sigma b^\sigma c^{-\sigma} b^\sigma a^\sigma = c^\sigma - (ab + ba)^\sigma + (abc^{-1}ba)^\sigma$$
$$= (c - ab - ba + abc^{-1}ba)^\sigma,$$

by an application of (6), (4) and (1). Thus

$$(c^\sigma - a^\sigma b^\sigma) c^{-\sigma} (c^\sigma - b^\sigma a^\sigma) = (c - ab - ba + abc^{-1}ba)^\sigma, \tag{7}$$

for all $a, b, c$ such that $c \neq 0$. For $c = ab$ the right-hand side reduces to

$ab - ab - ba + ba = 0$, hence the left-hand side of (7) vanishes for $c = ab$. Thus $(ab)^\sigma$ is either $a^\sigma b^\sigma$ or $b^\sigma a^\sigma$, it only remains to show that the same alternative holds for all pairs.

Fix $a \in K$ and put $U_a = \{b \in K \mid (ab)^\sigma = a^\sigma b^\sigma\}$, $V_a = \{b \in K \mid (ab)^\sigma = b^\sigma a^\sigma\}$. They are clearly subgroups of $K^+$ whose union is $K$, hence for each $a \in K$, one of them must be all of $K$ (cf. Ex. (2) below). Now put $U = \{a \in K \mid U_a = K\}$, $V = \{A \in K \mid V_a = K\}$; then $U$, $V$ are again subgroups whose union is $K$, so one of them must be all of $K$, i.e. either $(ab)^\sigma = a^\sigma b^\sigma$ for all $a, b \in K$ or $(ab)^\sigma = b^\sigma a^\sigma$ for all $a, b \in K$. ■

This result is used in projective geometry to show that a bijective transformation of the line which preserves harmonic ranges necessarily has the form $x \mapsto ax^\sigma + b$, where $a \neq 0$ and $\sigma$ is an automorphism or an antiautomorphism (for $K = \mathbf{R}$ this is v. Staudt's theorem, cf. E. Artin [1957]).

## Exercises

(1) Let $k$ be a commutative field of characteristic 0. Verify that the field of fractions of the Weyl algebra $A_1(k)$ has centre $k$. What is the centre when $k$ has prime characteristic?

(2) In the proof of Th. 1.3 the fact was used that a group $G$ cannot be the union of two proper subgroups $H, K$. By considering the product of two elements, one not in $H$ and one not in $K$, prove this fact.

(3) Let $K$ be a skew field. Show that for any $c \in K$ the centralizer of the set of all conjugates of $c$ is either $K$ or the centre of $K$. Deduce that no non-central element $c$ can satisfy the identity $cx^{-1}cx = x^{-1}cxc$ for all $x \in K^\times$.

(4) Let $\sigma: K \to L$ be a mapping between fields such that $(x + y)^\sigma = x^\sigma + y^\sigma$, $1^\sigma = \lambda$, $(x^\sigma)^{-1} = \lambda^{-1}(x^{-1})^\sigma \lambda^{-1}$. Show that $x^\sigma = x^\tau \lambda$, where $\tau$ is a homomorphism or an antihomomorphism.

(5) Show that any non-central element in a skew field has infinitely many conjugates.

(6) Let $K$ be a skew field. Show that any conjugacy class of elements of $K$ outside the centre generates $K$ as a field. Show that any subfield containing $K^{\times\prime}$ coincides with $K$.

(7) (Herstein) Let $K$ be a skew field of prime characteristic and $G$ a finite subgroup of $K^\times$. Denoting by $P$ the prime subfield of $K$, show that the $P$-space spanned by $G$ is an algebra, hence a finite field. Deduce that $G$ must be cyclic.

(8) Let $K$ be a skew field. Show that any abelian normal subgroup of $K^\times$ is contained in the centre of $K$.

## 11.2 The Dieudonné determinant

The determinant is a fundamental concept, going back much further than the notion of matrix on which it is based (matrices were introduced by Cayley in the middle of the 19th century, whereas determinants had been used at the end of the 17th century by Leibniz). Today we regard the determinant of a square matrix as an alternating multilinear function of the columns of the matrix; its most important property is that its vanishing characterizes the linear dependence of its columns. Here the entries of the matrix are assumed to lie in a (commutative) field, but it is clear that the definition is unchanged when the entries are taken from any commutative ring. So it is natural to try to extend the definition to the non-commutative case; we have seen in **7.3** how this can be done for a finite-dimensional algebra over a field by means of the norm. The general definition is due to Dieudonné (1943); before presenting it let us examine the simplest case, that of a $2 \times 2$ matrix.

The columns of the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ over a skew field $K$ say, are linearly dependent iff the equations

$$\begin{aligned} ax + by &= 0, \\ cx + dy &= 0, \end{aligned} \tag{1}$$

have a non-trivial solution $(x, y)$. If $a = 0$, such a solution exists precisely when $b = 0$ or $c = 0$, so let us assume that $a \neq 0$. Then in any non-trivial solution, $y \neq 0$ and by eliminating $x$ from (1) we obtain $(ca^{-1}b - d)y = 0$, hence the condition for linear dependence is

$$ca^{-1}b - d = 0. \tag{2}$$

Depending on which entries of $A$ are zero, we can find various expressions whose vanishing characterizes the linear dependence of the columns of $A$, but a few trials make it clear that there is no *polynomial* in $a, b, c, d$ with this property. Thus we must expect a determinant function (if one exists) to be a rational function. A second point is that under any reasonable definition one would expect $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ and $\begin{pmatrix} b & 0 \\ 0 & a \end{pmatrix}$ to have the same determinant. This suggests that even for a skew field the values of the determinant must lie in an abelian group. For any field $K$ we define the *abelianized group* $K^{ab}$ as

$$K^{ab} = K^\times / K^{\times\prime},$$

where $K^{\times\prime}$ is the derived group of the multiplicative group $K^\times$. Thus for a commutative field $K^{ab}$ reduces to $K^\times$. It is clear that $K^{ab}$ is universal for homomorphisms of $K^\times$ into abelian groups (cf. also Th. 2.2.7).

As usual we write $\mathbf{GL}_n(K)$ for the group of all invertible $n \times n$ matrices over

$K$. Let us recall that any $m \times n$ matrix $A$ over $K$ may be interpreted as the matrix of a linear mapping from an $m$-dimensional to an $n$-dimensional vector space over $K$. By choosing suitable bases in these spaces we can ensure that $A$ takes the form $I_r \oplus 0$, where $r$ is the rank of $A$. Thus there exist $P \in \mathbf{GL}_m(K)$, $Q \in \mathbf{GL}_n(K)$ such that

$$PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}. \tag{3}$$

This was proved in the remarks after Prop. 9.2.3. In particular, (3) shows that a square matrix over any field is invertible iff it is left (or equivalently, right) regular (see also Cor. 2.3 below). Such a matrix is also called *non-singular* and a non-invertible matrix is called *singular*.

Without a determinant we cannot define $SL_n$, but we have the group $\mathbf{E}_n(K)$ generated by all elementary matrices $B_{ij}(a) = I + aE_{ij}$ (cf. **4.7**). We observe that multiplication by an elementary matrix corresponds to an elementary operation on matrices; more precisely, left multiplication by $B_{ij}(a)$ amounts to adding $a$ times the $j$th row to the $i$th row and right multiplication by $B_{ij}(a)$ means adding the $i$th column multiplied by $a$ to the $j$th column.

As a first result we show that we can pass from $AB$ to $BA$ by elementary transformations, provided that the matrices are 'enlarged' by forming their diagonal sum with a unit matrix. Here it is not necessary to assume that the coefficients lie in a field.

LEMMA 2.1 (Whitehead's lemma) *Let $R$ be any ring and $n \geqslant 1$. For any $A, B \in \mathbf{GL}_n(R)$, $AB \oplus I$ and $BA \oplus I$ lie in the same (left or right) coset of $\mathbf{E}_{2n}(R)$, a fact which may be expressed by writing*

$$\begin{pmatrix} AB & 0 \\ 0 & I \end{pmatrix} \equiv \begin{pmatrix} BA & 0 \\ 0 & I \end{pmatrix} (\mathrm{mod}\, \mathbf{E}_{2n}(R)). \tag{4}$$

*Proof*. What we must show is that

$$A^{-1}B^{-1}AB \oplus I \in \mathbf{E}_{2n}(R). \tag{5}$$

In the first place we note that for any $C \in \mathbf{GL}_n(R)$,

$$\begin{pmatrix} 0 & C \\ -C^{-1} & 0 \end{pmatrix} = \begin{pmatrix} I & C \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ -C^{-1} & I \end{pmatrix} \begin{pmatrix} I & C \\ 0 & I \end{pmatrix} \in \mathbf{E}_{2n}(R), \tag{6}$$

for each matrix on the right can be written as a product of $n^2$ elementary matrices. Hence we have

$$\begin{pmatrix} C^{-1} & 0 \\ 0 & C \end{pmatrix} = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \begin{pmatrix} 0 & C \\ -C^{-1} & 0 \end{pmatrix} \in \mathbf{E}_{2n}(R), \tag{7}$$

for the matrices on the right are instances of (6). Now we have

$$\begin{pmatrix} A^{-1}B^{-1}AB & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} A^{-1} & 0 \\ 0 & A \end{pmatrix}\begin{pmatrix} B^{-1} & 0 \\ 0 & B \end{pmatrix}\begin{pmatrix} AB & 0 \\ 0 & (AB)^{-1} \end{pmatrix} \in \mathbf{E}_{2n}(R). \quad \blacksquare$$

As we shall soon see, in the case of a skew field $\mathbf{E}_n(K)$ is the derived group $\mathbf{GL}_n(K)$ (except when $n = 2 = |K|$), in particular, $E_n$ is a normal subgroup of $GL_n$; for a commutative field $k$, $\mathbf{E}_n(k) = \mathbf{SL}_n(k)$ and the result was proved in Prop. 4.7.2.

We can embed $\mathbf{GL}_n(K)$ in $\mathbf{GL}_{n+1}(K)$ by mapping $A$ to $A \oplus 1$. In this way we obtain an ascending chain

$$\mathbf{GL}_1(K) \subset \mathbf{GL}_2(K) \subset \cdots,$$

whose union is again a group, written $\mathbf{GL}(K)$ and called the *stable* general linear group. Its elements may be thought of as infinite matrices which differ from the unit matrix only in a finite square. Similarly the union of the groups $\mathbf{E}_n(K)$ is a group $\mathbf{E}(K)$.

In order to obtain a definition for the determinant we shall need to refine the expression (3) for a matrix. A square matrix is called *lower unitriangular* if all the entries on the main diagonal are 1 and those above it are zero; it is clear from the definition that such a matrix is a product of elementary matrices and hence is invertible. Moreover, the lower unitriangular matrices form a group under multiplication. An *upper unitriangular matrix* is defined similarly. We observe that left multiplication by a lower unitriangular matrix amounts to adding left multiples of certain rows to later rows, and right multiplication by an upper unitriangular matrix comes to adding right multiples of certain columns to later columns.

We can now describe a decomposition which applies to any matrices over a skew field. Our account follows that of Draxl [1983], with some simplifications.

THEOREM 2.2 (Bruhat normal form) *Let $K$ be a skew field and $A \in {}^mK^n$. Then $A$ can be written in the form*

$$A = LMU, \tag{8}$$

*where $L$ is an $m \times m$ lower unitriangular matrix, $U$ an $n \times n$ upper unitriangular matrix and $M$ is an $m \times n$ matrix with at most one non-zero entry in any row or column. Moreover, any other such decomposition of $A$ leads to the same matrix $M$.*

The matrix $M$ is called the *core* of $A$ and (8) is the *Bruhat decomposition*. For example, when $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then the Bruhat decomposition is

$$\begin{pmatrix} 1 & 0 \\ ca^{-1} & 1 \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & d - ca^{-1}b \end{pmatrix}\begin{pmatrix} 1 & a^{-1}b \\ 0 & 1 \end{pmatrix} \text{ if } a \neq 0, \qquad \begin{pmatrix} 1 & 0 \\ db^{-1} & 1 \end{pmatrix}\begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}$$

if $a = 0 \neq b$,

$$\begin{pmatrix} 0 & 0 \\ c & 0 \end{pmatrix}\begin{pmatrix} 1 & c^{-1}d \\ 0 & 1 \end{pmatrix} \text{ if } a=b=0\neq c, \qquad \begin{pmatrix} 0 & 0 \\ 0 & d \end{pmatrix} \text{ if } a=b=c=0.$$

*Proof.* Suppose that the first non-zero row of $A$ is the $i$th row and that its first non-zero entry is $a_{ij}$. By adding left multiples of the $i$th row to each succeeding row we can reduce every entry in the $j$th column, except $a_{ij}$, to zero. These operations correspond to left multiplication by a certain lower unitriangular matrix. Next we add right multiples of the $j$th column to succeeding columns, so as to reduce all entries in the $i$th row except $a_{ij}$ to zero; these operations will correspond to right multiplication by an upper unitriangular matrix. As a result $a_{ij}$ is the only non-zero element in its row and column, and all rows above the $i$th are zero. Next we take the first non-zero row after the $i$th, say the $k$th row and with the first non-zero entry $a_{kl}$ in this row we continue the process. After at most $m$ steps $A$ has been reduced to the form $M$ where each row and each column has at most one non-zero entry, with a left factor which is lower and a right factor which is upper unitriangular. This is the required decomposition (8).

If $A=L'M'U'$ is another such decomposition, then on writing $P=L^{-1}L'$, $Q=U'U^{-1}$, we have

$$PM'Q=M,$$

where $P$ is again lower and $Q$ upper triangular. This tells us that we can pass from $M'$ to $M$ by adding left multiples of rows to later rows and right multiples of columns to later columns. If $m'_{rs}$ is a non-zero entry of $M'$, the application of these operations only affects entries in the $r$th row or $s$th column and leaves $m'_{rs}$ itself unchanged. Hence in any operation on $M'$, the $(i,j)$-entry is affected only if $M'$ has either a non-zero entry in the $i$th row before the $j$th column, or a non-zero entry in the $j$th column above the $i$th row. In either case $m'_{ij}=0$, while $m'_{rs}$ remains unchanged. Hence $m_{rs}=m'_{rs}$ and it follows that $m_{ij}=0$. Therefore $M'=M$, as we had to show. ■

We remark that the matrices $L$, $U$ in (8) are not generally unique. For example, we have

$$\begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}\begin{pmatrix} 0 & u \\ v & 0 \end{pmatrix}\begin{pmatrix} 1 & -v^{-1}au \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} 0 & u \\ v & 0 \end{pmatrix}.$$

For a truly unique form see Ex. (4).

COROLLARY 2.3 *For any $m\times n$ matrix $A$ over a skew field $K$ the following conditions are equivalent:*

(a) *$A$ is left regular*: $XA=0\Rightarrow X=0$,

(b) *$A$ has a right inverse*: $AB=I$ *for some* $B\in{}^nK^m$.

*Moreover, when* (a), (b) *hold, then $m\leqslant n$, with equality if and only if* (a), (b) *are equivalent to their left–right analogues.*

This follows because either of (a), (b) clearly holds for $A$ iff it holds for its core, and it holds for the core iff each row has a non-zero entry. ■

Let us define a *monomial matrix* as a square matrix with precisely one non-zero entry in each row and each column; e.g. the core of any invertible matrix is a monomial matrix. If all the non-zero entries in a monomial matrix are 1, we have a *permutation matrix*; this may also be defined as the matrix obtained by permuting the rows (or equivalently, the columns) of the unit matrix. The determinant of a permutation matrix is 1 or $-1$ according as the permutation is even or odd. Sometimes it is more convenient to have matrices of determinant 1; this can be accomplished by using a *signed* permutation matrix, i.e. a matrix obtained from the unit matrix by a series of operations which consist in interchanging two columns and changing the sign of one of them. By (6) such a matrix is a product of elementary matrices.

Any monomial matrix $M$ may be written in the form

$$M = DP, \tag{9}$$

where $D$ is a diagonal matrix and $P$ is a signed permutation matrix. By Cor. 2.3 any matrix over a skew field is invertible precisely when its core is a monomial matrix.

With the help of Th. 2.2 we can also identify $E_n$:

PROPOSITION 2.4 *For any skew field $K$ and any $n \geqslant 1$, we have*

$$\mathbf{GL}_n(K) = \mathbf{D}_n(K) \cdot \mathbf{E}_n(K), \tag{10}$$

*where $\mathbf{D}_n(K)$ is the group of all diagonal matrices in $\mathbf{GL}_n(K)$. Moreover, for any $n \geqslant 2$,*

$$\mathbf{E}_n(K) = \mathbf{GL}_n(K)', \tag{11}$$

*except when $n = 2$, $K = \mathbf{F}_2$, when $\mathbf{E}_2(\mathbf{F}_2) = \mathbf{GL}_2(\mathbf{F}_2)$.*

*Proof.* By Th. 2.2, any invertible matrix can be written as $LDPU$, where $L$ is lower, $U$ is upper unitriangular, $D$ is diagonal and $P$ is a signed permutation matrix. Now $P$ can be written as a product of elementary matrices, using (6); moreover, $LD = DL'$, where $L'$ is again lower unitriangular, hence our matrix takes the form $D \cdot F$, where $F \in \mathbf{E}_n(K)$, and this proves (10).

To establish (11), let us write $(A, B) = A^{-1}B^{-1}AB$ and $(H, K)$ for the subgroup generated by all $(A, B)$, $A \in H$, $B \in K$. Let us write $G_n$ for $\mathbf{GL}_n(K)$ and similarly for $D_n$, $E_n$. We first show that $(G_n, G_n) \subseteq E_n$; by (10) this will follow if we can show that $(D_n, D_n) \subseteq E_n$ and $(D_n, E_n) \subseteq E_n$. The first inclusion follows from Lemma 2.1 (because $n \geqslant 2$), while the second results from the formula

$$\begin{pmatrix} u^{-1} & 0 \\ 0 & v^{-1} \end{pmatrix}\begin{pmatrix} 1 & -a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix}\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a - u^{-1}av \\ 0 & 1 \end{pmatrix}.$$

In the other direction we have

$$\begin{pmatrix}1 & a\\ 0 & 1\end{pmatrix}=\left(\begin{pmatrix}1 & 0\\ 0 & v\end{pmatrix},\begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix}\right)=\begin{pmatrix}1 & b(1-v)\\ 0 & 1\end{pmatrix}, \tag{12}$$

provided that $a=b(1-v)$. If $K\neq \mathbf{F}_2$, there is an element $v\neq 0,1$ and putting $b=a(1-v)^{-1}$, we can use (12) to express $B_{ij}(a)$ as a commutator. Hence $E_n\subseteq G_n'$ whenever $K\neq \mathbf{F}_2$. If $n\geqslant 3$, we have

$$\begin{pmatrix}1 & a & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}=\left(\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & a & 1\end{pmatrix}\right),$$

so we again have $E_n\subseteq G_n'$ and (11) follows. When $K=\mathbf{F}_2$ and $n=2$, then $\mathbf{GL}_2(\mathbf{F}_2)$ is the symmetric group of degree 3 and is equal to $\mathbf{E}_2(\mathbf{F}_2)$, as is easily verified. ■

We now define for any skew field $K$, a mapping $\delta:\mathbf{GL}(K)\to K^{ab}$ from the stable linear group to the abelianized group of $K$, as follows: for $A\in\mathbf{GL}_n(K)$, $\delta(A)=\prod_{i=1}^{n}\bar{d}_i$, where $d_i$ are the diagonal elements of $D$ in the expression (9) for the core of $A$ and $\bar{d}_i$ is its residue class mod $K^{\times\prime}$. $\delta(A)$ is called the *Dieudonné determinant*, or simply the *determinant* of $A$. To obtain its properties we shall need to find how it is affected by permutations of its rows; for simplicity we consider the effect of signed permutations.

LEMMA 2.5 *For any $A\in\mathbf{GL}_n(K)$ and any signed permutation matrix $P$,*

$$\delta(PA)=\delta(A).$$

*Proof.* By induction it will be enough to prove the result when $P$ is the signed permutation matrix corresponding to a transposition, $(r,s)$ say where $r<s$. Denote this matrix by $P_0$ and take a Bruhat decomposition (8) of $A$, where the core is factorized as in (9):

$$A=LDPU,$$

and denote the $(s,r)$-entry of $L$ by $b$. We have

$$\begin{aligned}P_0A=P_0LDPU&=L'B_{rs}(-b)P_0DPU\\&=L'D'B_{rs}(c)P_0PU,\end{aligned} \tag{13}$$

where $L', D'$ are again lower unitriangular and diagonal respectively, and $D'$ differs from $D$ by an interchange of the $r$th and $s$th diagonal elements. If $c=0$ (which is the case iff $b=0$) or if the permutation corresponding to $P_0P$ preserves the order of $r,s$, then the matrix on the right takes the form $L'D'P_0PB_{rs}(c)U$; this is again in Bruhat normal form and so we have $\delta(P_0A)=\delta(A)$ in this case.

Suppose now that $c \neq 0$ and that $P_0P$ inverts the order $r, s$. Then the formula

$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c^{-1} & 1 \end{pmatrix}\begin{pmatrix} c & 0 \\ 0 & c^{-1} \end{pmatrix}\begin{pmatrix} 1 & -c^{-1} \\ 0 & 1 \end{pmatrix}$$

shows that, on writing $D_i(c)$ for the matrix differing from the unit matrix only in the $(i, i)$-entry, which is $c$, we have

$$B_{rs}(c)P_0 = B_{sr}(c^{-1})D_r(c)D_s(c^{-1})B_{rs}(-c^{-1}).$$

Inserting this in the expression (13) for $P_0A$, we find

$$\delta(P_0A) = \overline{c \cdot c^{-1}}\delta(A) = \delta(A),$$

and this proves the assertion in all cases. The conclusion follows by induction. ■

We can now establish the main property of the determinant:

THEOREM 2.6 *For any skew field $K$, the determinant function $\delta$ is a homomorphism giving rise to an exact sequence*

$$1 \longrightarrow \mathbf{E}(K) \longrightarrow \mathbf{GL}(K) \xrightarrow{\delta} K^{ab} \longrightarrow 1.$$

*In particular, the determinant is unchanged by row or column operations.*

*Proof.* Let $A \in \mathbf{GL}_n(K)$; we first show that $\delta(BA) = \delta(A)$ for any $B \in \mathbf{E}_n(K)$, and by induction it is enough to prove this for $B = B_{ij}(a)$. If $i > j$, this matrix is lower unitriangular, so $BA$ and $A$ have the same core and hence the same determinant. If $i < j$, let $P_0$ be the signed permutation matrix corresponding to the transposition $i, j$. Using Lemma 2.5 and the case just proved, we have

$$\delta(B_{ij}(a)A) = \delta(P_0B_{ij}(a)A) = \delta(B_{ji}(a)P_0A) = \delta(P_0A) = \delta(A),$$

hence the result holds generally. The same argument applies for multiplying by an elementary matrix on the right. Now Lemma 2.1 shows that

$$\delta(AB) = \delta(BA) \qquad \text{for any } A, B. \tag{14}$$

Here it is of course necessary to consider $\delta$ as defined on the stable linear group.

Now take any two matrices $A$, $B$ with Bruhat decompositions $A = L_1M_1U_1$, $B = L_2M_2U_2$. We have, by what has been proved and (14),

$$\delta(AB) = \delta(L_1M_1U_1L_2M_2U_2) = \delta(M_1U_1L_2M_2) = \delta(M_2M_1U_1L_2) = \delta(M_2M_1).$$

Further, it is clear from the definition of $\delta$ that $\delta(M_2M_1) = \delta(M_1M_2) = \delta(M_1)\delta(M_2)$, so we obtain

$$\delta(AB) = \delta(A)\delta(B). \tag{15}$$

Hence $\delta$ is a homomorphism. Clearly its image is $K^{ab}$; its kernel includes $\mathbf{GL}(K)' = \mathbf{E}(K)$ because $K^{ab}$ is abelian. Conversely, if $\delta(A) = 1$, then the core of $A$ has the form $DP$, where $P$ is a signed permutation matrix and so $1 = \delta(A) = \delta(D)$. By (7) we can apply elementary operations to reduce $D$ to the form $D_1(c)$, where $c$ is a product of the diagonal elements of $D$. But by hypothesis, $\bar{c} = 1$, hence $D$ has been reduced to $I$ and so $A \in \mathbf{E}_n(K)$, as we had to show. ■

## Exercises

(1) Show that the transpose of every invertible matrix over a field $K$ is invertible iff $K$ is commutative. (Hint: try a $2 \times 2$ matrix with $(1,1)$-entry 1.)

(2) Use Th. 2.2 to show that $\mathbf{GL}_n(R) = \mathbf{D}_n(R)\mathbf{E}_n(R)$ for any local ring $R$.

(3) Show that $\mathbf{GL}_2(\mathbf{F}_2) = \mathbf{E}_2(\mathbf{F}_2) = S_3$.

(4) Let $K$ be a skew field. Show that $A \in \mathbf{GL}_n(K)$ can be written as $A = LDPU$, where $L$ is lower, $U$ is upper unitriangular, $D$ is diagonal, $P$ is a permutation matrix and $PUP^{-1}$ is also upper triangular. Moreover, such a representation is unique. (Draxl [1983]; this is known as the *strict* Bruhat normal form.)

(5) Show that if $P$ is the permutation matrix obtained by applying a permutation $\sigma$ to the rows of $I$, then it can also be obtained by applying $\sigma^{-1}$ to the columns of $I$.

(6) Let $K$ be a skew field with centre $C$. Show that $\delta$ restricted to $C$ reduces to the usual determinant, provided that no element of $C$ other than 1 is a product of commutators.

## 11.3 Free fields

We have seen that free rings and free algebras may be defined by a universal property; this is not to be expected for fields, since fields do not form a variety (cf. Th. 1.3.7). Nevertheless, in the commutative case the rational function field $k(x_1, \ldots, x_d)$ may be regarded as a 'free' field in the sense that all other fields generated by $d$ elements over $k$ can be obtained from it by specialization. Moreover, it is the field of fractions of the polynomial ring $k[x_1, \ldots, x_d]$ and as such it is uniquely determined. By contrast, the free algebra $k\langle x_1, \ldots, x_d \rangle$ has more than one field of fractions (cf. Ex. (2)), but this leaves the question whether in the general case there exists a field that is universal for specializations. A full study of these questions is beyond the scope of this book (cf. Cohn [1985] Ch. 7), but it is possible to prove the existence of free fields in a relatively straightforward way, and this will now be done.

For a general theory of fields it is convenient to invert matrices rather than elements. We shall not need to enter into details, but we have to consider which matrices can become invertible under a homomorphism to a field. Clearly we

can confine ourselves to square matrices. If $A$ is an $n \times n$ matrix over a ring $R$, and $A$ can be written in the form

$$A = PQ, \qquad \text{where } P \in {}^nR^r,\ Q \in {}^rR^n, \tag{1}$$

then it is clear that under any homomorphism from $R$ to a field we again have a factorization as in (1), hence the image of $A$ cannot have rank greater than $r$, and so cannot be invertible whenever $r < n$. The least possible value of $r$ in (1) for a given $A$ is called the *inner rank* of $A$ over $R$ and is denoted by $\rho A$. It is easily verified that over a field the inner rank reduces to the usual rank. Thus an $n \times n$ matrix over any ring cannot become invertible under a homomorphism to a field, unless its inner rank is $n$.

A square matrix over any ring $R$ is said to be *full* if its inner rank equals the number of rows. The above remarks show that in studying matrices that can be inverted under a homomorphism to a field, we can confine our attention to full matrices. Our aim in this section is to show that for the tensor ring $F$ there exists a field $U$ containing $F$ such that any full matrix over $F$ can be inverted over $U$. This field $U$ is a *universal field of fractions* of $F$, in the sense that there is a specialization from $U$ to any other field which is obtained by a homomorphism from $F$ (Cohn [1985], Ch. 7).

We begin with some remarks on the inner rank. We recall from **4.4** of A.2 that a ring $R$ is called *weakly finite* if for any square matrices of the same size $A, B, AB = I$ implies $BA = I$. If $R$ is a weakly finite ring which is non-trivial, then the unit matrix in $R$ must be full, i.e. if $A$ is $r \times s$ and $B$ is $s \times r$ and $AB = I$, then $r \leqslant s$. For if $r > s$, we can adjoin zero columns to $A$ and zero rows to $B$ to obtain square matrices. Now

$$(A \quad 0)\begin{pmatrix} B \\ 0 \end{pmatrix} = I, \qquad \text{hence} \qquad \begin{pmatrix} B \\ 0 \end{pmatrix}(A \quad 0) = I.$$

Comparing $(r,r)$-entries, we obtain $0 = 1$, which contradicts the fact that $R$ is non-trivial.

LEMMA 3.1 *Let $R$ be weakly finite ring and consider a partitioned matrix over $R$:*

$$A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}, \qquad \textit{where } A_1 \text{ is } r \times r.$$

*If $A_1$ is invertible, then $\rho A \geqslant r$, with equality if and only if $A_3 A_1^{-1} A_2 = A_4$.*

*Proof.* Clearly the inner rank is unchanged on passing to an associated matrix. Hence we can make the transformations

$$A \to \begin{pmatrix} I & A_1^{-1}A_2 \\ A_3 & A_4 \end{pmatrix} \to \begin{pmatrix} I & A_1^{-1}A_2 \\ 0 & A_4 - A_3A_1^{-1}A_2 \end{pmatrix} \to \begin{pmatrix} I & 0 \\ 0 & A_4 - A_3A_1^{-1}A_2 \end{pmatrix},$$

and these transformations leave the inner rank unchanged. If $\rho A = s$, this matrix can be written as

$$PQ = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}(Q_1 \quad Q_2),$$

where $P_1$ is $r \times s$ and $Q_1$ is $s \times r$. Thus $I = P_1 Q_1$, hence $r \leqslant s$ by weak finiteness, and this shows that $\rho A \geqslant r$. When equality holds, we have $Q_1 P_1 = I$, but $P_1 Q_2 = 0$, so $Q_2 = 0$. Similarly $P_2 = 0$ and hence $A_4 - A_3 A_1^{-1} A_2 = P_2 Q_2 = 0$. The converse is clear. ■

The existence proof for free fields is based on a lemma of independent interest, the specialization lemma. This may be regarded as an analogue of the GPI-theorem, which is used in the proof; its commutative counterpart is the elementary result that a polynomial vanishing for all values in an infinite field must be the zero polynomial.

LEMMA 3.2 (Specialization lemma) *Let $D$ be a skew field with infinite centre $C$ and such that $[D:C]$ is infinite. Then any full matrix over the tensor ring $D_C\langle X\rangle$ is invertible for some choice of $X$ in $D$.*

*Proof.* Let $A = A(x)$ be any full $n \times n$ matrix over $D_C\langle X\rangle$ and denote by $r$ the supremum of its ranks as its arguments range over $D$. We have to show that $r = n$, so let us assume that $r < n$. By a translation $x \mapsto x + a$ $(x \in X, a \in D)$ we may assume that the maximum rank is assumed at the point $x = 0$, and by an elementary transformation we may take the principal $r \times r$ minor to be invertible. Thus if

$$A(x) = \begin{pmatrix} B_1(x) & B_2(x) \\ B_3(x) & B_4(x) \end{pmatrix},$$

where $B_1$ is $r \times r$, then $B_1(0)$ is invertible. Given $a \in D^X$ and any $\tau \in C$, we have $\rho A(\tau a) \leqslant r$, hence by Lemma 9.2.4 the rank of $A(ta)$ over $D(t)$ is at most $r$, and the same holds over $D((t))$, the field of formal Laurent series in $t$ as central indeterminate. Now $B_1(ta)$ is a polynomial in $t$ with matrix coefficients and constant term $B_1(0)$, a unit, hence $B_1(ta)$ is invertible over the power series ring $D[\![t]\!]$. By Lemma 3.1, the equation

$$B_4(ta) = B_3(ta)B_1(ta)^{-1}B_2(ta) \tag{2}$$

holds over $D[\![t]\!]$, for all $a \in D^X$. This means that the matrix

$$B_4(tx) - B_3(tx)B_1(tx)^{-1}B_2(tx) \tag{3}$$

vanishes when the elements of $X$ are replaced by any values in $D$. Now (3) is a power series in $t$ with coefficients that are matrices over $D_C\langle X\rangle$. Thus the

coefficients are generalized polynomial identities (or identically 0), so by Amitsur's GPI-theorem (Th. 9.8.4), the expression (3) vanishes as a matrix over $D_C\langle X\rangle[\![t]\!]$. It follows that for $r<n$, $A(tx)$ is non-full over $D_C\langle X\rangle[\![t]\!]$. Hence we can on replacing $x\in X$ by $t^{-1}x$ write $A(x)$ as a product

$$A(x)=PQ, \qquad \text{where } P \text{ is } n\times r \text{ and } Q \text{ is } r\times n, \tag{4}$$

and $P, Q$ have as entries Laurent series in $t$ over $D_C\langle X\rangle$. Now $D_C\langle X\rangle$ is a semifir, by Th. 10.9.2 and by the inertia lemma (Lemma 10.9.3) we can, on replacing $P$ by $PU^{-1}$ and $Q$ by $UQ$, for suitably chosen $U$, assume that the entries of $P, Q$ are power series in $t$.

We now have (4), where $P, Q$ have power series in $t$ as entries. Putting $t=0$, we find that $A(x)$ is non-full over $D_C\langle X\rangle$. This is a contradiction and the conclusion follows. ■

In this lemma the condition $[D:C]=\infty$ is clearly necessary; whether the condition that $C$ be infinite is needed is not known. In its first version (Cohn 1972) the inertia theorem was needed for its proof (cf. Cohn [1985], Lemma 5.9.5); its use has now been replaced by the much easier inertia lemma.

We can now prove the existence of free fields:

THEOREM 3.3 *Let $D$ be a skew field with centre $C$ and $X$ any set. Then $D_C\langle X\rangle$ can be embedded in a field $U$, generated by $D_C\langle X\rangle$, such that every full matrix over $D_C\langle X\rangle$ becomes invertible over $U$.*

*Proof.* Suppose first that $[D:C]=\infty$, $|C|=\infty$, and consider the mapping

$$D_C\langle X\rangle \to D^{D^X},$$

where $p\in D_C\langle X\rangle$ is mapped to $(p_f)$, with $p_f=p(xf)$, for any $f\in D^X$. With each square matrix $A$ over $D_C\langle X\rangle$ we associate the subset $\mathscr{D}(A)$ of $D^X$ defined by

$$\mathscr{D}(A)=\{f\in D^X \mid A(xf) \text{ is invertible}\}.$$

$\mathscr{D}(A)$ is called the *singularity support* of $A$. Of course $\mathscr{D}(A)=\varnothing$ unless $A$ is full, but by Lemma 3.2, $\mathscr{D}(A)\neq\varnothing$ whenever $A$ is full. If $P, Q$ are any invertible matrices, then $P\oplus Q$ is invertible, hence $A(x)\oplus B(x)$ becomes singular precisely when $A(x)$ or $B(x)$ becomes singular; thus

$$\mathscr{D}(A)\cap\mathscr{D}(B)=\mathscr{D}(A\oplus B).$$

It follows that the family of sets $\mathscr{D}(A)$, where $A$ is full, is closed under finite intersections. Hence it is contained in an ultrafilter $\mathscr{F}$ on $D^X$ (cf. **1.6**), and we have a homomorphism to an ultrapower

$$D_C\langle X\rangle \to D^{D^X}/\mathscr{F}, \tag{5}$$

where by definition, every full matrix $A$ over $D_C\langle X\rangle$ is invertible on $\mathscr{D}(A)$ and

hence invertible in the ultrapower. Hence the subfield of the ultrapower generated by $D_C\langle X\rangle$ is the required field $U$.

In the general case we take indeterminates $r, s, t$ and define $D_1 = D(r)$, $D_2 = D_1(s)$. On $D_2$ we have an automorphism $\alpha: f(s) \mapsto f(rs)$, with fixed field $D_1$. We now form $E = D_2(t; \alpha)$; the centre of $E$ is the centre of $D_1$, namely $C(r)$ (cf. Th. 9.3.5). This is infinite and $E$ has infinite dimension over $C(r)$, because the powers $s^n$ are linearly independent. It is clear that we have an embedding $D_C\langle X\rangle \to E_{C(r)}\langle X\rangle$, and it follows from the inertia lemma that this embedding is *honest*, i.e. full matrices are mapped to full matrices. Hence on taking the field $U$ constructed earlier for $E_{C(r)}\langle X\rangle$, we obtain a field over which every full matrix over $D_C\langle X\rangle$ becomes invertible. ■

The field $U$ whose existence has been established in Th. 3.3 is denoted by $D_C \ll X \gg$ and is called the *universal field of fractions of* $D_C\langle X\rangle$ or also the *free field* over $D$ with centre $C$ (its centre can be shown to be $C$). The existence of such a universal field of fractions can be shown more generally for any tensor ring $K_L\langle X\rangle$, where $K$ is any skew field and $L$ any subfield of $K$. This is a special case of the fact that every semifir has a universal field of fractions over which every full matrix can be inverted (Cohn, [1985] Ch. 7).

We remark that any automorphism of $D_C\langle X\rangle$ is honest and therefore extends to an automorphism of $U$. Further, by representing derivations as homomorphism from $D_C\langle X\rangle$ to $\mathfrak{M}_2(U)$ we see that for any automorphisms $\alpha, \beta$ of $U$, any $(\alpha, \beta)$-derivation of $D_C\langle X\rangle$ extends to one of $U$.

## Exercises

(1) Show that the $n \times n$ unit matrix over a ring $R$ is full iff $R^n$ cannot be generated by less than $n$ elements. Show also that a weakly finite ring ($\neq 0$) has IBN (Hint: see Prop. 4.4.6 of A.2.)

(2) Let $E = k(t)$ be the field of functions in one variable $t$, with the endomorphism $\alpha_r: f(t) \mapsto f(t^r)$ ($r \geqslant 1$). Show that if $r > 1$, the subalgebra of $E[x; \alpha_r]$ generated by $x$ and $y = xt$ is free on $x, y$. Using Ex. (2) of **9.3**, obtain for each $r > 1$ a field of fractions $L_r$ of the free algebra $F$. Show that these fields are pairwise non-isomorphic as $F$-rings (J. L. Fisher).

(3) Show that an endomorphism $\theta$ of $D_C\langle X\rangle$ can be extended to an endomorphism of the free field iff $\theta$ is honest.

(4) Show that every honest homomorphism is injective; give an example of an endomorphism of the free algebra $k\langle X\rangle$ which is injective but not honest.

(5) Verify that over a skew field the inner rank agrees with the rank.

(6) Let $K$ be a skew field with infinite centre. Show that for any square matrix $A$ over

$K$ there is an element $\alpha$ in $K$ such that $A - \alpha I$ is non-singular. For a finite field $F$ find a matrix $A$ such that $A - xI$ is singular for all values of $x$ in $F$. (For infinite fields with finite centre the question remains open.)

(7) Show that over a PID, a square matrix is regular iff it is full. Give an example of a square matrix over a free algebra which is regular but not full. (Hint: try a $3 \times 3$ matrix with a $2 \times 2$ block of zeros.)

## 11.4 Valuations on skew fields

Valuations may be defined on skew fields as in the commutative case, but there have so far been fewer applications. This is no doubt due to the inherent difficulties in handling general valuations; however in special cases they become more tractable and offer the prospect of a means of gaining information on skew fields. Here we present a part of the general theory that runs parallel to the commutative case, together with some illustrations.

Let $K$ be a skew field. A subring $V$ of $K$ is said to be *total* if for every $a \in K^\times$, either $a$ or $a^{-1}$ lies in $V$. If for every $a \in K^\times$, $a^{-1}Va = V$, then $V$ is called *invariant*. Now a *valuation ring* of $K$ is an invariant total subring of $K$. In any valuation ring $V$ in $K$ the set $\mathfrak{m}$ of all its non-units is easily seen to be an ideal, hence $V$ is a local ring with $\mathfrak{m}$ as maximal ideal. The set $U$ of all units in $V$ is a normal subgroup of $K^\times$; we shall denote the quotient $K^\times/U$ by $\Gamma$ and call it the *value group* of $V$, with natural homomorphism $v: K^\times \to \Gamma$. We shall use additive notation for $\Gamma$; our main concern will be with the case when $\Gamma$ is abelian. Given $a, b \in K^\times$, we shall write $v(a) \geqslant v(b)$ iff $ab^{-1} \in V$, or equivalently (because $V$ is invariant), $b^{-1}a \in V$. This relation is a total ordering on $\Gamma$; for if $v(a) \geqslant v(b)$, $v(b) \geqslant v(c)$, then $ab^{-1}, bc^{-1} \in V$, hence $ac^{-1} \in V$ and so $v(a) \geqslant v(c)$. Clearly $v(a) \geqslant v(a)$ and if $v(a) \geqslant v(b)$ and $v(b) \geqslant v(a)$, then $ab^{-1}, ba^{-1} \in V$, hence $ab^{-1} \in U$ and so $v(ab^{-1}) = 0$, hence $v(a) = v(b)$. Under this ordering $\Gamma$ becomes an ordered group, for if $v(a) \geqslant v(b)$, then $ab^{-1} \in V$, hence for any $c \in K^\times$, $ac(bc)^{-1} = ab^{-1} \in V$, $ca(cb)^{-1} = cab^{-1}c^{-1} \in V$, hence $v(ac) \geqslant v(bc)$, $v(ca) \geqslant v(cb)$. Moreover, since $v$ is a homomorphism, we have $v(ab) = v(a) + v(b)$, and the fact that $ab^{-1} \in V \Rightarrow ab^{-1} + 1 \in V$ implies that $v(a) \geqslant v(b) \Rightarrow v(a + b) \geqslant v(b)$. We thus have

**V.1** $v(x) \in \Gamma$ for $x \in K^\times$,
**V.2** $v(x + y) \geqslant \min\{v(x), v(y)\}$,
**V.3** $v(xy) = v(x) + v(y)$.

Here we need to exclude the values $x, y = 0$, but it is more convenient to allow $x = 0$ and define $v(0)$ as $\infty$. Then **V.1–3** continue to hold if we define (as usual) $\infty + \alpha = \alpha + \infty = \infty + \infty = \infty$, $\alpha < \infty$ for all $\alpha \in \Gamma$. As in the commutative case we have equality in **V.2** whenever $v(x) \neq v(y)$ ('all triangles are isosceles').

A function $v$ from $K$ to an ordered group $\Gamma$ (with $v(0) = \infty$), satisfying **V.1–3**

is called a *valuation* on $K$. Given such a valuation, we can define

$$V = \{x \in K \mid v(x) \geqslant 0\},$$

and it is easily verified that $V$ is a valuation ring on $K$. In this way valuation rings on $K$ and valuations correspond to each other; to make the correspondence bijective we define two valuations $v, v'$ on $K$ with value groups $\Gamma, \Gamma'$ to be *equivalent* if there is an isomorphism $\varphi: \Gamma \to \Gamma'$ such that

$$v(x)\varphi = v'(x) \qquad \text{for all } x \in K^\times.$$

With this definition we have

THEOREM 4.1 *On any field there is a natural bijection between valuation rings and equivalence classes of valuations on $K$.*

The proof is an easy consequence of the above remarks and may be left to the reader. ■

Valuations on skew fields were introduced by O. F. G. Schilling in 1945. In the commutative case there is a third notion, equivalent to the above two, namely that of a *place*; this can also be defined for skew fields, but will not be needed here (cf. Ex. (2)).

We note that $K$ itself is a valuation ring in $K$; it corresponds to the *trivial* valuation, defined by $v(x) = 0$ for all $x \neq 0$, with trivial value group. Of course we shall mainly be interested in non-trivial valuations.

The simplest (non-trivial) type of an ordered group is the infinite cyclic group. It can be shown that a valuation has the infinite cyclic group as value group precisely when its valuation ring is a principal ideal domain; such a valuation is said to be *principal*. For example, the usual $p$-adic valuation on $\mathbf{Q}$ is principal, and in Ch. 9 of A.2 we saw that every valuation on a rational function field $k(t)$ which is trivial on $k$ is principal.

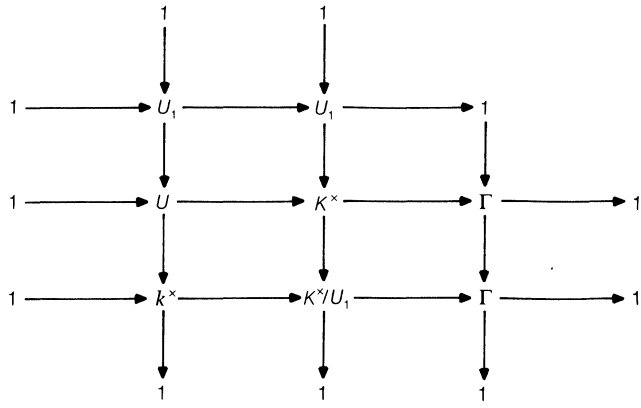
Let $K$ be any field with a valuation $v$ and write $V$ for its valuation ring, $\mathfrak{m}$ for its maximal ideal and $U$ for the group of units in $V$. It is clear from V.2 that every element of the form $1 + x$, where $x \in \mathfrak{m}$, is a unit; such a unit is called a 1-*unit* (Einseinheit). Thus $u$ is a 1-unit whenever $v(u-1) > 0$. The group of all 1-units is written $1 + \mathfrak{m}$ or $U_1$. Let us denote $V/\mathfrak{m}$, the residue class field of $V$, by $k$. Then we have a group isomorphism

$$k^\times \cong U/U_1, \tag{1}$$

while the value group of $v$ is given by

$$\Gamma \cong K^\times/U. \tag{2}$$

These isomorphisms may be combined into the following commutative diagram with exact rows and columns:




(3)

When $V$ is principal, $\Gamma$ is infinite cyclic; hence the horizontal sequences split and so

$$K^\times/U_1 \cong k^\times \times \Gamma. \tag{4}$$

For this reason the rows of (3) add nothing to our knowledge in the cases usually encountered; but in general, especially with a non-abelian value group, (3) provides more information about $K^\times$.

In constructing valuations it is helpful to know that any valuation on an Ore domain has a unique extension to its field of fractions.

PROPOSITION 4.2 *Let $R$ be a right Ore domain with field of fractions $K$. If $v$ is a valuation on $R$, satisfying* **V**.1–3, *then $v$ has a unique extension to $K$.*

*Proof.* If $v$ is to have an extension to $K$, then for $p = as^{-1} \in K$ we must have

$$v(p) = v(a) - v(s). \tag{5}$$

Suppose that $as^{-1} = a_1 s_1^{-1}$. Then there exist $u, u_1 \in R$ such that $su_1 = s_1 u \neq 0$, $au_1 = a_1 u$; hence $a = 0$ iff $a_1 = 0$, and when $a, a_1 \neq 0$, then $-v(a_1) + v(a) = v(u) - v(u_1) = -v(s_1) + v(s)$ and so $v(a) - v(s) = v(a_1) - v(s_1)$. This shows the definition (5) to be independent of the particular representation $as^{-1}$ of $p$. Now it is easily verified that $v$ so defined satisfies **V**.1–3 on $K$. ■

*Examples of valuations.*

1. Let $K$ be any field with a valuation $v$. We can extend $v$ to the rational function field $K(x)$ by defining $v$ on any polynomial $f = \sum x^i a_i (a_i \in K)$ by the rule

$$v(f) = \min\{v(a_i)\}, \tag{6}$$

and using Prop. 4.2 to extend $v$ to $K(x)$. This is called the *Gaussian extension*

of $v$; the value group is unchanged while the residue class field undergoes a simple transcendental extension. The same construction works if instead of $K(x)$ we use $K(x;\alpha)$, the skew function field with respect to an automorphism $\alpha$ of $K$, provided that $v(a^\alpha) = v(a)$ for all $a \in K$.

If we enlarge the value group $\Gamma$ by an element $\delta$ having no non-zero multiple in $\Gamma$ (e.g. by forming the direct product of $\Gamma$ and the infinite cyclic group on $\delta$, with lexicographic ordering), and instead of (6) define

$$v(f) = \min\{v(a_i) + i\delta\}, \tag{7}$$

we obtain an extension, called the *x-adic extension* (provided that $\delta > 0$), with the same residue class field and enlarged value group.

2. Consider the free field $k\langle\!\langle x, y\rangle\!\rangle$; let $E$ be the subfield generated over $k$ by $y_i = x^{-i}yx^i (i \in \mathbf{Z})$. The conjugation by $x$ defines an automorphism $\alpha$ which maps $E$ into itself, the 'shift automorphism' $y_i \mapsto y_{i+1}$. Moreover, $k\langle\!\langle x, y\rangle\!\rangle$ may be obtained as the skew function field $E(x;\alpha)$. Taking the $x$-adic extension of the trivial valuation on $E$, we obtain a principal valuation on $k\langle\!\langle x, y\rangle\!\rangle$ with residue class field $E$. We note that whereas the general aim of valuation theory is to obtain a simpler residue class field, $E$ is actually more complicated than the original field. As we shall see, in order to simplify the residue class field we must allow a more complicated value group.

3. In a skew field it may happen that an element is conjugate to its inverse: $y^{-1}xy = x^{-1}$. For example, let $\alpha$ be the automorphism of the rational function field $F = k(x)$ defined by $f(x) \mapsto f(x^{-1})$ and put $E = F(y;\alpha)$. If $v$ is any valuation on $E$, then $v(x) = 0$, for if $v(x) \neq 0$, say $v(x) > 0$, then $v(x^{-1}) < 0$, but $x^{-1} = y^{-1}xy$, hence $v(x^{-1}) = -v(y) + v(x) + v(y) > 0$, a contradiction. In fact fields exist in which every element outside the centre is conjugate to its inverse (e.g. the existentially closed fields, cf. Cohn [1977]). It is clear that such a field can have no non-trivial valuation.

One of the main tools of the commutative theory is Chevalley's extension theorem (cf. A.2, **8.5**), which allows one to construct extensions for any valuations defined on a subfield. Such a result is not to be expected in general; but an analogue exists when the value group is abelian, and it is no harder to prove.

A valuation on a field $K$ is said to be *abelian* if its value group is abelian. For any field $K$ we denote the derived group of $K^\times$ by $K^c$. It is clear that any abelian valuation on $K$ is trivial on $K^c$. As an almost immediate consequence we have

LEMMA 4.3 *Let $K$ be a skew field with a valuation $v$ and valuation ring $V$. Then $v$ is abelian if and only if $V \supseteq K^c$ or equivalently, $v(a) = 0$ for all $a \in K^c$. Moreover, any subring $A$ of $K$ such that $A \supseteq K^c$ is invariant, and any ideal in $A$ is invariant.*

*Proof.* The second sentence follows from the fact that $v$ is abelian iff the unit

group of $V$ contains $K^c$. To prove the last sentence, take $a \in A^\times, b \in K^\times$; then $b^{-1}ab = a \cdot a^{-1}b^{-1}ab \in A$, and similarly for any ideal of $A$. ■

To state the analogue of Chevalley's theorem (Lemma 4.4) we require the notion of domination. On any field $K$ we consider the pairs $P = (R, \mathfrak{a})$ consisting of a subring $R$ of $K$ and a proper ideal $\mathfrak{a}$ of $R$. Given two such pairs $P_i = (R_i, \mathfrak{a}_i)(i = 1, 2)$, we say that $P_1$ *dominates* $P_2$, in symbols $P_1 \geqslant P_2$ if $R_1 \supseteq R_2$ and $\mathfrak{a}_1 \supseteq \mathfrak{a}_2$. If the pair $(R, \mathfrak{a})$ is such that $R \supseteq K^c$, then every element of $K^c$ is a unit in $R$ (because $K^c$ is a group), and so $K^c \cap \mathfrak{a} = \varnothing$. The essential step in our construction is the following

LEMMA 4.4 *Let $K$ be a skew field, $R$ a subring containing $K^c$ and $\mathfrak{a}$ a proper ideal in $R$. Then there is a subring $V$ with a proper ideal $\mathfrak{m}$ such that $(V, \mathfrak{m})$ is maximal among pairs dominating $(R, \mathfrak{a})$, and any such maximal pair $(V, \mathfrak{m})$ consists of a valuation ring and its maximal ideal.*

The proof is similar to the commutative case (Lemma 8.4.3 of A.2); we briefly recall it to show where changes are needed.

The pairs dominating $(R, \mathfrak{a})$ form an inductive family, hence a maximal pair exists by Zorn's lemma. If $(V, \mathfrak{m})$ is a maximal pair, then $\mathfrak{m}$ is a maximal ideal in $V$, and since $V \supseteq K^c$, $V$ and $\mathfrak{m}$ are invariant. To show that $V$ is a total subring in $K$, take $c \in K^\times$; if $c \notin V$, then $V[c] \supset V$, so if the ideal $\mathfrak{m}'$ generated by $\mathfrak{m}$ in $V[c]$ is proper, we have $(V[c], \mathfrak{m}') > (V, \mathfrak{m})$, contradicting the maximality of the latter. Hence $\mathfrak{m}' = V[c]$ and we have an equation

$$a_0 + a_1 c + \cdots + a_m c^m = 1, \qquad a_i \in \mathfrak{m}. \tag{8}$$

Here we were able to collect powers of $c$ on the right because of the invariance of $\mathfrak{m}$, using the equation $cb = cbc^{-1} \cdot c$.

Similarly if $c^{-1} \notin V$, we have

$$b_0 + b_1 c^{-1} + \cdots + b_n c^{-n} = 1, \qquad b_j \in \mathfrak{m}. \tag{9}$$

We assume that $m, n$ are chosen as small as possible and supppose that $m \geqslant n$, say. Multiplying (9) on the right by $c^m$, we obtain

$$(1 - b_0)c^m = b_1 c^{m-1} + \cdots + b_n c^{m-n}. \tag{10}$$

By the invariance of $V$, $xc = c \cdot x\gamma$, for all $x \in V$, where $\gamma = \gamma(c)$ is an automorphism of $V$ which maps $\mathfrak{m}$ into itself. If we multiply (8) by $1 - b_0$ on the left and (10) by $a_m \gamma^m$ on the right and substitute into (8), we obtain an equation of the same form as (8) but of degree less than $m$, a contradiction. This proves $V$ to be total, and hence a valuation ring; from the maximality it is clear that $\mathfrak{m}$ is the maximal ideal in $V$. ■

We now have the following general form of the extension theorem:

THEOREM 4.5 *Let $K \subseteq L$ be an extension of skew fields. Given an abelian valuation $v$ on $K$, there is an extension of $v$ to $L$ if and only if there is no equation*

$$\sum a_i c_i = 1, \qquad \text{where } a_i \in K, v(a_i) > 0 \text{ and } c_i \in L^c. \tag{11}$$

*Proof.* If (11) holds, then any abelian extension $w$ of $v$ to $L$ must satisfy $w(a_i c_i) = w(a_i) = v(a_i) > 0$, hence $w(1) \geqslant \min_i\{w(a_i c_i)\} > 0$, a contradiction. Conversely, if no equation (11) holds, this means that if $V$ is the valuation ring of $v$, with maximal ideal $\mathfrak{m}$, then $\mathfrak{m}L^c$ is a proper ideal in $VL^c$, and by Lemma 4.4 there is a maximal pair $(W, \mathfrak{n})$ dominating $(VL^c, \mathfrak{m}L^c)$. Now $W$ is a valuation ring satisfying $W \cap K \supseteq V$, $\mathfrak{n} \cap K \supseteq \mathfrak{m}$, hence $W \cap K = V$ and so $W$ defines the desired extension. ■

To make the valuations more tractable we shall require an abelian value group and commutative residue class field. It is convenient to impose an even stronger condition, as in the next result:

PROPOSITION 4.6 *Let $K$ be a skew field with a valuation $v$, having valuation ring $V$, maximal ideal $\mathfrak{m}$ and group of 1-units $U_1$. Then the following conditions are equivalent:*

(a) $K^\times/U_1$ *is abelian,*

(b) $K^c \subseteq 1 + \mathfrak{m} = U_1$,

(c) $v(1-a) > 0$ *for all* $a \in K^c$.

*Moreover, when (a)–(c) hold, then the value group and residue class field are commutative.*

The equivalence proof is an almost immediate consequence of the definitions, and the last sentence is clear from a glance at the diagram (3). ■

A valuation satisfying the equivalent conditions of Prop. 4.6 will be called *quasi-commutative*. The following criterion for quasi-commutative extensions is an easy consequence:

THEOREM 4.7 *Let $K$ be a skew field with a quasi-commutative valuation $v$, and let $L$ be an extension field of $K$. Then $v$ can be extended to a quasi-commutative valuation of $L$ if and only if there is no equation in $L$ of the form*

$$\sum a_i p_i + \sum b_j(q_j - 1) = 1, \tag{12}$$

*where $a_i, b_j \in K$, $v(a_i) > 0$, $v(b_j) \geqslant 0$, $p_i, q_j \in L^c$.*

*Proof.* If there is a quasi-commutative extension $w$ of $v$ to $L$, then for $a_i, b_j, p_i, q_j$ as above we have

$$w(\sum a_i p_i + \sum b_j(q_j - 1)) \geqslant \min\{v(a_i) + w(p_i), v(b_j) + w(q_j - 1)\}$$
$$> 0,$$

because $v(a_i) > 0$, $w(q_j - 1) > 0$. It follows that no equation of the form (12) can hold. Conversely, assume that there is no equation (12) and consider the set $\mathfrak{Q}$ of all expressions $\sum a_i p_i + \sum b_j(q_j - 1)$, where $a_i, b_j, p_i, q_j$ are as before. It is clear that $\mathfrak{Q}$ is closed under addition and contains the maximal ideal corresponding to the valuation $v$. Moreover, $\mathfrak{Q}$ is invariant in $L$, i.e. $u^{-1}\mathfrak{Q}u = \mathfrak{Q}$ for $u \in L^\times$, because $u^{-1}a_i p_i u = a_i \cdot a_i^{-1}u^{-1}a_i u \cdot u^{-1}p_i u \in VL^c$, and similarly for the other terms. In the same way we can verify that $\mathfrak{Q}$ admits multiplication. We define

$$W = \{a \in L \mid a\mathfrak{Q} \subseteq \mathfrak{Q}\}. \tag{13}$$

It is clear that $W$ is a subring of $L$ containing $L^c$ and the valuation ring $V$ of $v$, for if $c \in V$ and we multiply the expression on the left of (12) by $c$, we obtain $\sum ca_i p_i + \sum cb_j(q_j - 1)$; this is of the same form, since $v(ca_i) = v(c) + v(a_i) > 0$ and $v(cb_j) \geqslant 0$. Moreover, $\mathfrak{Q}$ is an ideal in $W$, for we have $c\mathfrak{Q} \subseteq \mathfrak{Q}$ for all $c \in W$, by the definition of $W$, and $\mathfrak{Q}c = c \cdot c^{-1}\mathfrak{Q}c = c\mathfrak{Q} \subseteq \mathfrak{Q}$. Since $1 \notin \mathfrak{Q}$ by hypothesis, $\mathfrak{Q}$ is a proper ideal of $W$. Thus $W$ is a subring of $L$ containing $L^c$ and the valuation ring of $v$. If $(T, \mathfrak{p})$ is a maximal pair dominating $(W, \mathfrak{Q})$, then by Lemma 4.5, $T$ is a valuation ring and $T \supseteq W \supseteq L^c$, while $1 + \mathfrak{p} \supseteq 1 + \mathfrak{Q} \supseteq L^c$. Hence the valuation $w$ defined by $T$ extends $v$ and is quasi-commutative, by Prop. 4.6. ■

Let $D$ be a skew field with centre $C$ and let $X$ be any set. If $D$ has a quasi-commutative valuation $v$, one can use Th. 4.7 to extend $v$ to a quasi-commutative valuation of the free field $D_C\langle\langle X \rangle\rangle$, but this requires more detail on how free fields are formed. In essence one uses the specialization lemma to show that if an equation (12) exists, then $X$ can be specialized to values in $D$ so as to yield an equation (12) in $D$, which is a contradiction (cf. Cohn (1987), (1989)).

## Exercises

(1) Show that any total subring of a field is a local ring.

(2) A *place* of a field $K$ in another, $L$, is defined as a mapping $f: K \to L \cup \{\infty\}$ such that $f^{-1}(L) = V$ is an invariant subring of $K$, the restriction $f|V$ is a homomorphism and $xf = \infty$ implies $x \neq 0$ and $(x^{-1})f = 0$. Show that $V$ is a valuation ring and that conversely, every valuation ring $V$ on $K$ leads to a place of $K$ in the residue class field of $V$. Define a notion of equivalence of places and show that there is a natural bijection between valuation rings on $K$ and equivalence classes of places.

(3) Verify that a valuation on a field $K$ has a value group $\mathbf{Z}$ iff its valuation ring is a PID.

In the following exercises a valuation is said to be *over* $K$, if its restriction to $K$ is the trivial valuation.

(4) Let $F$ be the tensor $D$-ring on $X$ and $U$ its universal field of fractions. Form $U(t)$ with a central indeterminate $t$ and define a homomorphism $\lambda: U \to U(t)$ as the identity on $D$ and mapping $x \in X$ to $xt$. Let $v_0$ be the $t$-adic valuation on $U(t)$ over $U$ and put

$v(p) = v_0(p\lambda)$ for $p \in U$. Verify that $v$ is a valuation on $U$; find its value group and residue class field.

(5) Let $K$ be a field with a valuation $v$ whose value group $\Gamma$ is a subgroup of $\mathbf{R}$. Define an extension of $v$ to the rational function field $K(x)$ by (7), where $\delta \in \mathbf{R}, \delta > 0$, and find the new value group and residue class field. Distinguish the cases $\Gamma \cap \delta\mathbf{Z} = 0, \Gamma \cap \delta\mathbf{Z} \neq 0$.

(6) Let $E$ be the field of fractions of the Weyl algebra on $k$ (generated by $u, v$ with $uv - vu = 1$), where $\operatorname{char} k = 0$. Writing $t = u^{-1}$, verify that $vt = t(v + t)$. Show that the $t$-adic valuation on $E$ over $k(v)$ is quasi-commutative.

## 11.5 Pseudo-linear extensions

We recall that (for any ring $K$) a $K$-ring $A$ is essentially a ring $A$ with a homomorphism $K \to A$. If $K$ is a field, this means that $A$ contains a copy of $K$ except when $A = 0$. Every $K$-ring $A$, for a field $K$, may be regarded as a left or right $K$-module; we shall denote the corresponding dimensions by $[A:K]_L$ and $[A:K]_R$ and note that they need not be equal, even when $A$ itself is a field. Below we shall give an example of a field extension in which the right dimension is two and the left dimension is infinite (Artin's problem, cf. Cohn (1961)). For examples in which both dimensions are finite but different see Schofield (1985).

We note the dimension formula familiar from the commutative case; the proof is essentially the same as in that case (cf. A.2, Th. 4.1.1).

PROPOSITION 5.1 *Let $D \subseteq E$ be skew fields and $V$ a left $E$-space. Then*

$$[V:D] = [V:E][E:D]_L,$$

*whenever either side is finite.* ■

For any skew field extension $E/D$ the left and right dimensions are easily seen to be equal if $D$ is finite-dimensional over its centre. The same conclusion still holds if $E$ is finite-dimensional over its centre:

PROPOSITION 5.2 *Let $D \subseteq E$ be skew fields and assume that $E$ is finite-dimensional over its centre. Then*

$$[E:D]_L = [E:D]_R. \tag{1}$$

*Proof.* Let $C$ be the centre of $E$, so that $E$ is a finite-dimensional $C$-algebra. Write $A = DC = \{\sum x_i y_i | x_i \in D, y_i \in C\}$; then $A$ is a subring containing $C$ in its centre, hence a $C$-algebra and also a $D$-ring. Since it is generated by $C$ over $D$, we may choose a basis of $A$, as left $D$-space, consisting of elements of $C$. This is also a basis of $A$ as right $D$-space, because $C$ is the centre of $E$, so

$$[A:D]_L = [A:D]_R. \tag{2}$$

Now $A$ is a subalgebra of the division algebra $E$, so $A$ is also a skew field and by Prop. 5.1,

$$[E:C] = [E:A]_L[A:C] = [E:A]_R[A:C]. \tag{3}$$

Since $[E:C]$ is finite, so is $[A:C]$; dividing (3) by $[A:C]$, we find that $[E:A]_L = [E:A]_R$. If we now multiply by (2) and use Prop. 5.1 and its left–right dual, we obtain the required formula (1). ■

The study of finite-dimensional field extensions is greatly complicated by the lack of commutativity. Let us consider the simplest case, of a quadratic extension $E/D$, $[E:D]_R = 2$. For any $u \in E \setminus D$ the pair $1, u$ is a right $D$-basis of $E$. Thus every element of $E$ can be uniquely expressed in the form $ua + b$, where $a, b \in D$. In particular we have

$$cu = uc^\alpha + c^\delta \qquad \text{for all } c \in D, \tag{4}$$

and

$$u^2 + u\lambda + u = 0 \qquad \text{for certain } \lambda, \mu \in D. \tag{5}$$

Here $c^\alpha, c^\delta$ are uniquely determined by $c$ and a calculation as in **9.3** shows $\alpha$ to be an endomorphism of $D$ and $\delta$ an $\alpha$-derivation. Moreover, the structure of $E$ is completely determined by $D$ and (4), (5).

Conversely, if $D$ is any field with an endomorphism $\alpha$ and an $\alpha$-derivation $\delta$, then for given $\lambda, \mu \in D$ it is possible to write down necessary and sufficient conditions for a quadratic extension of $D$ to be defined by (4) and (5) (cf. Ex. (3)).

Generalizing the above discussion, we may define a *pseudo-linear* extension of right dimension $n$, with generator $u$, as an extension field $E$ of $D$ with right $D$-basis $1, u, u^2, \ldots, u^{n-1}$ such that (4) holds, and in place of (5),

$$u^n + u^{n-1}\lambda_1 + \cdots + \lambda_n = 0, \qquad \text{for certain } \lambda_i \in D. \tag{6}$$

What has been said shows that every extension of right dimension 2 is pseudo-linear; in higher dimensions the pseudo-linear extensions form a special class. We note the following formula for the left dimension:

PROPOSITION 5.3 *Let $E/D$ be a pseudo-linear extension of right dimension $n$, with commutation formula* (4). *Then*

$$[E:D]_L = 1 + [D:D^\alpha]_L + [D:D^\alpha]_L^2 + \cdots + [D:D^\alpha]_L^{n-1}. \tag{7}$$

*In particular, any pseudo-linear extension $E/D$ satisfies*

$$[E:D]_L \geqslant [E:D]_R,$$

*with equality if and only if $\alpha$ is an automorphism of $D$.*

*Proof.* Take a generator $u$ of $E/D$ and write $E_0 = D$, $E_i = uE_{i-1} + D$ $(i \geqslant 1)$.

Then by induction on $r$ we have

$$E_r = D + uD + \cdots + u^r D.$$

Moreover, each $E_i$ is a left $D$-module, by (4), so we have a tower of left $D$-modules

$$D = E_0 \subset E_1 \subset \cdots \subset E_{n-1} = E,$$

and (7) will follow if we prove

$$[E_r/E_{r-1}:D]_L = [D:D^\alpha]_L^r. \tag{8}$$

Let $\{e_\lambda | \lambda \in I\}$ be a left $D^\alpha$-basis for $D$. We claim that the elements

$$u^r e_{\lambda_{r-1}}^{\alpha^{r-1}} \cdots e_{\lambda_1}^{\alpha} e_{\lambda_0} \tag{9}$$

where $\lambda_0, \lambda_1, \ldots, \lambda_{r-1}$ range independently over $I$, form a basis of $E_r \pmod{E_{r-1}}$. This will prove (8) and hence (7).

Any $c \in K$ can be written as a linear combination of the $e_\lambda$ with coefficients in $K^\alpha$, say $c = \sum c_{\lambda_0}^{\alpha} e_{\lambda_0}$. If we repeat the process on $c_{\lambda_0}$ we obtain $c_{\lambda_0} = \sum c_{\lambda_0\lambda_1}^{\alpha} e_{\lambda_1}$. Hence

$$c = \sum c_{\lambda_0\lambda_1}^{\alpha^2} e_{\lambda_1}^{\alpha} e_{\lambda_0},$$

and after $r$ steps we find

$$c = \sum c_{\lambda_0 \cdots \lambda_{r-1}}^{\alpha^r} e_{\lambda_{r-1}}^{\alpha^{r-1}} \cdots e_{\lambda_0}.$$

Therefore

$$\begin{aligned} u^r c &\equiv \sum u^r c_{\lambda_0 \cdots \lambda_{r-1}}^{\alpha^r} e_{\lambda_{r-1}}^{\alpha^{r-1}} \cdots e_{\lambda_0} \\ &\equiv \sum c_{\lambda_0 \cdots \lambda_{r-1}} u^r e_{\lambda_{r-1}}^{\alpha^{r-1}} \cdots e_{\lambda_0} \qquad (\mathrm{mod}\, E_{r-1}). \end{aligned}$$

Hence the elements (9) span $E_r \pmod{E_{r-1}}$. To prove their linear independence, assume that

$$\sum c_{\lambda_0 \cdots \lambda_{r-1}} u^r e_{\lambda_{r-1}}^{\alpha^{r-1}} \cdots e_{\lambda_0} \equiv 0 \qquad (\mathrm{mod}\, E_{r-1}).$$

We obtain

$$\sum u^r c_{\lambda_0 \cdots \lambda_{r-1}}^{\alpha^{r-1}} e_{\lambda_{r-1}}^{\alpha^{r-1}} \cdots e_{\lambda_0} \equiv 0 \qquad (\mathrm{mod}\, E_{r-1}).$$

Since the $e_\lambda$ are left linearly independent over $D^\alpha$, we can equate the coefficients of $e_{\lambda_0}$ to 0 and using induction on $r$ we find that $c_{\lambda_0 \cdots \lambda_{r-1}} = 0$. ■

In order to obtain a quadratic extension satisfying (4) and (5) let us assume that $\lambda = 0$, $\alpha\delta + \delta\alpha = 0$, and that $\delta^2$ is the inner $\alpha^2$-derivation induced by $-\mu$. Further assume that $\mu^\alpha = \mu$, $\mu^\delta = 0$ and that $D$ contains no element $a$ satisfying

$$a \cdot a^\alpha + a^\delta + \mu = 0. \tag{10}$$

We form the skew polynomial ring $R = D[t; \alpha, \delta]$ and consider $f = t^2 + \mu$.

For any $c \in D$ we have

$$\begin{aligned} cf &= c(t^2 + \mu) = (tc^\alpha + c^\delta)t + c\mu \\ &= t^2 c^{\alpha^2} + tc^{\alpha\delta} + tc^{\delta\alpha} + c^{\delta^2} + c\mu \\ &= t^2 c^{\alpha^2} + \mu c^{\alpha^2} - c^{\delta^2} + c^{\delta^2} \\ &= fc^{\alpha^2}. \end{aligned}$$

Hence $fR$ is a two-sided ideal. Moreover, $f$ is irreducible, for if $f$ could be factorized, we would have a product of two linear factors which may both be taken monic, without loss of generality. Thus

$$t^2 + \mu = (t - a)(t - b) = t^2 - t(a^\alpha + b) - a^\delta + ab.$$

Therefore $b = -a^\alpha$ and $aa^\alpha + a^\delta + \mu = 0$, but this contradicts the fact that (10) has no solution in $D$. It follows that $E = R/fR$ is an integral domain of right dimension 2 over $D$, hence a field. Its right dimension is 2, since it has the right $D$-basis $1, u$, where $u$ is the residue class of $t$, while the left dimension is $1 + [D:D^\alpha]_L$, and this will be greater than 2 provided that $\alpha$ is not an automorphism.

As an example let us take any commutative field $k$, put $E = k\langle\!\langle x, y \rangle\!\rangle$, the free field on $x$ and $y$, and take $D$ to be the subfield generated over $k$ by $x, y^2$ and $xy - yx^2$. On $E$ we have a $k$-linear endomorphism $\alpha: x \mapsto x^2$, $y \mapsto -y$. To show this one has either to verify directly that $\alpha$ as endomorphism of $k\langle x, y\rangle$ is honest, or show that $E$ admits an extension in which $x$ has a square root. For then, by iterating the process one obtains a field $P$ containing $E$ in which $x$ has a $2^n$th root for all $n \geqslant 0$; hence the mapping $x \mapsto x^2$, $y \mapsto -y$ is an automorphism of $P$ and the restriction to $E$ is the required endomorphism. This is fairly plausible, but we shall not give a formal proof here (cf. Cohn [1977] **5.6**).

It is easily checked that $\alpha$ maps $D$ into itself; moreover $D$ admits the inner $\alpha$-derivation $\delta: a \mapsto ay - ya^\alpha$ induced by $y$. If we can show that $y \notin D$ and that $\alpha$ is not an automorphism of $D$, we have a quadratic extension $E/D$ with left dimension $> 2$; in fact we shall find that $[E:D]_L = \infty$.

Consider the $(\alpha, 1)$-derivation $\gamma$ on $E$ such that $x^\gamma = 0, y^\gamma = 1$. We have $(ab)^\gamma = a^\gamma b + a^\alpha b^\gamma$, by definition, hence $x^\gamma = 0$, $(y^2)^\gamma = y + y^\alpha = 0$, $(xy - yx^2)^\gamma = x^2 - x^2 = 0$. Hence $\gamma$ vanishes on $D$, but $y^\gamma = 1$, so $y \notin D$.

Finally to show that $[E:D]_L = \infty$, we first note that if $[D:D^\alpha]_L = n$, then $[E:E^\alpha]_L$ is finite. For let $u_1, \ldots, u_n$ be a left $D^\alpha$-basis for $D$. We claim that $u_i, u_i u_j, u_i y u_j$ span $E$ as left $E^\alpha$-space. Any $p \in E$ has the form $p = a + yb$, where $a, b \in D$, say $a = \sum a_i^\alpha u_i$, $b = \sum b_i^\alpha u_i$; hence $p = \sum a_i^\alpha u_i + y\sum b_i^\alpha u_i = \sum a_i^\alpha u_i + \sum b_i y u_i - \sum b_i^\delta u_i = \sum a_i^\alpha u_i + \sum b_{ij}^\alpha u_j y u_i - \sum c_{ij}^\alpha u_j u_i$, for suitable $b_{ij}, c_{ij} \in D$, This proves our claim, in particular it shows that $[E:E^\alpha]_L \leqslant 2n^2 + n$, if $[D:D^\alpha] = n$. Now $E^\alpha = k\langle\!\langle x^2, y \rangle\!\rangle$ and so the elements $xy^r$ $(r = 1, 2, \ldots)$ are left $E^\alpha$-independent. This is intuitively plausible and can be proved with the methods of Cohn

[1977], Lemma 5.5.5. It follows that $[E:E^\alpha]_L = \infty$; hence $[D:D^\alpha]_L = \infty$ and by Prop. 5.3, $[E:D]_L = \infty$.

## Exercises

(1) Let $E/D$ be a skew field extension. Show that if $D$ is finite-dimensional over its centre, then $[E:D]_L = [E:D]_R$.

(2) Verify that every cyclic division algebra is a pseudo-linear extension of a maximal commutative subfield.

(3) Show that (4) and (5) define an algebra of right dimension 2 over $D$ iff $c^{\alpha\delta} + c^{\delta\alpha} = \lambda c^{\alpha^2} - c^\alpha\lambda$, $c^{\delta^2} + c^\delta\lambda = \mu c^{\alpha^2} - c\mu$, $\lambda^\delta = \mu - \mu^\alpha - \lambda(\lambda - \lambda^\alpha)$, $\mu^\delta = \mu(\lambda^\alpha - \lambda)$, and this extension is a field iff $c\cdot c^\alpha + c\lambda + \mu \neq 0$ for all $c \in D$.

## Further exercises on Chapter 11

(1) Let $D$ be a skew field with centre $C$ and $k$ a subfield of $C$ such that $C$ is algebraic over $k$. Show that if $D$ is finite-dimensional over $C$ and finitely generated as $k$-algebra then $[D:k]$ is finite.

(2) Let $P = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ be a square matrix written in block form, where $A$ is invertible. Show that $\delta(P) = \delta(A)\cdot\delta(D - CA^{-1}B)$.

(3) Show that over a local ring, $A = LUP$, where $L$ is lower unitriangular, $U$ is upper triangular and $P$ is a permutation matrix.

(4) Show that the core of a matrix $A$ (over a field) has diagonal form iff each principal minor of $A$ is invertible, and when this is so, then the factors $L$ and $U$ in (8) (as well as $M$) are unique.

(5) Let $F = D_C\langle X\rangle$ be the tensor ring and $G = D_C\langle\langle X\rangle\rangle$ the corresponding power series ring, and with a central indeterminate $t$ form the power series ring $F[\![t]\!]$. Show that $G$ can be embedded in $F[\![t]\!]$ by mapping $x \mapsto xt$ for $x \in X$ and leaving $D$ fixed.

Given any full matrix $A = A(x)$ over $F$, show that $A(xt)$ is invertible over $U((t))$, where $U$ is the universal field of fractions of $F$. Deduce that $A(xt)$ is full over $F[\![t]\!]$, and hence that $A(x)$ is full over $G$. (This proof that the embedding $F \to G$ is honest yields a special case of the inertia theorem, cf. Cohn [1985], Th. 2.9.15.)

(6) Let $A$ be a full matrix over $k\langle X\rangle$. Show that there exists $n = n(A)$ such that for every central division algebra $D$ over $k$ of degree at least $n$, $A$ is non-singular for some set of values of $X$ in $D$.

(7) Let $K$ be a field with a valuation $v$, Given any matrix $A$ over $K$, show that

$A = PLDUQ$, where $P, Q$ are signed permutation matrices, $L$ is lower and $U$ upper unitriangular and $D = \mathrm{diag}(d_1, \ldots, d_r)$ a diagonal matrix with $v(d_1) \leqslant v(d_2) \leqslant \cdots$. Show that if $A$ is square and $v$ is abelian, then $v(\delta(A)) = \sum v(d_i)$.

(8) Show that in a quadratic extension $E/D$, subject to (4) and (5), $\alpha$ may be extended to an endomorphism $\alpha'$ of $E$ by $u^{\alpha'} = -u - \lambda$, and $\delta$ is then the inner $\alpha'$-derivation induced by $u$. Verify that $\alpha'$ is an automorphism of $E$ iff $\alpha$ is an automorphism of $D$.

# Bibliography

This is primarily a list of books where the topics are pursued further (and which were often used as sources). A second list contains articles having a bearing on the text as well as those referred to in the text. References in the text are by name and date, the latter enclosed in square brackets for books and round brackets for papers.

## I. Books


Albert, A. A. 1939, *Structure of Algebras*, AMS Colloquium publications No. 24, Providence, RI.

Artin, E. 1957, *Geometric Algebra*, Interscience, New York.

Artin, E. 1965, *Collected Papers*, Addison-Wesley, Reading, Mass.

Barbilian, D. 1956, *Teoria Aritmetica a Idealilor* (*în inele necomutative*), Ed. Acad. Rep. Pop. Romine, Bucharest.

Barwise, J. (Ed.) 1977, *Handbook of Mathematical Logic*, North-Holland, Amsterdam.

Bass, H. 1962, *The Morita Theorems*, Oregon Lectures.

Bass, H. 1968, *Algebraic K-theory*, Benjamin, New York.

Bell, J. L. and Slomson, A. B. 1971, *Models and Ultraproducts: An Introduction*, North-Holland, Amsterdam.

Bourbaki, N. 1974, *Algebra* I, Ch. 1–3, Addison-Wesley, Reading, MA.

Bourbaki, N. 1990, *Algebra* II, Ch. 4–7, Springer, Heidelberg.

Burnside, W. 1911, *Theory of Groups of Finite Order*, 1955 Dover, New York

Chevalley, C. 1951, *Introduction to the Theory of Algebraic Functions of One Variable*. Amer. Math. Soc., New York.

Cohn, P. M. 1966, *Morita Equivalence and Duality*, Queen Mary College Math. Notes.

Cohn, P. M. 1977, *Skew Field Constructions*, LMS Lecture Notes No. 27, Cambridge University Press.

Cohn, P.M. 1981, *Universal Algebra* (2nd edn), Reidel, Dordrecht.

Cohn, P. M. 1985, *Free Rings and their Relations* (2nd edn) LMS Monographs No. 19, Academic Press, London.

Cohn, P. M. 1991, *Algebraic Numbers and Algebraic Functions*, to appear.

Dickson, L. E. 1901, *Linear Groups, with an Exposition of the Galois Field Theory*, 1958 Dover, New York.

Draxl, P. K. 1983, *Skew Fields*, LMS Lecture Notes No. 83, Cambridge University Press.

Faith, C. 1981, *Algebra I: Rings, Modules and Categories*, Grundlehren d. Math. Wiss 190, Springer, Heidelberg.

Goodearl, K. R. and Warfield, Jr., R. B. 1989, *An Introduction to Non-commutative Noetherian Rings*, LMS Student Texts No. 16, Cambridge University Press.
Gorenstein, D. 1982, *Finite Simple Groups*, Plenum, New York.
Hall, Jr. M. 1959, *The Theory of Groups*, Macmillan, New York.
Hartshorne, R. 1977, *Algebraic Geometry*, Grad. Texts in Math. 52, Springer, Heidelberg.
Herstein, I. N. 1968, *Noncommutative Ring Theory*, Carus Math. Monographs No. 15, Math. Association of America.
Herstein, I. N. 1976, *Rings with Involution*, Chicago Lectures in Math., Chicago University Press.
Hilbert, D. 1897, *Bericht über die Theorie der algebraischen Zahlkörper*, Jahrber, DMV iv, reprinted in Vol. 1 of Collected Works.
Hilton, P. J. and Stammbach, U. 1971, *A Course in Homological Algebra*, Grad. Texts in Math. 4, Springer, Heidelberg.
Jacobson, N. 1953, *Lectures in Abstract Algebra* I–III (1951, 1953, 1964), van Nostrand, Princeton, NJ.
Jacobson, N. 1964, *Structure of Rings* (2nd edn) AMS Colloq. Publs. No. 37, Providence, RI.
Jacobson, N. 1985, 1989, *Basic Algebra* I–II (2nd edn) Freeman, San Francisco.
Jategaonkar, A. V. 1986, *Localization in Noetherian Rings*, LMS Lecture Notes No. 98, Cambridge University Press.
Klein, F. 1884, *Lectures on the Icosahedron*, 1956 Dover, New York.
Lam, T. Y. 1978, *Serre's Conjecture*, Springer Lecture Notes in Math. No. 635, Springer, Berlin.
Lam, T. Y. 1980, *The Algebraic Theory of Quadratic Forms* Adv. Book Progr. Benjamin/Cummings Reading, MA.
Lang, S. 1984, *Algebra* (2nd edn) Addison-Wesley, Reading, MA.
Mac Lane, S. 1963, *Homology*, Grundlehren d. Math. Wiss. 114, Springer, Heidelberg.
McConnell, J. C and Robson, J. C. 1987, *Non-commutative Noetherian Rings*, Wiley, Chichester.
Mitchell, B. 1965, *Theory of Categories*, Academic Press, New York.
Neukirch, J. 1986, *Class Field Theory*, Grundl. d. Math. Wiss. 280, Springer, Heidelberg.
Pierce, R. S. 1982, *Associative Algebras*, Grad. Texts in Math. 88, Springer, Heidelberg.
Procesi, C. 1973, *Rings with Polynomial Identities*, Dekker, New York.
Reiner, I. 1976, *Maximal Orders*, LMS Monographs No. 5, Academic Press, London.
Rowen, L. H. 1980, *Polynomial Identities in Ring Theory*, Academic Press, New York.
Rowen, L. H. 1988, *Ring Theory* I, II, Academic Press, New York.
Scharlau, W. 1985, *Quadratic and Hermitian Forms*, Grundl. d. Math. Wiss. 270, Springer, Heidelberg.
Schofield, A. H. 1985, *Representations of Rings over Skew Fields*, LMS Lecture Notes No. 92, Cambridge University Press.
Semple, J. G. and Roth, L. 1949, *Introduction to Algebraic Geometry* (New Printing 1987) Clarendon Press, Oxford.
Serre, J.-P. 1979, *Local Fields*, Grad. Texts in Math. 67, Springer, Heidelberg.
Stroyan, K. D. and Luxemburg, W. A. J. 1976, *Introduction to the Theory of Infinitesimals*, Academic Press, New York.
Weber, H. 1894, 1896, 1908, *Lehrbuch der Algebra*, I–III (1960 reprint), Chelsea, New York.
Weil, A. 1967, *Basic Number Theory*, Grundl. d. Math. Wiss. 144, Springer, Heidelberg.
Welsh, D. J. A. 1976, *Matroid Theory*, LMS Monographs No. 8, Academic Press, London.
Weyl, H. 1946, *The Classical Groups*, Princeton University Press.
White, N. (ed.) 1986, *Theory of Matroids, Encyclopedia of Maths. and its Applications*, No. 26, Cambridge University Press.

## II. Papers

Amitsur, S. A. 1965, Generalized polynomial identities and pivotal polynomials, *Trans. Amer. Math. Soc.* **114** 210–226.

Amitsur, S. A. 1972, On central division algebras, *Isr. J. Math.* **12** 408–420.

Auslander, M. and Goldman, O. 1960, The Brauer group of a commutative ring, *Trans. Amer. Math. Soc.* **97** 367–409.

Azumaya, G. 1950, Corrections and supplementaries to my paper concerning Krull-Schmidt's theorem, *Nagoya Math. J.* **1** 117–124.

Bass, H. 1960, Finitistic dimension and a generalization of semiprimary rings, *Trans. Amer. Math. Soc.* **95** 466–488.

Bergman, G. M. 1974, Modules over coproducts of rings, *Trans. Amer. Math. Soc.* **200** 1–32.

Bergman, G. M. 1974′, Coproducts and some universal ring constructions, *Trans. Amer. Math. Soc.* **200** 33–88.

Bergman, G. M. 1978, The diamond lemma for ring theory, *Adv. in Math.* **29** 178–218.

Chase, S. U. 1960, Direct products of modules, *Trans. Amer. Math. Soc.* **97** 457–473.

Cohn, P. M. 1961, Quadratic extensions of skew fields, *Proc. London Math. Soc.* (3) **11** 531–556.

Cohn, P. M. 1966, On the structure of the $\mathbf{GL}_2$ of a ring, *Publ. Math. IHES* **30** 5–53.

Cohn, P. M. 1973, Unique factorization domains, *Amer. Math. Monthly* **80** 1–17.

Cohn, P. M. 1987, Valuations on free fields, in '*Algebra, some current trends*', *Proc. Varna 1986*, ed. L. L. Avramov, K. B. Tchakerian, Springer Lecture Notes in Math. No. 1352, 75–87.

Cohn, P. M. 1989, The construction of valuations on skew fields, *J. Indian Math. Soc.*, **54**, 1–45.

Deligne, P. R. 1973, Variétés unirationnelles non rationnelles, *Sem. Bourbaki 1971/72*, Exp. 402, Springer Lecture Notes in Math. 317.

Dieudonné, J. 1943, Les déterminants sur un corps non-commutatif, *Bull. Soc. Math. France* **71** 27–45.

Goldie, A. W. 1958, The structure of prime rings under ascending chain conditions, *Proc. London Math. Soc.* (3) **8** 589–608.

Henkin, L. 1960, On mathematical induction, *Amer. Math. Monthly* **67** 323–338

Lenstra, jr. H. W. 1974, Rational functions invariant under a finite abelian group, *Invent. Math.* **25** 299–325.

Merkuryev, A. S. and Suslin, A. A. 1986, On the structure of the Brauer groups of fields, *Math. USSR Izvestiya* **27** 141–155.

Nagata, M. 1957, A remark on the unique factorization theorem, *J. Math. Soc. Japan* **9** 143–145.

Ore, O. 1931, Linear equations in non-commutative fields, *Ann. Math.* **32** 463–477.

Roganov, Yu. V. 1975, The dimension of a tensor product on a projective bimodule (Russian), *Mat. Zametki* **18** 895–902.

Sasiada, E. and Cohn, P. M. 1968, An example of a simple radical ring, *J. Algebra* **5** 373–377.

Schofield, A. H. 1985, Artin's problem for skew field extensions, *Math. Proc. Camb. Phil. Soc.* **97** 1–6.

Sierpiński, W. 1945, Sur les fonctions de plusieurs variables, *Fund. Math.* 7, **33** 169–173.

Swan, R. G. 1969, Invariant rational functions and a problem of Steenrod, *Invent. Math.* **7** 148–158.

Voskresenskii, V. E. 1973, Fields of invariants of abelian groups (Russian), *Uspekhi. Mat. Nauk SSSR* **28** 77–102.

Webber, D. B. 1970, Ideals and modules in simple Noetherian hereditary rings, *J. Algebra* **16** 239–242.

# List of notations

The number indicates the page where the term is first used or defined. Terms occurring in Vol. 1 or 2 are indicated by I, II respectively, instead of a page reference.

**Number systems**

| | |
|---|---|
| $\mathbf{N}$ | the natural numbers, I |
| $\mathbf{N}_0$ | the natural numbers together with 0, II |
| $\mathbf{Z}$ | the integers, I |
| $\mathbf{Q}$ | the rational numbers, I |
| $\mathbf{R}$ | the real numbers, I |
| $\mathbf{C}$ | the complex numbers, I |
| $\mathbf{H}$ | the real quaternions, 280 |
| $\mathbf{K} = \mathbf{Q}/\mathbf{Z}$ | the rational numbers mod 1, 132 |
| $\mathbf{Z}/m$ | the integers mod $m$, I |
| $\mathbf{Z}(p^\infty)$ | the group of all $p^n$th roots of 1 for $n = 1, 2, \ldots$, II |
| $\mathbf{U}_m$ | the group of $m$th roots of unity, II |
| $\mathbf{U}(n)$ | the group of units mod $n$, I |
| $\mathbf{F}_q$ | the field of $q$ elements, II |
| $\mathbf{Z}_p$ | the ring of $p$-adic integers, II |
| $\hat{\mathbf{Z}} = \varprojlim \mathbf{Z}/n$, | the profinite completion of $\mathbf{Z}$, 201 |

**Set theory**

| | |
|---|---|
| $\varnothing$ | the empty set, I |
| $\lvert X \rvert$ | the cardinal of a set $X$, I |
| $\mathscr{P}(X)$ | the power set (set of all subsets) of $X$, II |
| $Y'$ or $X \setminus Y$ | the complement of $Y$ in $X$, I, 28 |
| $Y^X$ | the set of all mappings from $X$ to $Y$, I |
| $\aleph_0$ | aleph-null, the cardinal of $N$, II |

**Number theory and field theory**

| | |
|---|---|
| $\max(a, b)$ | the larger of $a$ and $b$, I |
| $\min(a, b)$ | the smaller of $a$ and $b$, I |
| $a \mid b$ | $a$ divides $b$, I |
| $[x]$ | the greatest integer not exceeding $x$, II |
| $\delta_{ij}$ | Kronecker delta, I |
| $\varphi(m)$ | Euler function, I |
| $\wp(x) = x^p - x$, | 209 |

**Groups**

| | |
|---|---|
| $\mathrm{Sym}_n$ or $S_n$ | symmetric group of degree $n$, I |
| $\mathrm{Alt}_n$ or $A_n$ | alternating group of degree $n$, I |
| $\mathrm{sgn}(\sigma)$ | sign of the permutation $\sigma$, I |
| $\mathbf{C}_n$ | cyclic group of order $n$, I |
| $N \lhd G$ | $N$ is a normal subgroup of $G$, I |
| $(G:H)$ | index of $H$ in $G$, I |
| $\mathbf{GL}_n(R)$ | general linear group of a ring $R$, I |
| $\mathbf{SL}_n(R)$ | special linear group of $R$, II |
| $\mathbf{O}(V)$, $\mathbf{SO}(V)$, | the orthogonal (special orthogonal) group on a quadratic space $V$, I |
| $\mathbf{Sp}_{2m}(k)$ | the symplectic group over $k$, I |
| $\mathbf{P}^n(k)$ | projective $n$-space over $k$, 162 |
| $PL$ | for any linear group $L$, the corresponding projective linear group (i.e. the quotient by the centre), 162 |
| $G \ltimes A$ | semidirect product, II |
| $H_n(A, M)$ | homology group, 129 |
| $H^n(A, M)$ | cohomology group, 129 |
| $\mathrm{Aut}(G)$ | automorphism group, II |
| $\mathrm{Inn}(G)$ | group of inner automorphisms, II |
| $\mathrm{Der}(G, A)$ | derivation module, 131 |
| $\Phi(G)$ | Frattini subgroup, 135 |
| $F(G)$ | Fitting subgroup, 139 |
| $\gamma_i(G)$ | lower central series, 157 |
| $Z_i(G)$ | upper central series, 157 |
| $\mathrm{St}_p$ | stabilizer of $p$, I |
| $B_{ij}(c)$ | elementary matrix, 160 |
| $\mathrm{Gal}(E/k)$ | Galois group, II |
| $\mathbf{K}_1(A), \mathbf{SK}_1(A)$ | Whitehead group, 276 |

**Rings, modules and categories**

| | |
|---|---|
| $A^T$ | transpose of a matrix $A$, I |
| ${}^mV^n$ | the space of $m \times n$ matrices over $V$, I |
| ${}^mV = {}^mV^1$ | column vectors, I |
| $V^n = {}^1V^n$ | row vectors, I |
| $\mathfrak{M}_n(R)$ or $R_n$ | $n \times n$ matrix ring over $R$, I |
| $\mathscr{M}_R$ | category of right $R$-modules, II |
| $\mathrm{Lat}(M)$ | Lattice of submodules of $M$, II |
| $\mathrm{Hom}(U, V)$ | group of all homomorphisms from $U$ to $V$, II |
| $\mathrm{End}(U)$ | ring of all endomorphisms of $U$, II |
| $U \oplus V, \oplus U_i$ | direct sum, I |
| $U \otimes V$ | tensor product ($\hat{\otimes}$ graded), II |
| $\prod M_i$ | product, 70 |
| $\coprod M_i$ | coproduct, 70 |
| $\coprod M_i$ | biproduct, 71 |
| $R^+$ | additive subgroup of $R$, I |
| $R^\times$ | set of non-zero elements of $R$, II |
| $R_+$ | in an ordered ring $R$, the set of all elements $\geqslant 0$, II (also: augmentation ideal in a graded ring $R$), 42 |
| $\mathbf{U}(R)$ | group of units in $R$, II |
| $k[X]$ | polynomial ring over $k$, I |
| $k[\![X]\!]$ | power series ring over $k$, II |
| $k(X)$ | rational function field over $k$, II |
| $k((X))$ | field of formal Laurent series over $k$, II |
| $D_C\langle X\rangle$ | tensor ring over $D$, 44, 46 |
| $D_C \langle\!\langle X \rangle\!\rangle$ | universal field of fractions of tensor ring, 447 |
| $\mathbf{T}(U)$ | tensor ring on $U$, 44f. |
| $\mathbf{S}(U)$ | symmetric algebra on $U$, 50 |
| $\Lambda(U)$ | exterior algebra on $U$, 58 |
| $\Omega_E(R)$ | universal bimodule of derivations, 49 |
| $\mathbf{T}(M)$ | Top of a module $M$, 220 |
| $\mathbf{P}(M)$ | projective cover of $M$, 219 |
| $\mathbf{T}(A)$ | trace group, 238 |
| $\hat{M} = \mathrm{Hom}_{\mathbf{Z}}(M, \mathbf{Q}/\mathbf{Z})$ | dual module, 243 |
| $E(M)$ | injective hull, 93 |
| $A^0$ | opposite ring, II |
| $A^e$ | enveloping algebra, 248 |
| $[A:K]$ | dimension, II |
| $\sim$ | Brauer equivalent, 266 |
| $\lambda_a, \rho_a$ | left, right multiplication, II |
| $T_{A/k}, N_{A/k}$ | trace, norm, 274 |
| $(a, b; k)_n, (a, b; k\|_p$ | symbol, 296f. |

| | |
|---|---|
| $(a, F/k, \sigma)$ | cyclic algebra, 296 |
| $A^{\otimes n}$ | repeated tensor product, 294f. |
| $C(V)$ | Clifford algebra, 301 |
| $\langle a_1, \ldots, a_n \rangle$ | diagonal quadratic form, II |
| $V \perp V'$ | orthogonal sum, II |
| $\Gamma(V)$ | Clifford group, 307 |
| $\mathrm{pin}(V)$, $\mathrm{Spin}(V)$ | Spin group, 309 |
| $\hat{W}(k), W(k)$ | Witt ring, 323, 325 |
| $A_1(k)$ | Weyl algebra, 362 |
| $(X)_l, (X)_r$ | left, right annihilator, 370 |
| $S_n$ | standard polynomial, 375 |
| $F_{(n)}$ | generic matrix ring, 384 |
| $C_n$ | Capelli polynomial, 388 |
| $D_n$ | Razmyslov polynomial, 388 |
| $A^1$ | augmented algebra, II |
| $\mathbf{J}(R)$ | Jacobson radical, II |
| $U(R)$ | upper nilradical, 416 |
| $L(R)$ | lower nilradical, prime radical, 416 |
| $N(R)$ | Levitzki nilradical, 417 |
| $Z(R)$ | Singular ideal, 371 |
| $\mathbf{E}_n(K)$ | group generated by all elementary matrices, 437 |
| $\mathbf{D}_n(K)$ | group generated by all (invertible) diagonal matrices, 440 |
| $K^{ab}$ | ring or skew field abelianized, 50, 436 |

# Index


AB5-axiom, 84
Abelianize, 50, 81f., 436
Abelian category,74
Abelian valuation, 451
Absolutely flat, 243
ACC (= ascending chain condition), Noetherian, 349
Acyclic complex, 102, 105
Additive category, functor, 71, 78
Adic extension, 451
Adic topology, 364
Adjoint associativity, 89
Adjoint functor, pair, 81
Albert, A. A. (1906–72), 268, 283
Algebra, 2, 409
Algebraic closure, 169, 196
Algebraic dependence, 168ff
Algebraic extension, 170
Alternating form, 330
Amitsur, S. A. (1921–), 283, 386, 392, 406, 422
Amitsur–Levitzki theorem, 378
Andrunakievich, V. A. (1917–), 418
Anisotropic, 301, 325
Annihilator, 370
Antichain, 382
Anticommutative, 42
Antiderivation, 59
Arf invariant, 346
(Arf, C. 1910–)
Arity, 1
Artin, E., 197, 259, 310, 313, 315, 434
Artin, M. (1934–), 421
Artin's problem, 455
Artin–Schreier extension, 209
Augmentation ideal, 42, 129, 324
Augmentation map, 409
Augmented algebra, 409
Auslander, M. (1926–), 238
Automorphism, 3, 165
Automorphism class group, 127
Azumaya, G. (1920–), 217, 238, 259
Azumaya algebra, 238, 421

Baer injectivity criterion, 90
Baer product, sum, 132, 285f.
Baer nilradical, 416
Balanced bifunctor, 110
Balanced map, 89
Bar resolution, 130
Basic ring, 256
Basis, 21, 150
Bass, H. (1932–), 222, 239
Becker, E. 316
Bergman, G. M. (1943–), 27, 39, 405
Bezout domain, 354, 428
Bicentral(izer), 399
Bidimension, didim, 249
Bifunctor, 110
Binary tetrahedral group, 298
Binet–Cauchy identity, 62
(Binet, J. P. M., 1786–1856)
Biproduct, 71
Birkhoff, G. (1911–), 11, 18
Boundary, 102, 130
Bourbaki, N. (190•–), 398
Box principle, 34
Brauer, R. (1902–77), 263, 433
Brauer class, group, 267, 272
Brauer equivalent, 238
Brauer–Wall group, 328
Bruhat normal form, 438
(Bruhat, F., 1929–)
Burnside, W., 57, 147, 413

Capelli polynomial, 388
(Capelli, A., 1858–1916)
Cardinal number, 34
Carrier, 2
Cartan, H. (1904–), 340, 433

Cartan–Brauer–Hua theorem, 433
Cauchy, A. L., 62
Cayley, A., 436
cd, 97
Central algebra, 259, 421
Central extension, 128
Central polynomial, 386ff.
Central series, 136f., 157
Centre of an abelian category, 232
$C_1$-field, 277
Chain map, complex, 102
Change-of-rings, 91, 115
Characteristic, 329, 353, 432
Characteristic class, 111
Characteristic polynomial, 54
Characteristic subgroup, 135
Chase, S. U. (1932–), 245
Chevalley, C., 160, 278, 398, 452
Classical groups, 160
Clifford algebra, group, 301, 307
Clone, 14
Co-boundary, chain, cycle, 130
Cochain complex, 102
Cofinite, 29
Cogenerator, 94, 243
Coherent ring, 245
Cohn, P. M., 280, 412, 428ff., 455, 458
Cohomological dimension, cd, 97
Cohomology group, 129, 249
Coimage, cokernel, 73
Coinduced extension, 91
Comaximal, 412
Comma category, 69
Commutator, 155
Commutator ideal, 50
Compactness theorem, 31
Comparison theorem (complexes), 105
Compatible, 3
Complement, 141, 165
Completely primary ring, 66
Complex, 74
Complex-skew polynomial ring, 362
Compound matrix, 62
Cone, 310
Congruence, 6
Conjugate, 150, 155, 222, 263
Connecting homomorphism, 104
Constant, 46, 182
Constant operation, 1
Copower, coproduct, 70
Core of a field, 310f.
Core of a matrix, 438
Core of a right ideal, 402
Corestriction, cor, 292ff.
Correspondence, 4
Coset diagram, 152
Crossed homomorphism, 131
Crossed product, 283
Cycle, 102, 130
Cyclic algebra, 295
Cyclic conjugate, 150

DCC (= descending chain condition), Artinian, 258
Dedekind's Lemma, 205, 284
Defective quadratic form, 345
Degree, 40f., 169, 172, 359
Degree of central simple algebra, 268
Deligne, P. R. (1945–), 174
Denominator set, 351
Dense action, 258, 399
Dense functor, 79
Density theorem, 259, 400
Dependence of modules, 368
Dependence relation, 20
Derivation, 46ff., 131, 181ff., 360
Derived functor, 106ff.
Derived operation, 14
Determinant, 304, 441
Diagonal functor, 83
Diagram of groups, 151
Diamond lemma, 26
Dicks, W. (1947–), 116
Dickson, L. E. (1874–1954), 160, 299, 342
Dieudonné, J., 340, 342, 425, 436
Dieudonné determinant, 441
Differential module, 101f.
Dilworth's theorem, 382
Dimension, dim, 96f., 169, 395
Direct family, limit, 83
Direct power, product, 3
Discriminant of a quadratic form, 304
Divisible module, 245f.
Division algebra, 259
Dominate, 452
Draxl, P. K. (1944–83), 277, 283, 366
Dual module, 234, 243

Eckmann, B. (1917–), 93
Eilenberg, S. (1913–), 48, 121, 228
Eilenberg trick, 122
Elementary matrix, 160
Elementary reduction, expansion, 148f.

Endomorphism, 3
Ens (category of sets), 69
Enveloping algebra, 248
Epic, epimorphism, 73
Equivalence of categories, 79
Essential extension, 90, 367
Essential homomorphism, 218
Euclidean domain, 161
Euclidean field, 313, 330
Euler characteristic, 110
Exact functor, 78f.
Exact homology sequence, 104
Exact sequence, 74
Exchange property, lemma, 20ff.
Exponent, 206, 288
Extension, 75, 111, 124
Exterior algebra, 58

Factor set, 127
Factor theorem, 8
Faithful functor, 79, 256
Faithful module, 239, 401
Fibre of a mapping, 5
Fibre product, sum, 76
Field of fractions, 352
Filter, 28
Filtered algebra, 66
Final object, 69
Finitary, 1
Finite, 34
Finite intersection property, 28
Finitely presented, related, 240
Fir, 98, 425f.
Fitting, H. (1906–38), 138
Fitting's lemma, 214
Fitting subgroup, 138
Five-lemma, 109
Flanders, H. (1925–), 193
Flat module, 114, 240ff.
Forgetful functor, 81
Formally real, 310
Formanek, E. (1942–), 387
Franke, E., 66
Frattini, G., 135, 138
Frattini argument, 137
Frattini subgroup, 135
Free algebra, 14, 17, 45
Free field, 447, 451
Freyd, P. (1936–), 90, 257
Frobenius' theorem, 281
Full action, 399
Full functor, 79
Full matrix, 444
Fully invariant, 16, 381
Fundamental involution, 307

Galois connexion, 6
Galois descent, 202
Galois extension, 197
Gaussian extension, 450
General linear group, 160
Generalized polynomial identity, GPI, 389
Generating set, 3
Generator, 93, 227
Generic division algebra, 386
Generic matrix ring, 384
Global dimension, gl. dim., 96
Goldie, A. W. (1920–), 353
Goldie ring, 371, 374
Goldie's theorem, 373
Goldman, O. (1925–86), 238
Golod, E. S., 56, 58, 417
Golod–Shafarevich theorem, 56
GPI-theorem, 391
Graded algebra, module, 40f.
Graded tensor product, 302
Graph of a mapping, 4
Grassmann, H. G. (1809–77), 63
Green, J. A. (1926–), 217
Grothendieck, A. (1928–), 84, 323
Group word, 148

Hall, P., 12, 140f., 144f., 156
Hall subgroup, 141
Hall's theorem, 145
Hamel basis, 24, 39
(Hamel, G. 1877–1954)
Hamilton, W. R., 280
Harrison, D. K. (1931–), 329
Hasse, H. (1898–1980), 273, 283
Hasse invariant, 328
Hattori, A. 239
Hereditary ring, 98
Herstein, I. N. (1923–88), 270, 392, 418, 435
Higgins, P. J. (1926–), 394
Higman, G. (1917–), 394
Hilbert basis theorem, 361
Hilbert polynomial, series, 53f.
Hilbert–Serre theorem, 53
Hilbert syzygy theorem, 118
Hilbert's 'Theorem 90', 204
Hochschild group, 249

(Hochschild, G. P. 1916–)
Hölder, O. 134
Homogeneous, 41
Homological dimension, hd, 96
Homology group, 102, 129, 249
Homomorphism, 3
Homotopy, 102
Honest homomorphism, 447
Hopfian group, 154
(Hopf, H., 1894–1971)
Hotzel, E., 425
Hua, L.-K. (1909–85), 433
Hua's identity, theorem, 433f.
Hyperbolic plane, 309, 331

IBN, invariant basis number, 426
Idempotent, 221
Identity, 16
Image, im, 5, 73
Index of central simple algebra, 268
Index of quadratic form, 325
Index reduction factor, 268f.
Induced extension, 91
Induction algebra, 32ff.
Inertia lemma, 429
Inertia theorem, 429, 459
Inexact, 81
Infinitesimal, 37, 343
Inflation map, 133, 291
Initial object, 69
Injective dimension, 97
Injective hull, 93
Injective object, module, 85, 91
Inner rank, 444
Intermediate value property, 315
Invariant element, 355
Inverse family, limit, 83f.
Inverting homomorphism, 350
Involution, 59, 281
Isomorphism theorems, 8f.
Iwasawa, K. (1917–), 162

Jacobson, N., 361, 398, 413, 425
Jacobson radical, 404, 411
Jategaonkar, A. V., 405

Kaplansky, I. (1917–), 431
Kaplansky's PI-theorem, 407
Kernel, ker, 5, 73
Kharchenko, V. K., 298
Klein quadric, theorem, 64
(Klein, F., 1849–1925)
Knebusch, M., 316
Köthe, G. (1905–89), 270
Köthe's conjecture, 417
Krull–Schmidt theorem, 216f., 226
Krull topology, 200
Kummer extension, 207
Kurosh, A. G. (1908–71), 57

Lambek, J. (1922–), 101
Laplace expansion, 62
Large submodule, 367
Latyshev, V. N., 383
Law, 16
Leibniz, G. W., 436
Leibniz's formula, 51
Leicht, J. B., 329
Length, 12
Lenstra, Jr., H. W., 171
Levitzki, J. (1904–56), 378
Levitzki nilradical, 417
Lifting, 221
Limit, 83
Linear category, functor, 87
Linear dependence, 20f.
Linear group, 160
Linearly disjoint, 186
Line coordinates, complex, 64
Local ring, 213
Localization, 350
Loop functor, 96
Lorenz, F., 329
Lukasiewicz, J. (1878–1956), 13
Lüroth's theorem, 173
(Lüroth, J., 1844–1910)

Mac Lane's criterion, 192
(Mac Lane, S. 1909–)
Magnus' theorem, 159, 166
(Magnus, W., 1912–)
Malcev, A. I., 257
Mapping cone, cylinder, 122
Martindale III, W. S., 392
Matroid, 39
Merkuryev, A. S., 297, 347
Metacyclic group, 135
Mitchell, B. (1934–), 257
Modular law, 124
Modular right ideal, 410
Monic, monomorphism, 72
Monomial matrix, 146, 440
Morita context, 235
Morita equivalence, 227

Morita invariant, 232
(Morita, K.)
m-system, 414
Multilinear, 376
Multiplicative set, 350
Multiplicator, 133
Multivector, 58

Nagata, M., 64f.
Nagata–Higman theorem, 394
Nakayama, T., 259
Natural number, 32
Negative, 310
Negative-definite, 316
Neumann, J. von, 243
Newman, M. H. A. (1897–1984), 26f.
Nielsen transformation, 166
(Nielsen, J., 1890–1959)
Nil ideal, 222, 414
Nilpotent group, 137, 157
Nilpotent ideal, 414
Nilradical, 415
Noether, E., 170, 176, 350
Noether's equations, 203
Noether's problem, 170
Non-standard model, 37
Norm, 307
Normal basis theorem, 205
Normal extension, 180
Normalizer, 136
Norm residue symbol, 298f.
Noughtary, 1

Obstruction, 111
Operation, 1
Operator domain, 1
Order, 363
Ordered ring, 316
Ore, O. (1899–1968), 350ff., 361, 393
Ore domain, 352
Ore set, 351
Orthogonal group, 306ff.

Peano, G. (1858–1932), 32
Peano axioms, 32
Perfect closure, 178
Perfect group, 335
Perfect ring, 226
Permutation matrix, 440
Petrich, M. (1932–), 425
PI-algebra, 375
PI-degree, 421
PID, principal ideal domain, 355
Pin ($V$), 307
Place, 454
Platonov, V. P., 277
Plücker coordinates, 64
(Plücker, J. V., 1801–68)
Polynomial identity, PI, 375, 389
Positive (cone), 310
Positive-definite, 316
Posner's theorem, 420
(Posner, E. C. 1933–)
Prefix notation, 13
Presentation matrix, 239
Primary algebra, 289
Prime ideal, 414
Prime radical, 416
Prime ring, 354, 367, 413
Primitive group, 162
Primitive ring, 402
Principal ideal domain, 355
Principle of induction, 32
Procesi, C. (1941–), 421
Product in categories, 70
Profinite group, 84, 200
Progenerator, 237
Projection operator, 14, 70
Projective cover, 218
Projective dimension, 96
Projective linear group, space, 162, 332, 343
Projective object, module, 85, 91
Projective resolution, 94, 105
Pseudolinear extension, 456
Pullback of a map, 91
Pullback of a pair, 76
Pure extension, sequence, 257
Purely inseparable extension, 176
Pushout, 76
Pythagorean field, 330

Quadratic form, space, 300
Quasi-algebraically closed, 277
Quasi-commutative valuation, 453
Quasi-Galois extension, 180
Quasi-inverse, regular, 405
Quaternion algebra, group, 280
Quillen, D. (1940–), 119
Quotient algebra, 7
Quotient category, 229
Quotient object, 73
Quotient ring, 352

$p$-radical extension, 176
Rank (free algebra), 52
Rank (free group), 150
Rank (module), 369, 395
Rank (endomorphism), 396
Razmyslov polynomial, 388
Razmyslov, Yu. P., (1951–), 387f.
Real closed, 312
Recursion, 36f.
Reduced norm, trace, 274
Reduced product, 29
Reduced ring, 189, 415, 418
Rees, D. (1918–), 425
Rees matrix algebra, 423
Regev's theorem, 383
(Regev, A.)
Regular element, 349, 374
Regular extension, 193
Regular ring, 243
Relation, 20
Remak, R. (1888–1942?), 217
Representable, 80
Residually-$P$, 10
Residue class field, 213
Resolution, 94, 105
Restriction, res, 133, 290
Retraction, 75, 229
Ring, 45
Roganov's theorem, 116
(Roganov, Yu. V.)
Rosset, S., 379
Row, 11
Rowen, L. H., 420
Row-finite, 239, 396
Ryabukhin, Yu. M., 418

Sandwich matrix, 423
Sąsiada, E., 412
Schanuel's lemma, 96
(Schanuel, S., 1933–)
Scharlau, W., 313
Schelter, W., 422
Schilling, O. F. G. (1911–73), 449
Schmidt, F. K. (1901–77), 193
Schmidt, O. Yu. (1891–1956), 217
Schofield, A. H. (1957–), 261, 455
Schopf, E., 93
Schreier, O., 209, 310
Schreier's extension theorem, 127
Schreier's formula, 153
Schreier's subgroup theorem, 153
Schröter, K., 12
Schur–Zassenhaus theorem, 142
Section, 75, 229
Self-injective ring, 374
Semi-Artinian module, 393
Semidirect product, 126
Semifir, 428
Semihereditary, 122
Semilinear, 202
Semiperfect ring, 222
Semiprime ideal, 415
Semiprime ring, 367, 414
Semiprimitive ring, 404, 411
Separability idempotent, 249
Separable algebra, 189, 249, 421
Separable closure, 178
Separable degree, 176
Separating congruences, 10
Separating transcendence basis, 192
Serre, J.-P. (1926–), 53, 298
Shafarevich, I. R. (1923–), 56
Siegel, C. L. (1896–1981), 338
Sierpiński, W., 15
Signature, 2, 316
Similar algebras, 266
Similar elements, 358
Simple algebra, 7, 423
Simple module, 410
Singular submodule, 370
Skew field, 259
Skew Laurent polynomial, 363
Skew polynomial ring, 360
Skolem–Noether theorem, 263
(Skolem, T., 1887–1963)
Snake lemma, 101
Socle, 398, 403
Span, -ning set, 21, 23
Specialization lemma, 445
Speiser, A. (1885–1970), 203
Spin group, 309
Spinor kernel, norm, 308
Spitzlay, K. J., 316
Split exact, 75
Split extension, 125, 250
Split quadratic space, 324
Splitting field, 268
Sporadic simple group, 160
Stabilizer, 162
Stable linear group, 438
Stably free module, 119
Staircase lemma, 379
Stallings, J. R. (1935–), 239
Standard identity, polynomial, 375

Standard involution, 59
Standard resolution, 130
Staudt, Chr. von (1798–1867), 435
Steinitz, E., 197
Sturm, J. C. F. (1803–55), 318
Sturm's theorem, 321
Subdirect product, 10
Subdirectly reducible, 10
Subobject, 72
Successor function, 32
Superalgebra, 43
Supernatural number, 200
Suslin, A. A., 119, 297
Swan, R. G. (1933–), 171
Sylvester, J. J., 316
Sylvester–Franke theorem, 66
Symbol, 296, 347
Symmetric algebra, 50
Symmetry, 306
Symplectic group, space, 330f.
Syzygy theorem, 118

Takahasi, M., 166
Tamagawa, T., 337
Tannaka–Artin problem, 277, 366
(Tannaka, T.)
Tensor algebra, 44f.
Tensor product (graded), 302
Tensor ring, 428
$T$-ideal, 381
Top of a module, 220
Topology of pointwise convergence, 399
Torsion, Torsion free, 358
Torsion product, Tor, 113
Total divisor, 355
Total quotient ring, 352
Totally positive, 312
Trace, 316
Trace group, function, 238
Trace ideal, 234
Transcendence basis, degree, 169
Transduction, 67
Transfer, 147, 292
Transgression, 133
Transitive action, 162
Translation ring, 363
Transvection, 160, 333
Treur, J. (1952–), 433
Trivial algebra, 2
Trivializable relation, 426
Tsen's theorem (Tsen, C. C,), 279

Ultrafilter, 28
Ultrapower, product, 29
Uniform module, rank, 368f.
Unipotent, 338
1-unit, 449
Unital module, 409
Unitriangular matrix, 438
Universal derivation bimodule, 48
Universal field of fractions, 444, 447

Valency, 12
Valuation ring, 448
Value group, 448
Variety, 16
Versor, 307
Villamayor, O., 253
Vinberg, E. B., 56
Voskresenski, V. E., 171

Wall, C. T. C. (1936–), 328
Warning, E., 278
Watts, C. E., 121, 228
Weakly finite, 444
Weak (global) dimension, 114
Webber, D. B., 234
Wedderburn, J. H. M., 217, 265f., 297
Wedderburn decomposition theorem, 412
Wedderburn first structure theorem, 401
Wedderburn theorem (finite fields), 265
Wedderburn's principal theorem, 252
Weyl algebra, 362
(Weyl, H., 1885–1955)
Whaples, G., 259
Whitehead group, 276
Whitehead lemma, 437
(Whitehead, J. H. C., 1904–60)
Wielandt, H., 138, 144, 159
Windmill lemma, 78
Witt chain equivalence theorem, 324
Witt decomposition, index, 325
Witt identity, 156
Witt invariant, 328
Witt (Grothendieck) ring, 323f.
Word, word algebra, 12

Yoneda's lemma, 120
(Yoneda, N.)

Zassenhaus, H., 142
Zelinsky, D. (1923–) 224, 253
Zero morphism, object, 72
Zieschang, H., 166