Graduate Algebra: Noncommutative View

Louis Halle Rowen

Graduate Studies in Mathematics

Volume 91

American Mathematical Society

Graduate Algebra:
Noncommutative View

Graduate Algebra: Noncommutative View

Louis Halle Rowen

Graduate Studies
in Mathematics
Volume 91

American Mathematical Society
Providence, Rhode Island

2000 *Mathematics Subject Classification.* Primary 16–01, 17–01;
Secondary 17Bxx, 20Cxx, 20Fxx.

For additional information and updates on this book, visit
www.ams.org/bookpages/gsm-91

Library of Congress Cataloging-in-Publication Data

Rowen, Louis Halle.
Graduate algebra : commutative view / Louis Halle Rowen.
p. cm. — (Graduate studies in mathematics, ISSN 1065-7339 : v. 73)
Includes bibliographical references and index.
ISBN 978-0-8218-0570-1 (alk. paper)
1. Commutative algebra. 2. Geometry, Algebraic. 3. Geometry, Affine. 4. Commutative rings. 5. Modules (Algebra). I. Title. II. Series.

QA251.3.R677 2006
512′.44—dc22 2006040790

---

10 9 8 7 6 5 4 3 2 1 13 12 11 10 09 08

To the memory of my beloved mother
Ruth Halle Rowen, April 5, 1918 – January 5, 2007

Contents

Introduction

As indicated in the title, this volume is concerned primarily with noncommutative algebraic structures, having grown from a course introducing complex representations of finite groups via the structure of group algebras and their modules. Our emphasis is on algebras, although we also treat some major classes of finite and infinite groups. Since this volume was conceived as a continuation of Volume 1 (*Graduate Algebra: Commutative View*, Graduate Studies in Mathematics, volume 73), the numeration of chapters starts with Chapter 13, Part IV, and we use the basics of rings and modules developed in Part I of Volume 1 (Chapters 1–3). Nevertheless, Chapters 13–15 and 18 can largely be read independently of Volume 1.

In the last one hundred years there has been a vast literature in noncommutative theory, and our goal here has been to find as much of a common framework as possible. Much of the theory can be cast in terms of representations into matrix algebras, which is our major theme, dominating our treatment of algebras, groups, Lie algebras, and Hopf algebras. A secondary theme is the description of algebraic structures in terms of generators and relations, pursued in the appendices of Chapter 17, and leading to a discussion of free structures, growth, word problems, and Zelmanov's solution of the Restricted Burnside Problem.

One main divergence of noncommutative theory from commutative theory is that left ideals need not be ideals. Thus, the important notion of "principal ideal" from commutative theory becomes cumbersome; whereas the principal left ideal Ra is described concisely, the smallest ideal of a noncommutative ring $\mathrm{Q}R$ containing an element a includes all elements of the form

$$r_{1,1}ar_{1,2} + \cdots + r_{m,1}ar_{m,2}, \quad \forall r_{i,1}, r_{i,2}, \in R,$$

where m can be arbitrarily large. This forces us to be careful in distinguishing "left" (or "right") properties from two-sided properties, and leads us to rely heavily on modules.

There are many approaches to structure theory. We have tried to keep our proofs as basic as possible, while at the same time attempting to appeal to a wider audience. Thus, projective modules (Chapter 25) are introduced relatively late in this volume.

The exposition is largely self-contained. Part IV requires basic module theory, especially composition series (Chapter 3 of Volume 1). Chapter 16 draws on material about localization and Noetherian rings from Chapters 8 and 9 of Volume 1. Chapter 17, which goes off in a different direction, requires some material (mostly group theory) given in the prerequisites of this volume. Appendix 17B generalizes the theory of Gröbner bases from Appendix 7B of Volume 1. Chapter 18 has applications to field theory (Chapter 4 of Volume 1).

Parts V and VI occasionally refer to results from Chapters 4, 8, and 10 of Volume 1. At times, we utilize quadratic forms (Appendix 0A) and, occasionally, derivations (Appendix 6B). The end of Chapter 24 draws on material on local fields from Chapter 12. Chapters 25 and 26 require basic concepts from category theory, treated in Appendix 1A.

There is considerable overlap between parts of this volume and my earlier book, *Ring Theory* (student edition), but the philosophy and organization is usually quite different. In *Ring Theory* the emphasis is on the general structure theory of rings, via Jacobson's Density Theorem, in order to lay the foundations for applications to various kinds of rings.

The course on which this book is based was more goal-oriented — to develop enough of the theory of rings for basic representation theory, i.e., to prove and utilize the Wedderburn-Artin Theorem and Maschke's Theorem. Accordingly, the emphasis here is on semisimple and Artinian rings, with a short, direct proof. Similarly, the treatment of Noetherian rings here is limited mainly to Goldie's Theorem, which provides most of the nontechnical applications needed later on.

Likewise, whereas in *Ring Theory* we approached representation theory of groups and Lie algebras via ring-theoretic properties of group algebras and enveloping algebras, we focus in Part V of this volume on the actual groups and Lie algebras.

Thanks to Dror Pak for pointing me to the proofs of the hook categories, to Luda Markus-Epstein for material on Stallings foldings, to Alexei Belov for gluing components in the Wedderburn decomposition, and to Sue Montgomery for a description of the current state of the classification of

finite dimensional Hopf algebras. Steve Shnider, Tal Perri, Shai Sarussi, and Luie Polev provided many helpful comments. Again, as with Volume 1, I would like to express special gratitude to David Saltman, in particular for his valuable suggestions concerning Chapter 24 and Chapter 25, and also to Uzi Vishne. Thanks to Sergei Gelfand for having been patient for another two years. And, of course, many thanks to Miriam Beller for much of the technical preparation of the manuscript.

Needless to say, I am deeply indebted to Rachel Rowen, my helpmate, for her steadfast support all of these years.

List of Symbols

Warning: Some notations have multiple meanings.

Prerequisites

As mentioned in the Introduction, most of Part IV of Volume 2 is self-contained, modulo some basic results on rings and modules. In Chapter 17, we need a few extra general basic results, mostly concerning finitely generated groups, which we list here.

Finitely generated (f.g.) groups.

A fair part of Chapter 17 concerns f.g. groups, introduced briefly in Volume 1, namely on p. 13 and Exercises 0.23–0.27. Often we look for f.g. subgroups of a given f.g. group. The following straightforward facts often come in handy. Recall that a subgroup H has **finite index** G if H has finitely many cosets in G, the number of which is designated as $[G{:}H]$.

Remark 00.1. Any subgroup H of finite index in a f.g. group G is also f.g. (This was stated in Exercise 0.27 of Volume 1, with an extensive hint.) The same proof shows, more precisely, that if G is generated by t elements and $[G{:}H] = m$, then H is generated by tm elements.

LEMMA 00.2. *For any $n \in \mathbb{N}$, any f.g. group G has finitely many subgroups of index n.*

Proof. We elaborate on Exercise 0.25 of Volume 1. For any subgroup H of index n, we have a homomorphism $\psi_H{:}\, G \to S_n$, given by left multiplication on the cosets of H. But any element a of $\ker \psi_H$ satisfies $aH = H$, implying $\ker \psi_H \subseteq H$, and thus $H = \psi_H^{-1}(\overline{H})$ for some subgroup $\overline{H}$ of S_n.

Working backwards, since G is f.g., there are only finitely many homomorphisms from G to S_n, which has finitely many possible subgroups $\overline{H}$. Since any subgroup H of index n can be recovered in this way, we have only finitely many possibilities for H. □

PROPOSITION 00.3. *If H is a f.g. normal subgroup of G, and K is a subgroup of finite index in H, then K contains a f.g. normal subgroup of G that has finite index in H. (The special case for $H = G$ was given in Exercise 0.25 of Volume 1.)*

Proof. For each $g \in G$, gKg^{-1} is a subgroup of $gHg^{-1} = H$ of the same index as K; by the lemma, there are only finitely many of these, so, by Exercise 0.24 of Volume 1, $\bigcap_{g\in G} gKg^{-1}$ is a normal subgroup of G having finite index in H. □

Groups of fractions.

In the proof of Theorem 17.61 we also need the following easy special case of the construction of Exercise 8.26 of Volume 1:

Definition 00.4. Suppose $(A, +)$ is a torsion-free Abelian group. The group $A_{\mathbb{Q}}$ is defined as follows:

Define an equivalence on $A \times \mathbb{N}^+$ by putting $(a, m) \sim (b, n)$, iff $an = bm$. Writing $\frac{a}{m}$ for the equivalence class $[(a, m)]$, we define $A_{\mathbb{Q}}$ to be the set of equivalence classes, endowed with the operation

$$\frac{a}{m} + \frac{b}{n} = \frac{an + bm}{mn}.$$

Remark 00.5. $A_{\mathbb{Q}}$ is a group, and in fact is a $\mathbb{Q}$-module in the natural way, namely

$$\frac{u}{v}\frac{a}{m} = \frac{ua}{vm}, \qquad a \in A,\ u \in \mathbb{Z},\ m, v \in \mathbb{N}^+.$$

There is a group injection $A \to A_{\mathbb{Q}}$ given by $A \mapsto \frac{a}{1}$. Furthermore, any automorphism σ of A extends naturally to an automorphism of $A_{\mathbb{Q}}$ via the action $\sigma(\frac{a}{m}) = \frac{\sigma(a)}{m}$.

(The verifications are along the lines of those in the proof of Proposition 12.18 of Volume 1. Alternatively, once we have tensor products from Chapter 18, we could view $A_{\mathbb{Q}}$ as $A \otimes_{\mathbb{Z}} \mathbb{Q}$.)

Jordan decomposition.

The Jordan decomposition of Theorem 2.75 of Volume 1 has an easy but useful application in nonzero characteristic:

PROPOSITION 00.6. *Over a field of characteristic $p > 0$, any $n \times n$ matrix T has a power whose radical component is 0.*

Proof. Write the Jordan decomposition $T = T_{\mathbf{s}} + T_{\mathbf{n}}$, where the semisimple component $T_{\mathbf{s}}$ and the nilpotent component $T_{\mathbf{n}}$ commute. Then, as in Corollary 4.69 of Volume 1,

$$T^{p^k} = (T_{\mathbf{s}} + T_{\mathbf{n}})^{p^k} = T_{\mathbf{s}}^{p^k} + T_{\mathbf{n}}^{p^k}$$

for each k, but $T_{\mathbf{n}}^{p^k} = 0$ whenever $p^k > n$, so we conclude for such k that $T^{p^k} = T_{\mathbf{s}}^{p^k}$ is semisimple. □

Galois theory.

We also need a fact from Galois theory, which was missed in Volume 1.

PROPOSITION 00.7. *Suppose F is a finite field extension of $\mathbb{Q}$, and $a \in F$ is integral over $\mathbb{Z}$. If $|\sigma(a)| \leq 1$ for every embedding $\sigma\colon F \to \mathbb{C}$, then a is a root of unity.*

Proof. The minimal monic polynomial $f_a \in \mathbb{Z}[\lambda]$ of a over $\mathbb{Z}$ has some degree n; its coefficients are sums of products of conjugates of a, and so by hypothesis have absolute value $\leq n$. But there are at most $(2n+1)^n$ possibilities for such a polynomial; moreover, the hypothesis also holds for each power of a, which must thus be a root of one of these polynomials. We conclude that there are only finitely many distinct powers of a, which means a is a root of unity. □

The trace bilinear form.

We need a result about the **trace bilinear form** on the matrix algebra $M_n(F)$ over a field F, given by $\langle x, y\rangle = \operatorname{tr}(xy)$. Clearly this form is symmetric and also nondegenerate, for if $x = (a_{ij})$ with $a_{i_0 j_0} \neq 0$, then $\operatorname{tr}(x e_{j_0 i_0}) = a_{i_0 j_0} \neq 0$. The **discriminant** of a base $\mathcal{B} = \{b_1, \ldots, b_{n^2}\}$ of $M_n(F)$ is defined as the determinant of the $n^2 \times n^2$ matrix $(\operatorname{tr}(b_i b_j))$. In view of Remark 4B.5 of Volume 1, the discriminant of any base $\mathcal{B}$ is nonzero (since there exists an orthogonal base with respect to the trace bilinear form).

LEMMA 00.8. *Suppose $\{b_1, \ldots, b_n\}$ is a base of $M_n(F)$ over F. Then for any $\alpha_1, \ldots, \alpha_n^2 \in F$, the system of n^2 equations $\{\operatorname{tr}(b_i x) = \alpha_i : 1 \leq i \leq n^2\}$ has at most one solution for $x \in M_n(F)$.*

Proof. Write $x = \sum_{j=1}^{n^2} \gamma_j b_j$. Then $\alpha_i = \sum_{j=1}^{n^2} \gamma_j \operatorname{tr}(b_i b_j)$, $1 \leq i \leq n^2$, can be viewed as n^2 equations in the γ_j; since the discriminant $\det(\operatorname{tr}(b_i b_j))$ is nonzero, one can solve these equations using Cramer's rule.

To prove uniqueness, suppose there were two matrices x_1 and x_2 such that $\operatorname{tr}(b_i x_1) = \operatorname{tr}(b_i x_2)$, $1 \leq i \leq n^2$. Then $\operatorname{tr}(b_i(x_1 - x_2)) = 0$ for each i, which implies $x_1 - x_2 = 0$ since the trace form is nondegenerate; thus, $x_1 = x_2$. □

Part IV

The Structure of Rings

Introduction to the Structure of Rings

Whereas much of commutative theory stems from the polynomial algebra $F[\lambda_1, \ldots, \lambda_n]$ over a field F, the matrix algebra $M_n(F)$ is arguably the most important example of a noncommutative algebra, since the tools of matrix theory are available, including the trace, determinant, transpose, and so forth. Much of representation theory involves comparing a given algebraic structure to such a matrix algebra, via suitable homomorphisms.

Our goal in this part is to introduce and develop enough structure theory of algebras and modules to enable us to carry out the applications in the next two parts, especially to representation theory. Since our main objective is representations of finite degree, we focus on finite-dimensional (f.d.) algebras over a field F.

We lay out the territory in Chapter 13, studying matrices (and related notions) from a structural point of view; any f.d. algebra can be embedded into a matrix algebra via the **regular representation**. We also study matrices as rings of endomorphisms. In Chapter 14, we introduce semisimple rings, leading to the Wedderburn-Artin theorem describing semisimple rings as direct products of matrix rings over division rings. Since any semisimple ring is Artinian, we are led in Chapter 15 to the general theory of Artinian rings, which is described elegantly by means of Jacobson's structure theory.

The material described above provides the structural basis for most of our applications. Nevertheless, there are many important algebras which

are not Artinian. Some of the general structure-theoretical results are given in Appendix 15A, and the Noetherian theory is developed in Chapter 16.

One general method of studying algebraic structures, prevalent in much of group theory, is through generators and relations; in Chapter 17 we take an excursion through this avenue, leading to presentations of free algebras and groups, and various questions as to the efficacy of such presentations. Finally, in Chapter 18 we bring in another basic tool, the tensor product, which has a myriad of applications.

Chapter 13

Fundamental Concepts in Ring Theory

In Chapter 2 of Volume 1, we encountered the ring $M_n(R)$ of $n \times n$ matrices over a commutative ring R. This ring plays a key role in ring theory, parallel to that of the symmetric group S_n in group theory. Familiar computations from linear algebra courses show that $M_n(R)$ is a ring, even when R is not commutative, although we shall also see this from a structural point of view. Much of this chapter involves the basic ring-theoretical properties of $M_n(R)$.

Since the study of matrix rings involves modules, which play a key role throughout the noncommutative theory, we start by considering the relevant properties of modules, in particular simple modules; cf. Definition 3.1 of Volume 1. Idempotents also are a key concept in the study of matrices.

Recall that by **ideal** we mean two-sided ideal. As we shall see, the ideals of R and $M_n(R)$ are in 1:1 correspondence. Thus, when D is a division ring, the ring $M_n(D)$ has no proper nonzero ideals, and also is both left and right Artinian and Noetherian.

The main thrust of the next few chapters is to prove the ultra-important Wedderburn-Artin-Hopkins-Levitzki Theorem, that any simple left Artinian ring has the form $M_n(D)$ for a suitable division ring D. Our strategy in obtaining this theorem is to find an alternate description of matrix rings, motivated by the familiar observation from linear algebra that $M_n(F)$ can be identified with the linear transformations of $F^{(n)}$. The ring of maps $\mathrm{End}_R M$ from an R-module M to itself is an important generalization, and one has the ring-theoretic version of Cayley's Theorem, that every ring R is a subring of the ring $\mathrm{End}_R R$. This provides a way of displaying a ring explicitly; using Schur's Lemma, we prove (Theorem 13.48) that any ring which is a direct

sum of minimal left ideals must be a finite direct product of matrices over division rings.

Picking up on the idea of displaying a ring as a subring of $\mathrm{End}\, M$ leads us to one of the major concepts of algebra, namely **representations**. We content ourselves in this chapter with the regular representation. Finally, we turn to polynomial rings and related constructions; in Appendix 13A, we digress and provide two famous examples — skew polynomial rings and Weyl algebras.

R generally denotes a (not necessarily commutative) ring. As in Volume 1, **module** will mean "left module" (although one could just as well build the theory with right modules). A left ideal L of a ring R is just an R-submodule of R; we denote this by $L \leq R$.

Simple modules.

Recall that an R-module M is **simple** if it has no submodules other than 0 and M. We also recall an observation, to be used dozens of times.

Remark 13.0 (Remark 1.9 of Volume 1). Suppose M is any R-module and $a \in M$. There is a module homomorphism $f_a \colon R \to M$ given by right multiplication by a, i.e., $r \mapsto ra$, and $\ker f_a = \mathrm{Ann}_R\, a$. In particular, if the module M is simple, then $M = Ra = f_a(R)$, and $\mathrm{Ann}_R\, a$ is a maximal left ideal of R for each $a \in M$.

This was used in the proof of:

Proposition 13.1 (Proposition 3.5 of Volume 1). *An R-module $M \neq 0$ is simple iff $M \cong R/L$, where L is a maximal left ideal of R. In particular, all rings have simple modules, since, by Zorn's Lemma, any left ideal is contained in a suitable maximal left ideal.*

We also need elementary criteria for a module to be simple.

Proposition 13.2. *The following conditions on a module M are equivalent:*

(i) *M is simple.*
(ii) *$Ra = M$, for each $0 \neq a \in M$.*
(iii) *$b \in Ra$, for every $0 \neq a, b \in M$.*
(iv) *For every $0 \neq a, b \in M$, there is suitable r in R such that $ra = b$.*

Proof. (i) $\Rightarrow$ (ii) By Remark 13.0.

(ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) Obvious.

(iv) $\Rightarrow$ (i) Suppose $0 \neq L \leq M$. Take $a \neq 0$ in L. Then for any b in M we can find $r \in R$ with $b = ra \in L$, proving that $L = M$. □

Matrix rings

We are ready to begin to analyze matrix rings, drawing on module theory when necessary. First of all, let $M_{m,n}(R)$ denote the set of $m \times n$ matrices over a ring R. Recalling the matrix units e_{ij} preceding Remark 2.27 of Volume 1, we see that $M_{m,n}(R)$ is a free R-module, the module operation being

$$r \sum_{i,j} r_{ij} e_{ij} = \sum_{i,j} r r_{ij} e_{ij} \qquad (r \in R).$$

Furthermore, any matrix (r_{ij}) can be written uniquely in the form $\sum_{i,j} r_{ij} e_{ij}$, so the module $M_{m,n}(R)$ is free, having a base comprised of the set of matrix units $\{e_{ij} : 1 \leq i \leq m,\ 1 \leq j \leq n\}$.

Usually $m = n$, and matrix units are tacitly assumed to be $n \times n$ matrices. Thus $M_{n,n}(R) = M_n(R)$ is a free R-module of rank n^2.

$M_n(R)$ also has the familiar ring structure, with multiplication of any two matrices given by

$$\Big(\sum_{i,j=1}^{n} r_{ij} e_{ij}\Big)\Big(\sum_{i,j=1}^{n} s_{ij} e_{ij}\Big) = \sum_{i,j=1}^{n} \Big(\sum_{k=1}^{n} r_{ik} s_{kj}\Big) e_{ij}.$$

It is easy to check associativity via the matrix unit condition, showing that $M_n(R)$ is indeed a ring; cf. Exercise 1. The unit element of $M_n(R)$ being $\sum_{i=1}^{n} e_{ii}$, there is a ring injection $R \to M_n(R)$ via $r \mapsto \sum_{i=1}^{n} r e_{ii}$.

$M_{m,n}(R)$ does not have a natural ring structure for $m \neq n$, but can be viewed as a left module over $M_m(R)$ and as a right module over $M_n(R)$, via the usual matrix operations. (In fact, $M_{m,n}(R)$ is then a bimodule; cf. Definition 13.35′ below.)

Although expeditious for calculations, the set of matrix units of $M_n(F)$ depends on the choice of base of $F^{(n)}$ (viewing matrices as linear transformations); a change of base would provide a different set of matrix units. This motivates us to formalize the notion of matrix unit.

Definition 13.3. A set of $n \times n$ **matrix units** of an arbitrary ring W is a set of elements $\{e_{ij} : 1 \leq i, j \leq n\}$ satisfying the following two basic properties:

(MU1) $e_{ik} e_{\ell j} = \delta_{k\ell} e_{ij}$, where $\delta_{k\ell}$ denotes the Kronecker delta (1 if $k = \ell$, and 0 otherwise).

(MU2) $\sum_{i=1}^{n} e_{ii} = 1$.

Before verifying that (MU1) and (MU2) provide everything needed to write W as a matrix ring $M_n(R)$ with respect to this given set of matrix units, we introduce a related notion.

Idempotents.

Definition 13.4. An **idempotent** of a ring R is an element e such that $e^2 = e$. The idempotents 0,1 are called the **trivial idempotents** of R. Idempotents $e_i,\ i \in I$, are called **orthogonal** if $e_i e_j = 0 = e_j e_i$ for all $i \neq j \in I$. We call $\{e_1, \ldots, e_n\}$ a **1-sum set** of orthogonal idempotents if they are orthogonal and $\sum_{i=1}^{n} e_i = 1$.

For any idempotent e, clearly $1-e$ is an idempotent orthogonal to e, so the pair $\{e, 1-e\}$ comprises a 1-sum set of orthogonal idempotents.

PROPOSITION 13.5.

(i) *If e and f are orthogonal idempotents of the ring R, then, as modules,*

$$R(e+f) = Re \oplus Rf.$$

(ii) *If $e_1, \ldots, e_n$ are orthogonal idempotents of R, then*

$$R(e_1 + \cdots + e_n) \cong Re_1 \oplus \cdots \oplus Re_n$$

as R-modules.

(iii) *If in* (ii), $\sum_{i=1}^{n} e_i = 1$, *then* $R \cong Re_1 \oplus \cdots \oplus Re_n$.

Proof. (i) $Re = Re(e+f) \subseteq R(e+f)$, and likewise $Rf \subseteq R(e+f)$. Hence

$$Re + Rf \subseteq R(e+f),$$

and the opposite direction is clear since $r(e+f) = re+rf$. Finally, we show that $Re \cap Rf = 0$: if $r_1 e = r_2 f$, then $r_1 e = r_1 e^2 = r_2 f e = 0$.

(ii) Induction applied to (i).

(iii) $Re_1 \oplus \cdots \oplus Re_n \cong R(e_1 + \cdots + e_n) = R$. □

Example 13.5′. The diagonal matrix units $\{e_{11}, \ldots, e_{nn}\}$ comprise a 1-sum set of orthogonal idempotents of $M_n(R)$, so in particular

$$M_n(R) = \bigoplus_{i=1}^{n} M_n(R) e_{ii}.$$

Idempotents also help us study the ring-theoretic structure of R.

Remark 13.6. For any ring R with an idempotent e, we define the **Peirce decomposition**

$$R \;=\; eRe \;\oplus\; eR(1-e) \;\oplus\; (1-e)Re \;\oplus\; (1-e)R(1-e)$$

as Abelian groups.

The Peirce components eRe and $(1-e)R(1-e)$ are rings, with respective multiplicative units e and $(1-e)$. A richer way of viewing the Peirce decomposition is to write $e_1 = e$, $e_2 = 1-e$, and $R_{ij} = e_iRe_j$ for $j = 1, 2$. Then $R_{ij}R_{jk} = R_{ik}$ for all i, j, k, so R can be written in matrix notation as $\left(\begin{smallmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{smallmatrix}\right)$, with multiplication as in matrix multiplication.

This way of writing a ring has led to many interesting examples, especially when $R_{21} = 0$; cf. Exercise 16. See Exercise 9 for the general version of the Peirce decomposition.

Another famous source of idempotents is the idempotent linear transformation, such as the projection of a vector space to a subspace. On the other hand, let us see now that certain rings have no nontrivial idempotents at all.

Definition 13.7. A ring R is a **domain** if $ab \neq 0$ for every nonzero a, b in R.

In a domain R, since $e(1-e) = 0$ for any idempotent e, we have $e = 0$ or $1-e = 0$, proving that R has no nontrivial idempotents.

Matrix rings described in terms of matrix units.

We have already defined the natural injection $R \mapsto M_n(R)$ of rings given by $r \mapsto \sum_{i=1}^n re_{ii}$. Nevertheless, there is another way of extracting from $M_n(R)$ a ring isomorphic to R, which is at times more convenient since the notation is more concise.

Remark 13.8. (i) If e is an idempotent of a ring W, then eWe is a ring with multiplicative unit e. (Easy verification.) Also, $wEw = \{w \in W : ewe = w\}$, since $e(ewe)e = e^2we^2 = ewe$. (However, eWe is not a subring of W when $e \neq 1$, since the multiplicative units differ.)

(ii) R is isomorphic to the ring $e_{11}M_n(R)e_{11}$, via the homomorphism $r \mapsto e_{11}re_{11}$.

Although rather trivial, Remark 13.8 serves as our main tool in relating the structures of R and $M_n(R)$, and gains in profundity in the Morita theory (Appendix 25A).

PROPOSITION 13.9. *Suppose W is any ring having a set of matrix units $\{e_{ij} : 1 \leq i, j \leq n\}$ satisfying* (MU1) *and* (MU2) *of Definition 13.3. Then $W \cong M_n(R)$, where $R = e_{11}We_{11}$.*

Proof. Define $\varphi\colon W \to M_n(R)$ by $\varphi(w) = (r_{ij})$, where $r_{ij} = e_{1i}we_{j1} \in R$. This is clearly a homomorphism of the additive group structure, and noting that $\sum_{k=1}^n e_{k1}e_{1k} = \sum e_{kk} = 1$, we see that the i, j entry of $\varphi(ww')$ is

$$e_{1i}ww'e_{j1} = e_{1i}w\left(\sum_{k=1}^n e_{k1}e_{1k}\right)w'e_{j1} = \sum_{k=1}^n r_{ik}r'_{kj}.$$

Hence $\varphi(ww') = \varphi(w)\varphi(w')$, and φ is a homomorphism. To show that φ is onto, note that the matrix $(r_{ij}) = \varphi\left(\sum e_{i1}r_{ij}e_{1j}\right)$, since

$$e_{1i}(e_{i1}r_{ij}e_{1j})e_{j1} = e_{11}r_{ij}e_{11} = r_{ij}.$$

Finally, $w \in \ker\varphi$ iff each entry $r_{ij} = e_{1i}we_{j1}$ is 0, implying

$$e_{ii}we_{jj} = e_{i1}(e_{1i}we_{j1})e_{1j} = 0,$$

and thus

$$0 = \sum_{i,j=1}^n e_{ii}we_{jj} = \sum_{i=1}^n e_{ii}w\sum_{j=1}^n e_{jj} = 1w1 = w,$$

proving that $\ker\varphi = 0$. □

Thus the structure of matrix rings is encoded in the matrix units, as illustrated in the next result:

COROLLARY 13.10. *Suppose $W = M_n(R)$ and $\overline{W} = W/A$, where $A \lhd W$. Then $\overline{W}$ is isomorphic to the matrix ring $M_n(\bar{R})$ for some ring $\bar{R}$, and the natural homomorphism $w \mapsto \bar{w}$ induces a ring surjection $R \to \bar{R}$.*

Proof. Take a set of matrix units $\{e_{ij} : 1 \leq i, j \leq n\}$ of W, and define $\bar{e}_{ij} = e_{ij} + A \in \overline{W}$. Clearly $\{\bar{e}_{ij} : 1 \leq i, j \leq n\}$ satisfies (MU1) and (MU2), and so it is a set of matrix units. Using Proposition 13.9, we see that $\overline{W}$ is a matrix ring over $\bar{R} = \bar{e}_{11}\overline{W}\bar{e}_{11}$; we conclude by recalling that $R \cong e_{11}We_{11}$.

□

Lifting matrix units.

A subtle point which comes up later in the proof of Theorem 15.26 (but could be skipped for now) is whether we can reverse Corollary 13.10; namely given $W/A \cong M_n(T)$, can we "lift" the matrix structure from T, and write W as a matrix ring? Recall that an element $w \in W$ is **nilpotent** if $w^k = 0$ for some $k \geq 1$.

Definition 13.11. We say that a subset $A \subset W$ is **nil** if every element of A is nilpotent.

Remark 13.12. Suppose $A \subset W$ is nil.

(i) The only idempotent $e \in A$ is 0. (Indeed, for some $k \geq 1$, $0 = e^k = e$.)

(ii) $1 - a$ is invertible for each $a \in A$. (Indeed, if $a^k = 0$, then $(1-a)(1+a+\cdots+a^{k-1}) = 1 - a^k = 1$.)

PROPOSITION 13.13. *Suppose $A \lhd W$ is nil. Then any set $\{f_1, \ldots, f_n\}$ of orthogonal idempotents of W/A can be lifted to a set $\{e_1, \ldots, e_n\}$ of orthogonal idempotents of W (i.e., $f_i = e_i + A$). If moreover $W/A \cong M_n(T)$, then W can be written as $M_n(R)$ for a suitable ring R such that there is a surjection $R \to T$ inducing the given surjection $W = M_n(R) \to M_n(T)$.*

Proof. More explicitly, we shall "lift" any given set of matrix units $\{f_{ij} : 1 \leq i, j \leq n\}$ of $\overline{W}$ to a set of matrix units $\{e_{ij} : 1 \leq i, j \leq n\}$ of W, such that $f_{ij} = e_{ij} + A$ for each i, j. To see this, we start by lifting the idempotents $f_i = f_{ii}$ computationally, and then fill them in with the other matrix units.

STEP I. Any idempotent f_1 of $\overline{W} = W/A$ can be lifted to an idempotent of W. Indeed, writing $\bar{\ }$ for the image in $\overline{W}$, take any $w \in W$ such that $\overline{w} = f_1$. Then $w - w^2 \in A$ is nilpotent, so there is some k such that

$$0 = (w - w^2)^k = w^k - w^{k+1}v,$$

for suitable $v \in \mathbb{Z}[w]$. Hence

$$w^k = w^{k+1}v = ww^kv = ww^{k+1}vv = w^{k+2}v^2,$$

and iterating, we see that $w^k = w^{2k}v^k$. Hence, taking $e_1 = w^kv^k$ we have

$$e_1^2 = w^{2k}v^kv^k = w^kv^k = e_1;$$

i.e., e_1 is idempotent. Furthermore,

$$f_1 = f_1^k = \overline{w}^k = \overline{ww^{k+1}v} = f_1^{k+1}\overline{v} = f_1\overline{v},$$

implying that $f_1 = (f_1\overline{v})^k = \overline{w^kv^k} = \overline{e_1}$.

STEP II. Any orthogonal set of idempotents $f_1, \ldots, f_n$ of $\overline{W}$ can be lifted to orthogonal idempotents $e_1, \ldots, e_n$ of W. The proof is by induction on n, the case $n = 1$ being Step I. Inductively, we may assume that we have found $e_1, \ldots, e_{n-1}$. Also, by Step I we can find an idempotent $e \in W$ such that $\overline{e} = f_n$. Let $u = e_1 + \cdots + e_{n-1}$. Then u is idempotent and $\overline{eu} = 0$, so $eu \in A$. By Remark 13.12(ii), $1 - eu$ is invertible; let

$$\tilde{e} = (1-eu)^{-1}e(1-eu),$$

also idempotent with $\overline{\tilde{e}} = f_n$. Define

$$e_n = (1-u)\tilde{e} = (1-u)(1-eu)^{-1}e(1-eu).$$

Noting that $e(1-eu) = e - eu = e(1-u)$, we see that $e_nu = 0 = ue_n$, and

$$\begin{aligned} e_n^2 &= (1-u)(1-eu)^{-1}e(1-eu)(1-u)(1-eu)^{-1}e(1-u) \\ &= (1-u)(1-eu)^{-1}e(1-u)^2(1-eu)^{-1}e(1-u) \\ &= (1-u)(1-eu)^{-1}e(1-u)(1-eu)^{-1}e(1-u) \\ &= (1-u)\tilde{e}^2 = (1-u)\tilde{e} = e_n \end{aligned}$$

is idempotent. Also

$$\overline{e_n} = \left(1 - \sum_{i=1}^{n-1} f_i\right) f_n = f_n.$$

Finally, for all $1 \leq i \leq n-1$, $e_ne_i = e_n(ue_i) = 0$ and $e_ie_n = (e_iu)e_n = 0$.

STEP III. By Step II, we can lift $f_{11}, \ldots, f_{nn}$ to orthogonal idempotents $e_{11}, \ldots, e_{nn}$ of W. $1 - \sum_{i=1}^n e_{ii}$ is an idempotent of A and thus 0, by Remark 13.12(i), implying that $\sum e_{ii} = 1$; we need only fill in the matrix units e_{ij} for $i \neq j$. Take w_{ij} such that $\overline{w_{ij}} = f_{ij}$. Replacing w_{ij} by $e_{ii}w_{ij}e_{jj}$, we may assume that $e_{ii}w_{ij} = w_{ij} = w_{ij}e_{jj}$ for all i, j. For $i \neq 1$, $\overline{e_{11} - w_{1i}w_{i1}} = f_{11} - f_{11} = 0$, so in view of Remark 13.12(ii) we can put

$$e_{1i} = \left(1 - (e_{11} - w_{1i}w_{i1})\right)^{-1} w_{1i},$$

and put $e_{i1} = w_{i1}e_{11}$. Now

$$\begin{aligned} e_{1i}e_{i1} &= (1 - e_{11} + w_{1i}w_{i1})^{-1} w_{1i}w_{i1}e_{11} \\ &= (1 - e_{11} + w_{1i}w_{i1})^{-1} (1 - e_{11} + w_{1i}w_{i1})e_{11} = e_{11}. \end{aligned}$$

Moreover, $e_{ii} - e_{i1}e_{1i}$ is an idempotent of A and thus 0 by Remark 13.12(i); hence $e_{i1}e_{1i} = e_{ii}$. It is now easy to put $e_{ij} = e_{i1}e_{1j}$ for all other $i \neq j$, and check the matrix unit conditions. □

The structure of matrix rings.

We write $A \lhd R$ to indicate that A is a (two-sided) ideal of the ring R. The ideals of R and $M_n(R)$ are closely related:

PROPOSITION 13.14. *There is a lattice isomorphism*

$$\{\textit{Ideals of } R\} \to \{\textit{Ideals of } M_n(R)\}$$

given by $A \mapsto M_n(A)$.

Proof. One checks easily that if $A \lhd R$, then $M_n(A) \lhd M_n(R)$. Thus it suffices to show that every ideal B of $M_n(R)$ has the form $M_n(A)$, where $A \lhd R$. We define $A = \{r \in R : re_{11} \in B\}$, clearly an ideal of R. It remains to show that $M_n(A) = B$:

($\subseteq$) For any a in A and any i, j, we have

$$ae_{ij} = e_{i1}(ae_{11})e_{1j} \in B,$$

implying that $M_n(A) \subseteq B$.

($\supseteq$) Given any element $b = \sum b_{uv}e_{uv} \in B$ and any i, j, we have

$$b_{ij}e_{11} = e_{1i}be_{j1} \in B,$$

implying that each $b_{ij} \in A$. □

Let us consider some (but not all) of the left ideals of $W = M_n(R)$. Clearly $Ww \leq W$ for any w in W, and this already provides many examples; cf. Exercise 2.

Remark 13.15. Suppose $L \leq R$.

(i) $Le_{ik} = Le_{ij}e_{jk} \subseteq M_n(L)e_{jk}$, for all i, j, k.

(ii) $M_n(L)e_{uk} = \sum_{i=1}^n \sum_{j=1}^n Le_{ij}e_{uk} = \sum_{i=1}^n Le_{ik}$ for all u, k. In particular, $M_n(L)e_{kk} = \sum Le_{ik} = M_n(L)e_{uk}$, for all u, k.

(iii) $M_n(L)e_{ii} \cong M_n(L)e_{ij} = M_n(L)e_{jj}$. (The isomorphism is given by the map $ae_{ii} \mapsto ae_{ij}$, and the equality holds by (ii).)

Recall that a **division ring** (or **skew field**) is a ring D in which every nonzero element is invertible. Thus, each division ring D has no proper nonzero left ideals or right ideals. For example, any field is a division ring. To verify that a ring D is a division ring, it suffices to prove that each element of $D\backslash\{0\}$ is left invertible, since then $D\backslash\{0\}$ is a group; cf. [**Row3, Lemma 1.8**]. We return to division rings in Chapters 14 and 24; in particular, a noncommutative division ring is exhibited in Example 14.29.

The theory of vector spaces over fields extends to modules over division rings.

Remark 13.16. Any module M over a division ring D is free; indeed, any maximal independent subset is a base. This is an immediate consequence of Zorn's Lemma, but in fact we already proved it in Chapter 6 of Volume 1, since the verification of Example 6.3 did not use the commutativity of the underlying field. Hence, by Theorem 6.8 of Volume 1, any two bases of M over D have the same cardinality, which we denote as $[M : D]$. If F is a subfield of D, then the same argument as in the proof of Theorem 4.4 of Volume 1 shows that

$$[M : F] = [M : D][D : F].$$

Minimal left ideal will mean minimal as a *nonzero* left ideal. Thus a nonzero left ideal is simple as an R-module iff it is a minimal left ideal.

PROPOSITION 13.17. *Suppose $W = M_n(D)$, where D is a division ring. Then the left ideal $L = We_{jj}$ of W is minimal, for any $1 \leq j \leq n$.*

Proof. Note that $L = \sum_{i=1}^n De_{ij}$. It suffices to show for any given nonzero element $a = \sum_{u=1}^n d_{uj}e_{uj} \in L$ that $L \subseteq Wa$. Take v such that $d_{vj} \neq 0$. Then

$$d_{vj}e_{ij} = e_{iv}\sum_{u=1}^n d_{uj}e_{uj} \in Wa,$$

so $De_{ij} \subseteq Wa$ for each i; hence $L \subseteq Wa$. □

COROLLARY 13.18. *If D is a division ring, then $R = M_n(D)$ is a direct sum of minimal left ideals. Hence, R has composition length n as an R-module.*

Proof. $M_n(D) = \bigoplus_j M_n(D)e_{jj}$, so we use Remark 3.15 of Volume 1. □

Basic notions for noncommutative rings

When generalizing elementary definitions from commutative ring theory to noncommutative rings, we often formulate the definitions in terms of ideals, since ideals, being kernels of ring homomorphisms, lie at the foundation of the structure theory.

The opposite ring, and left-right duality.

The lack of left-right symmetry in the definition of module is a cause of concern — from time to time we need to use the right-handed versions of our previous theorems. Although intuition tells us that the right-handed and left-handed theories should be analogous, there is a formalism that gives us this correspondence at once.

Definition 13.19. The **opposite ring** R^{op} has the same additive structure as the original ring R, but with multiplication in the reverse order; i.e., the new product $a \cdot b$ is defined as the old product ba.

It is easy to see that R^{op} is a ring, since the ring axioms are left-right symmetric; also $(R^{\mathrm{op}})^{\mathrm{op}} = R$. Clearly $R^{\mathrm{op}} = R$ when R is commutative.

Remark 13.20. The left R-modules are precisely the right R^{op}-modules. Explicitly, if M is a left R-module, then M is also a right R^{op}-module, where the new product $a \cdot r$ is taken to be the old product ra. Thus, any general theorem for left modules applied to R^{op} would give the same theorem for right modules over R. Of course, when $R \cong R^{\mathrm{op}}$, then the lattices of left R-modules and of right R-modules are isomorphic.

Remark 13.21. There is a natural isomorphism $M_n(R)^{\mathrm{op}} \to M_n(R^{\mathrm{op}})$, sending a matrix to its transpose (since the transpose also reverses the order of multiplication). In particular, the right composition length $M_n(D)$ (for D a division ring) equals the left composition length of $M_n(D^{\mathrm{op}})$, which also is n. See Exercise 16 for an example of left-right asymmetry.

Simple rings.

Definition 13.22. A ring R is **simple** if it has no proper nonzero ideals.

Simple rings are the building blocks of the structure theory of rings — any homomorphism from a simple ring to an arbitrary ring is an injection (for the kernel, being an ideal, must be 0). The commutative simple rings are just the fields. However, in the noncommutative theory, the situation is much more intricate.

A note of caution: We have come across many definitions in which a ring R is defined to have some property P if R has this property P as an R-module. Left Artinian and left Noetherian are such properties; in other words, a ring is **left Artinian**, resp. **left Noetherian**, if it satisfies the DCC (rep. ACC) on left ideals. One exception to this linguistic rule is "simple." In view of Proposition 13.2, a ring R is simple as an R-module iff $Ra = R$ for every $0 \neq a \in R$, i.e., if R is a division ring. However, since the definition of simple rings involves lack of ideals, not of left ideals, there are many simple rings that are not division rings, to wit:

Example 13.23. Any division ring D is simple, so by Proposition 13.14, $R = M_n(D)$ is simple. (Nevertheless, R has many proper *left* ideals, namely Ra for any noninvertible $a \in R$.)

The ring $R = M_n(D)$ is also left Artinian and left Noetherian, by Corollary 13.18. The analogous argument for right modules (made precise by Remark 13.21) shows that R also is a right Artinian and right Noetherian ring. For convenience, a ring that is both left and right Artinian is called "Artinian"; likewise for "Noetherian." Thus $M_n(D)$ is a simple Artinian (and Noetherian) ring.

It is not easy to construct simple rings that do **not** have the form $M_n(D)$; one such example is given in Example 13A.1.

The center of a ring.

Definition 13.24. The **center** of a ring R, denoted $\mathrm{Cent}(R)$, is $\{c \in R : cr = rc \text{ for all } r \text{ in } R\}$.

The center enables us to utilize results from commutative algebra in the noncommutative theory.

Remark 13.24′. (i) $\mathrm{Cent}(R)$ is a commutative ring.

(ii) $Rc \lhd R$, for any $c \in \mathrm{Cent}(R)$.

(iii) If R is a simple ring, then $\mathrm{Cent}(R)$ is a field.

(Proof of (iii): If $0 \neq c \in \mathrm{Cent}(R)$, then $1 \in Rc$ by (ii), implying that $c^{-1} \in R$. Furthermore,

$$c^{-1}r = c^{-1}rcc^{-1} = c^{-1}crc^{-1} = rc^{-1},$$

for all r in R, proving that $c^{-1} \in \mathrm{Cent}(R)$.)

Direct products of rings

Definition 13.25. The **direct product** $\prod_{i\in I} R_i$ of rings $\{R_i : i \in I\}$ is the Cartesian product, with the ring operations taken componentwise. For a finite set $I = \{1, \dots, t\}$, we write $\prod_{i\in I} R_i$ as $R_1 \times \cdots \times R_t$, the unit element being $(1, \dots, 1)$ and the zero element being $(0, \dots, 0)$. As in Remark 2.14 of Volume 1, we always let $\pi_i \colon R \to R_i$ denote the natural projection onto the i-th component; in the other direction, there is a natural map $\nu_i \colon R_i \to R$ sending r to the element $(\dots, 0, r, 0, \dots)$, where each component except the i-th component is 0.

Direct products of rings are of utmost importance in ring theory. They can be described in terms of central idempotents, i.e., idempotents of $\mathrm{Cent}(R)$.

Remark 13.26. Suppose $R = \prod_{i\in I} R_i$. Each projection $\pi_i\colon R \to R_i$ is a surjection of rings. (However, $\nu_i\colon R_i \to R$ is not a ring homomorphism, since the unit element of R_i is sent to $(\dots, 0, 1, 0, \dots)$, which is not the unit element of R.) Let $e_i = \pi_i(1)$. Then the following properties hold:

(i) $e_ie_j = \pi_i(1)\pi_j(1) = \delta_{ij}e_i$. Hence $\{e_i : i \in I\}$ is a set of orthogonal idempotents, and furthermore are central since

$$re_i = \pi_i(r)e_i = \pi_i(r)\pi_i(1) = \pi_i(r) = \pi_i(1)\pi_i(r) = e_i\pi_i(r) = e_ir.$$

In the case when I is finite, then $\sum_{i\in I} e_i = 1$.

(ii) $R_i \cong Re_i$ as rings (where e_i is the unit element of Re_i).

(iii) If $L_i \le R_i$, then $\prod L_i \le \prod R_i$.

Conversely, we can describe the structure of R in terms of arbitrary central idempotents.

Remark 13.27. Suppose e is a central idempotent of a ring R. Then Re is a ring with multiplicative unit e, and there is a ring surjection $R \to Re$, given by $r \mapsto re$ whose kernel is $R(1-e)$. (For $re = 0$ iff $r = r(1-e)$.) Likewise, there is a ring surjection $R \to R(1-e)$ whose kernel is Re.

Together, $R \cong Re \times R(1-e)$, via the map $r \mapsto (re, r(1-e))$. (This isomorphism of rings can also be viewed via the Peirce decomposition of Remark 13.6, since $eR(1-e) = e(1-e)R = 0 = (1-e)Re$.)

Applying induction on t to Remark 13.27 yields

Remark 13.28. If $e_1, \dots, e_t$ are a 1-sum set of orthogonal central idempotents, then

$$R \cong R_1 \times \cdots \times R_t,$$

where R_i is the ring Re_i.

Remark 13.29. The Chinese Remainder Theorem (Theorem 0.3 of Volume 1) says that if $P_1, \dots, P_t$ are distinct maximal ideals of R, then

$$R/\bigcap_{i=1}^{t} P_i \cong R/P_1 \times \cdots \times R/P_t,$$

a direct product of simple images of R. (Our most common application is for R left Artinian, so that all R/P_i are simple Artinian.)

Digression 13.30. A reservation: Although the Chinese Remainder Theorem can be a very powerful tool, there are many instances for which it is inapplicable. For example, the intersection of all the maximal ideals of $\mathbb{Z}$ is 0, but any finite intersection is nonzero. Still, one can say something in such a situation.

If $A_i \lhd R$ and $\bigcap A_i = 0$, we say that R is a **subdirect product** of the R/A_i. In this case, by Remark 0.4 of Volume 1, there is a canonical injection $R \to \prod R/A_i$. For example, $\mathbb{Z}$ is a subdirect product of the fields $\mathbb{Z}/p\mathbb{Z}$, for all primes p. We return to this idea later, in Chapters 15 and 16.

Modules over finite direct products.

Recall that for modules, finite direct products and finite direct sums are the same, and here we write $\oplus$ instead of $\times$. The reader might feel at first that this notation is designed for maximal confusion, but actually there is a valid reason that becomes clearer if one considers infinite index sets. For modules, which generalize vector spaces, the direct sum is the more natural construction, whereas for rings, the direct product is more natural, since the direct sum of an infinite set of rings will fail to have a unit element.

PROPOSITION 13.31. *Suppose $R = R_1 \times R_2 \times \cdots \times R_t$.*

(i) *If M_i are R_i-modules for $1 \le i \le t$, then $M_1 \oplus \cdots \oplus M_t$ is an R-module, under the natural action*

$$(r_1, \dots, r_t)(a_1, \dots, a_t) = (r_1a_1, \dots, r_ta_t).$$

(ii) *Conversely, let $e_1, \dots, e_t$ be a 1-sum set of orthogonal central idempotents of R, with $R_i = Re_i$. If M is an R-module, then $M_i = e_iM$ is an R_i-module, and $M \cong M_1 \oplus \cdots \oplus M_t$ as R-modules.*

(iii) *Any left ideal L of $R_1 \times \cdots \times R_t$ has the form $L_1 \times \cdots \times L_t$ with $L_i = Le_i \le R_i$. Likewise, the ideals of $R_1 \times \cdots \times R_t$ have the form $A_1 \times \cdots \times A_t$ where $A_i \lhd R_i$.*

Proof. (i) is clear.

(ii) First we show the M_i are independent. Indeed, if $\sum_{i=1}^{t} a_i = 0$ for $a_i \in M_i$, then $a_i = e_ia_i$ and thus

$$0 = e_j \sum_{i=1}^{t} a_i = \sum_{i=1}^{t} e_je_ia_i = e_ja_j = a_j$$

for each j, proving independence.

Hence $M_1 \oplus \cdots \oplus M_t$ is identified with $\sum_{i=1}^{t} M_i \subseteq M$. On the other hand, for any $a \in M$ we have

$$a = \Big(\sum_{i=1}^{t} e_i\Big)a \in \sum e_iM = \sum M_i,$$

yielding equality.

(iii) If $a=(a_i)\in L$, then $a=\sum ae_i$. □

Note that for I infinite, the ideal structure of $\prod R_i$ is much more complicated, including ideals such as $\{(r_i) : \text{almost all } r_i \text{ are } 0\}$.

Remark 13.32. (i) Taking $M_i=e_iM$ as in Proposition 13.31, we note that $e_jM_i=e_je_iM=0$ for all $j\neq i$. Hence, the lattices of submodules of M_i as R-modules and as R_i-modules are the same. In particular, a submodule of M_i is simple as an R-module iff it is simple as an R_i-module.

(ii) Any simple submodule S of M is contained in some M_i. Indeed, e_iS must be S or 0. If each $e_iS=0$, then $S=(\sum e_i)S=0$, so some $e_iS=S$.

We are ready for the main point of this discussion.

Example 13.33. If $R=R_1\times\cdots\times R_t$ with each $R_i=M_{n_i}(D_i)$ for D_i a division ring, then R is the direct sum of minimal left ideals. Indeed, R is the direct sum of R_i as R-modules, and by Corollary 13.18, each R_i is the direct sum of minimal left ideals, which then are minimal as R_i-modules and thus as R-modules.

Furthermore, we have

PROPOSITION 13.34. *Suppose e_1,e_2 are orthogonal central idempotents of R. If $L_i\leq Re_i$ for $i=1,2$, then the only R-module map $L_1\to L_2$ is 0.*

Proof. For any map $f\colon L_1\to L_2$, we see for each $a\in L_1$ that

$$f(a)=e_2f(a)=e_2f(e_1a)=e_2e_1f(a)=0.$$ □

For example, for $R=F\times F$, there is no nonzero R-module map from $F\times 0$ to $0\times F$.

Note that idempotents have entered the theory both via matrix rings (as the e_{ii}) and via direct products (as central idempotents).

The structure of Hom(M,N)

As indicated in the Introduction, we turn now to the set of maps between modules over a given ring R, as a tool in the study of modules and of R itself. It is convenient to consider maps between different modules.

Definition 13.35. Given R-modules M,N, define $\mathrm{Hom}_R(M,N)$ to be the set of R-module homomorphisms from M to N, made into an Abelian group by pointwise addition, i.e., $(f+g)(a)=f(a)+g(a)$.

When understood, R may be deleted from the notation. $\mathrm{Hom}(M,M)$ is a ring whose multiplication is the composition of maps. To emphasize the ring structure, we write $\mathrm{End}_R M$ instead of $\mathrm{Hom}(M,M)$. End is short for "endomorphism," which means "homomorphism to itself."

Analogously, for right modules M,N over a ring W, we define the Abelian group $\mathrm{Hom}(M,N)_W$ of right W-module homomorphisms, and we write $\mathrm{End}\,M_W$ for the ring $\mathrm{Hom}(M,M)_W$. This notion is most useful when we consider left and right module structures simultaneously.

Definition 13.35′. Given rings R and W, we say that M is an R,W-**bimodule** if M is both a left R-module and right W-module, satisfying the associative law

$$(ra)w=r(aw),\ \forall a\in M,\ r\in R,\ w\in W. \tag{13.1}$$

This is a direct generalization from the commutative theory, since a module M over a commutative ring C is naturally a C,C-bimodule, where one defines ac to be ca for $c\in C$, $a\in M$. Here are other natural instances of bimodules.

Example 13.35″. (i) Any ring R itself is an R,R-bimodule; the subbimodules of R are precisely the (two-sided) ideals of R. Thus, bimodules are a way to involve the ring structure directly via module-theoretic techniques.

(ii) More generally, given a ring homomorphism $\psi: W\to R$, we can view R as an R,W-bimodule by taking r_1r_2w to be $r_1r_2\psi(w)$, evaluated in R.

(iii) As a special case of (ii), if R is a C-algebra, then we can view R as an R,C-bimodule. (In particular, any ring R is an $R,\mathbb{Z}$-bimodule.) Soon we shall come across other interesting examples of bimodules. Meanwhile, we want to focus on one particular aspect of modules.

Definition 13.36. For any module M and any $S\subseteq M$, define the **(left) annihilator**

$$\mathrm{Ann}_R S=\bigcap\{\mathrm{Ann}_R s : s\in S\}=\{r\in R : rS=0\}.$$

The module M is called **faithful** if $\mathrm{Ann}_R M=0$.

Remark 13.36′. (i) $\mathrm{Ann}_R M\lhd R$, for any nonzero R-module M. (Indeed, by definition, $\mathrm{Ann}\,M\leq R$, and $1\notin\mathrm{Ann}\,M$ since $1=1\cdot 1\neq 0$, so we must show that $rs\in\mathrm{Ann}\,M$ for all r in $\mathrm{Ann}\,M$ and s in R. But

$$rsM=r(sM)\subseteq rM=0,$$

yielding $rs\in\mathrm{Ann}\,M$.)

(ii) Every nonzero module M over a simple ring R is faithful; indeed, $\operatorname{Ann} M \lhd R$ implies $\operatorname{Ann} M = 0$.

The reason for the terminology "faithful" comes from the next observation. We write $\ell_r \colon M \to M$ for the left multiplication map $a \mapsto ra$.

PROPOSITION 13.37. *Suppose M is an R, W-bimodule. There is a ring homomorphism $\phi \colon R \to \operatorname{End} M_W$ given by $\phi(r) = \ell_r$. Furthermore, $\ker \phi = \operatorname{Ann}_R M$. Thus, ϕ is an injection iff M is faithful as an R-module, which is true in particular if $M = R$.*

Proof. Clearly, $\ell(r_1 + r_2) = \ell(r_1) + \ell(r_2)$. Likewise

$$\ell(r_1 r_2)(a) = r_1 r_2 a = \ell(r_1)(r_2 a) = \ell(r_1)\ell(r_2)(a),$$

proving that ϕ is a homomorphism. Also, $r \in \ker \phi$ iff $\ell(r) = 0$, iff $rM = 0$.

The last assertion is true because no element in R annihilates $1 \in R$ (since $r1 = r \neq 0$). □

Representations of rings and algebras

The conclusion of Proposition 13.37 is so important that we frame it with a definition.

Definition 13.38. A **representation** of a ring (resp. algebra) R is a homomorphism $\Phi \colon R \to \operatorname{End} M_W$, where M is a right module over the ring (resp. algebra) W. The representation Φ is **faithful** if $\ker \Phi = 0$.

Schur's Lemma.

We obtain some striking applications when M is simple.

PROPOSITION 13.39 (SCHUR'S LEMMA) .

(i) *If $f \colon M \to N$ is a map of modules with M simple, then either $f = 0$ or f is monic.*
(ii) *If $f \colon M \to N$ is a map of modules with N simple, then either $f = 0$ or f is onto.*
(iii) *Every nonzero map $f \colon M \to N$ between simple modules is an isomorphism.*

Proof. (i) $\ker f$ is a submodule of M, and thus is 0 or M.

(ii) Likewise, $f(M)$ is a submodule of N, and thus is 0 or N.

(iii) By (i) and (ii). □

PROPOSITION 13.40 (ANOTHER FORMULATION OF SCHUR'S LEMMA). *If M is a simple module, then $\operatorname{End}_R M$ is a division ring.*

Proof. Every nonzero element is invertible, by Proposition 13.39(iii). □

For a more explicit formulation over algebraically closed fields, see Proposition 14.28.

Here is another easy but useful application of Proposition 13.37.

Example 13.41. (i) Taking $R = M = W$, we have $R \cong \operatorname{End} R_R$. Indeed, Proposition 13.37 shows that $\phi \colon R \to \operatorname{End} R_R$ is an injection of rings, so it remains to verify that ϕ is onto; i.e., any right R-module map $f \colon R \to R$ is some ℓ_r. But taking $r = f(1)$, we see for all $a \in R$ that

$$\ell_r(a) = ra = f(1)a = f(1a) = f(a),$$

proving that $\ell_r = f$ as desired.

(ii) Similarly $\operatorname{End}_R R \cong R^{\mathrm{op}}$.

When M is free over W, we get matrix rings.

PROPOSITION 13.42 (GENERALIZING EXAMPLE 13.41).

(i) *The free right W-module $M = W^{(n)}$ is an $M_n(W), W$-bimodule, and the natural map $\phi \colon M_n(W) \to \operatorname{End} W^{(n)}{}_W$ is an isomorphism of rings.*
(ii) *Similarly, $M_n(W) \cong (\operatorname{End}_W W^{(n)})^{\mathrm{op}}$.*

Proof. (i) Let $R = M_n(W)$. M is a faithful R-module, where multiplication is the usual matrix multiplication between a matrix and a vector; hence ϕ of Proposition 13.37 is an injection. It remains to show ϕ is onto, i.e., any right W-module map $f \colon M \to M$ has the form ℓ_r. Let $e_1, \ldots, e_n$ be the standard W-base for M. Then f is given by its action on the base, i.e.,

$$f(e_j) = \sum_i e_i w_{ij}$$

for suitable w_{ij} in W. Let $r = (w_{ij}) \in M_n(W)$. For any $v = \sum_j e_j w_j \in M$, we have

$$rv = \sum_i \left(e_i \sum_j w_{ij} w_j \right) = \sum_j f(e_j) w_j = f\left(\sum e_j w_j \right) = f(v),$$

proving that $f = \ell_r$.

(ii) Same as (i), using the opposite ring to pass from right to left modules.

□

This result improves Proposition 2.30 (also see Remark 13.45 below) and provides an alternate proof that $M_n(W)$ is indeed a ring. In particular, if M is a f.g. right module over a division ring D, then $\operatorname{End} M_D \cong M_n(D)$ for some n. A useful extension is given in Exercise 20.

We can also go in the other direction, to recapture a ring from its matrix ring via End. Idempotents also play a key role here. Again, for technical reasons, we pass to the opposite ring, because we use left modules.

PROPOSITION 13.43.

(i) *If e is an idempotent of R, then $\operatorname{End}_R(Re) \cong (eRe)^{\mathrm{op}}$. (Note that we recover Example 13.41 by taking $e = 1$.)*

(ii) *If $R = M_n(W)$, then $W \cong (\operatorname{End}_R(Re_{11}))^{\mathrm{op}}$.*

Proof. (i) Let $W = eRe$ and $M = Re$, an R, W-bimodule and thus a $W^{\mathrm{op}}, R^{\mathrm{op}}$ bimodule. Proposition 13.37 gives us an injection $W^{\mathrm{op}} \to \operatorname{End}_R M$, which is onto in view of the map in the reverse direction given by $f \mapsto ef(e)$ for $f \in \operatorname{End}_R M$.

(ii) Immediate from (i), taking $e = e_{11}$. □

In the case W is a field, one way of understanding Proposition 13.43 is to view the module endomorphisms as linear transformations that commute with all left multiplications by elements of R. But the only such matrices are scalar matrices. Thus, Proposition 13.43 is a converse of sorts for Schur's Lemma.

Hom for direct sums.

Let us consider a more general situation: We take direct sums of arbitrary right W-modules.

PROPOSITION 13.44.

(i) *If $\{M_i : i \in I\}$ and $\{N_j : j \in J\}$ are right W-modules, then*

$$\operatorname{Hom}\left(\bigoplus_{i\in I} M_i, \bigoplus_{j\in J} N_j\right)_W \cong \bigoplus_{i,j} \operatorname{Hom}(M_i, N_j)_W$$

as additive groups.

(ii) $\operatorname{Hom}(M^{(m)}, N^{(n)})_W \cong M_{m,n}(\operatorname{Hom}(M, N)_W)$ *as groups, for any right W-modules M, N.*

Proof. Let $M = \oplus M_i$ and $N = \oplus N_j$. Given $f\colon M \to N$, define

$$f_{k\ell} = \pi_\ell f \nu_k \colon M_k \to N_\ell,$$

where $\nu_k\colon M_k \to \oplus M_i$ and $\pi_\ell\colon \oplus N_j \to N_\ell$ are the canonical embeddings and projections; cf. Remark 2.14 of Volume 1. This gives us a map $f \mapsto (f_{k\ell})$. Conversely, given $f_{ij}\colon M_i \to N_j$ for each i, j, define

$$f = \sum_{i,j} \nu_j f_{ij} \pi_i \colon M \to N. \tag{13.2}$$

(For any $a \in M$, $\pi_i(a)$ is zero for almost all i, so the sum in (13.2) is defined.) To see that these correspondences are inverses, note that

$$\pi_\ell\Big(\sum_{i,j} \nu_j f_{ij} \pi_i\Big)\nu_k = \pi_\ell \nu_\ell f_{k\ell} \pi_k \nu_k = f_{k\ell};$$

$$\sum_{i,j} \nu_j(\pi_j f \nu_i)\pi_i = \Big(\sum_j \nu_j \pi_j\Big) f \Big(\sum_i \nu_i \pi_i\Big) = f.$$

(ii) Immediate from (i). □

Remark 13.45. Now we can view Proposition 3.16 of Volume 1 more structurally. Suppose M and N are R, W-bimodules. Then $\operatorname{Hom}(M, N)_W$ is an R-module in the natural way, i.e.,

$$(rf)(a) = rf(a), \qquad (a \in M).$$

In this setting, Proposition 13.44 can be formulated in terms of maps of R-modules, when the M_i, N_j are R, W-bimodules. In particular,

$$\operatorname{Hom}(R^{(m)}, R^{(n)})_R \cong M_{m,n}(R)$$

as R-modules, with base $\{f_{ij} : 1 \le i \le m,\ 1 \le j \le n\}$.

COROLLARY 13.46. *Suppose $M_i = S_i^{(n_i)}$ for simple modules S_i. Then $\operatorname{Hom}(M_i, M_j) = 0$ unless $S_i \cong S_j$, in which case, for the division ring $D_i = \operatorname{End}_R S_i$,*

$$\operatorname{Hom}(M_i, M_j) \cong M_{n_i, n_j}(D_i)$$

as right D_i-modules, and

$$\operatorname{Hom}(M_i, M_i) \cong M_{n_i}(D_i)$$

as rings.

Proof. By Proposition 13.44,

$$\mathrm{Hom}(M_i, M_j) \cong M_{n_i,n_j}(A)$$

where $A = \mathrm{Hom}(S_i, S_j)$, which by Schur's Lemma is either 0 or D_i (when $S_i \cong S_j$). We conclude with Proposition 13.42. □

Putting everything together, we have

THEOREM 13.47. *Suppose $M_i = S_i^{(n_i)}$ for $1 \le i \le t$, where the S_i are simple pairwise nonisomorphic R-modules, and $M = M_1 \oplus \cdots \oplus M_t$. Then*

$$\mathrm{End}_R M \cong \prod_{i=1}^{t} M_{n_i}(D_i),$$

where $D_i = \mathrm{End}\, S_i$.

Proof. $\mathrm{End}_R M \cong \bigoplus_{i,j} \mathrm{Hom}(M_i, M_j) = \bigoplus_i \mathrm{Hom}(M_i, M_i)$, by Corollary 13.46. But then the nonzero components come in t $n_i \times n_i$ blocks along the diagonal, so we see as rings that

$$\mathrm{End}_R M \cong \prod_{i=1}^{t} \mathrm{Hom}(M_i, M_i) \cong \prod_{i=1}^{t} M_{n_i}(D_i). \qquad \square$$

We are ready for an important result that foreshadows Chapter 14.

THEOREM 13.48. *A ring R is a finite direct sum of minimal left ideals iff $R \cong \prod_{i=1}^{t} M_{n_i}(D_i)$ for suitable t, suitable $n_i \in \mathbb{N}$, and division rings D_i.*

Furthermore, $R \cong L^{(n)}$ for a single minimal left ideal L iff $t = 1$. In this case, $R \cong M_n(D)$ for the division algebra $D = \mathrm{End}_R L$.

Proof. ($\Leftarrow$) By Example 13.33.

($\Rightarrow$) $R \cong \mathrm{Hom}(R, R)$ by the right-hand version of Example 13.41, so we apply Theorem 13.47.

The second assertion is by Remark 13.15(iii) and Theorem 13.47. □

The regular representation of an algebra

We return to an instance of Definition 13.38, generalizing Example 13.41, that has resounding repercussions. Note the parallel to Cayley's Theorem from group theory.

Example 13.49 (The regular representation). For any algebra R over a commutative ring C, there is an algebra injection $\phi\colon R \to \mathrm{End}_C R$, sending each element $r \in R$ to the left multiplication map ℓ_r. (The justification comes from Proposition 13.37, taking $M = R$ and $W = C$, noting that $\ell_{cr}(a) = c\ell_r(a)$.)

In particular, when C is a field, we take a base $\{b_i : i \in I\}$ of R over F, and identify ℓ_r as a linear transformation.

The regular representation was utilized several times already in Volume 1 to obtain the rational form (preceding Theorem 2.65), the regular representation of a field (Definition 4.107), and our treatment of integral elements (Theorem 5.21). Several important noncommutative examples will be given in Appendix 13A, and the easiest example of a noncommutative division ring is the algebra of quaternions (Example 14.29). The computability of the regular representation depends on an appropriate choice of base B, as we see in those examples.

Note that the algebra structure of $\mathrm{End}_C R$ depends only on the structure of R as a C-module, but *not* on the given multiplication of R. Nevertheless, taking $C = \mathbb{Z}$, we see that any ring R is isomorphic to some subring of $\mathrm{End}_{\mathbb{Z}} R$.

Ironically, as in Cayley's Theorem, the main applications are obtained by arguing from the converse. Namely, suppose we want to exhibit an algebra structure for a C-module R having a designated element 1. It is enough to find a monic C-module map $\phi\colon R \to \mathrm{End}_C R$ satisfying $\phi(1) = 1$ and $\phi(r)\phi(s) = \phi(rs)$ for all r, s in R. (This is called *exhibiting R as a ring via the regular representation*.) Then associativity and distributivity of multiplication over addition follow from the corresponding properties of $\mathrm{End}_C R$, and do not need to be verified directly. It is convenient to formulate this in slightly greater generality.

Remark 13.50. Suppose W is a C-algebra and M is a right W-module. If R is any C-module and there is a 1:1 homomorphism $\rho\colon R \to \mathrm{End}\, M_W$, then R becomes a C-algebra, as seen by using Proposition 13.37 to transfer the necessary axioms from $\mathrm{End}\, M_W$ by means of ρ. In particular, for any subring W of R, we can embed R into $\mathrm{Hom}(R, R)_W$, since R is an R, W-bimodule; if M happens to be a free W-module of rank n, we have exhibited R as a subring of $M_n(W)$.

Supplement: Polynomial rings

Polynomial algebras play such an important role in commutative algebra that we would expect them to be prominent also in the noncommutative

theory. Although they are not as crucial in this setting, they do eventually play an important role. The definition is as expected:

Definition 13.51. (i) The **polynomial ring** $R[\lambda]$ (over a ring R) is defined to be the free R-module with base $\{\lambda^i : i \in \mathbb{N}\}$, also endowed with multiplication given by

$$\Big(\sum_{i\in\mathbb{N}} r_i\lambda^i\Big)\Big(\sum_{j\in\mathbb{N}} s_j\lambda^j\Big) = \sum_{k\in\mathbb{N}}\Big(\sum_{i+j=k} r_i s_j\Big)\lambda^k. \tag{13.3}$$

(ii) The **ring of formal power series** $R[[\lambda]]$ is defined to be the set of formal sums $\{\sum_{i\in\mathbb{N}} r_i\lambda^i : r_i \in R\}$, with addition defined componentwise and multiplication given by (13.3). (Note that the formula (13.3) suffices, since each coefficient in the product involves only a finite number of summands. Verification of associativity and distributivity is exactly the same as for polynomials.)

Thus, $R[\lambda]$ consists of all elements of $R[[\lambda]]$ for which almost all coefficients are 0, so we can view $R[\lambda] \subset R[[\lambda]]$. We would like to generalize the fact that $F[\lambda]$ is a PID for any field F.

Definition 13.52. A domain is a **principal left ideal domain**, or PLID, if every left ideal is principal.

Proposition 13.53. *If D is a division ring, then the polynomial ring $D[\lambda]$ satisfies the Euclidean algorithm, and thus is a PLID. In fact, any left ideal L of $D[\lambda]$ can be written $L = D[\lambda]g$, taking any $0 \neq g \in L$ of minimal degree. If L is a 2-sided ideal, then taking g monic, we have $g \in \mathrm{Cent}(D)[\lambda]$.*

Proof. This follows just as in the commutative theory, by examining the degree function on the polynomial. First we verify the Euclidean algorithm. Given polynomials $f, g \in D[\lambda]$, we need to show that $f = qg + r$, where $q, r \in D[\lambda]$ with $r = 0$ or $\deg r < \deg g$. It is easy to reduce to the case when f, g are monic, since any nonzero coefficient is invertible. If $\deg f = n \geq \deg g = m$, then $\deg(f - \lambda^{n-m}g) < n$, so by induction on n,

$$f - \lambda^{n-m}g = q_1 g + r,$$

implying $f = (\lambda^{n-m} + q_1)g + r$, as desired.

Next we show that $L = D[\lambda]g$. For any $f \in L$ we write $f = qg + r$ by the Euclidean algorithm. Then $r = f - qg \in L$, so by minimality of $\deg g$ we must have $r = 0$, i.e., $f = qg$.

If L is a two-sided ideal, then for any $d \in D$, $dg - gd \in L$ has lower degree than g, and thus must be 0; hence, the coefficients of g are in $\mathrm{Cent}(D)$. □

Remark 13.54. (i) If two cyclic modules R/Ra and R/Rb are isomorphic, then there are $u, v \in R$ such that $au = vb$. (Indeed, take $f\colon R \to R$ such that $f(Ra) = Rb$. Then $u = f(1)$ satisfies $au = f(a) \in Rb$, so $au = vb$ for some v in R.)

(ii) Continuing (i) for the case $R = D[\lambda]$, let $m = \deg a$, viewing a as a polynomial over D. The rank of R/Ra over D (cf. Remark 13.16) equals m, which likewise equals $\deg b$. Writing $u = qb + u'$ with $\deg u' < \deg b = m$, we have $aqb + au' = au = vb$, implying

$$au' = (v - aq)b.$$

Thus, replacing u by u' and v by $v - aq$, we obtain $\deg u = \deg v < m$ in (i).

Much of the theory of modules over a PID (Chapter 3) passes easily to the noncommutative setting; cf. Exercises 21–23. (For a fuller picture, see Example 13A.3 and the exercises to Appendix 13A.)

Appendix 13A. Ring constructions using the regular representation

Digressing slightly, let us illustrate the regular representation by constructing some examples, generalizing polynomial rings, that pervade the more general noncommutative ring theory.

Weyl algebras.

We start with a simple ring having surprising properties. Given any $a, b \in R$, we define the **ring commutator** $[a, b] = ab - ba$. One calculates easily that

$$[a, br] = [a, b]r + b[a, r]. \tag{13A.1}$$

In particular, if a and r commute, then $[a, br] = [a, b]r$.

Example 13A.1. Let F be a field. The **Weyl algebra** $\mathcal{A}_1(F)$ has the F-vector space structure of the commutative polynomial algebra $F[\lambda, \mu]$, but with multiplication satisfying

$$\mu\lambda - \lambda\mu = 1. \tag{13A.2}$$

Equation (13A.2) could be reformulated as

$$\mu\lambda = \lambda\mu + 1, \tag{13A.3}$$

which we could use to define multiplication on all of $\mathcal{A}_1(F)$ using the axioms of an associative F-algebra; we could perform multiplication simply by

juxtaposing monomials and then moving all occurrences of λ to the left of μ using Equation (13A.3). This approach would require direct verification of the properties of an algebra.

Alternatively, we construct $\mathcal{A}_1(F)$ explicitly by means of the regular representation as the F-subalgebra of $\mathrm{End}_F \mathcal{A}_1(F)$ generated by ℓ_λ and ℓ_μ, defined via their action on the base $\lambda^i\mu^j$ of $\mathcal{A}_1(F)$ as vector space over F. Thus ℓ_λ satisfies

$$\ell_\lambda(\lambda^i\mu^j) = \lambda^{i+1}\mu^j. \tag{13A.4}$$

The definition of ℓ_μ is a bit trickier. Indeed, Equation (13A.2) says that $[\mu, \lambda] = 1$. We claim that $[\mu, \lambda^i] = i\lambda^{i-1}$ for all $i \leq 1$. In fact, using Equation (13A.1), we get inductively

$$[\mu, \lambda^{i+1}] = [\mu, \lambda^i]\lambda + \lambda^i[\mu, \lambda] = i\lambda^i + \lambda^i = (i+1)\lambda^i,$$

as desired. Now,

$$[\mu, \lambda^i\mu^j] = [\mu, \lambda^i]\mu^j = i\lambda^{i-1}\mu^j. \tag{13A.5}$$

It follows that any polynomial f satisfies $[\mu, f] = \frac{\partial f}{\partial \lambda}$, or

$$\mu f = \frac{\partial f}{\partial \lambda} + f\mu. \tag{13A.6}$$

Thus we obtain $\ell_\mu(f) = \frac{\partial f}{\partial \lambda} + f\mu$. For this reason, $\mathcal{A}_1(F)$ is often written as $F[\lambda, \frac{\partial}{\partial \lambda}]$. (Analogously, for any $f \in \mathcal{A}_1(F)$ we have $[\lambda, f] = -\frac{\partial f}{\partial u}$.)

Focusing on the terms of leading degree in multiplication, we see that these behave just as for the commutative polynomial algebra, and we can obtain the following information:

(i) $\mathcal{A}_1(F)$ is a PLID and principal right ideal domain (by Exercises A5 and A8);

(ii) $\mathcal{A}_1(F)$ is not a division ring since the only invertible elements are the constants from F. Moreover, $\mathcal{A}_1(F)$ is not left Artinian, and in fact has no minimal left ideals, by Exercise 13.13.

(iii) On the other hand, when $\mathrm{char}(F) = 0$, the algebra $\mathcal{A}_1(F)$ is easily seen to be simple, by the following argument:

Suppose $0 \neq A \lhd \mathcal{A}_1(F)$. Take an element $f = \sum_{i,j} \alpha_{ij}\mu^i\lambda^j \neq 0$ in A, having smallest possible degree m in λ. If $m > 0$, then

$$[\mu, f] = \frac{\partial f}{\partial \lambda}$$

has smaller degree in λ, contrary to our choice of f, so we must conclude that $m = 0$; i.e., $f \in F[\mu]$. But then $-f' = [\lambda, f] \in A$; iterating, we obtain a nonzero constant in A. Hence $1 \in A$.

The structure of $\mathcal{A}_1(F)$ in nonzero characteristic is given in Exercise A1.

Example 13A.2. More generally, we define the n-th **Weyl algebra** $\mathcal{A}_n(F)$ to have the F-vector space structure of the commutative polynomial algebra

$$F[\lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_n],$$

where multiplication satisfies

$$[\mu_i, \lambda_j] = \delta_{ij}, \tag{13A.7}$$

the Kronecker delta.

One can prove that $R = \mathcal{A}_n(F)$ is an algebra by induction on n, via Remark 13.50, taking $W = \mathcal{A}_{n-1}(F)$ and $M = R$, viewed naturally as a free W-module with base $\{\lambda_n^i\mu_n^j : i, j \in \mathbb{N}\}$. Details are left to the reader.

As in the case $n = 1$, $\mathcal{A}_n(F)$ is also described as

$$F\left[\lambda_1, \dots, \lambda_n, \frac{\partial}{\partial \lambda_1}, \dots, \frac{\partial}{\partial \lambda_n}\right].$$

The Weyl algebra has an interesting connection to the Jacobian conjecture (cf. Appendix 6A in Volume 1) via the following conjecture:

Dixmier's Conjecture D_n: Every algebra homomorphism from $\mathcal{A}_n(F)$ to itself is surjective.

It is not hard to show that the Dixmier conjecture D_n implies the Jacobian conjecture for $F[\lambda_1, \dots, \lambda_n]$; cf. Exercise A2. Tsuchimoto [**Tsu**] and, independently, Belov and Konsevich, proved that the Jacobian conjecture for $F[\lambda_1, \dots, \lambda_{2n}]$ implies the Dixmier conjecture D_n. Both proofs involve reduction modulo p; the latter proof uses the theory of Azumaya algebras, developed in Appendix 25B.

Example 13A.3 (Skew polynomial rings). Given a ring R with a homomorphism $\sigma\colon R \to R$, we define $R[\lambda; \sigma]$ to be the ring having the same additive structure as the polynomial ring $R[\lambda]$, but with multiplication given by

$$(r\lambda^i)(s\lambda^j) = r\sigma^i(s)\lambda^{i+j}.$$

This is a ring, seen either by direct verification of the axioms or by means of the regular representation on $R[\lambda]$. The latter verification involves some subtlety because of the skewed multiplication. Note that $M = R[\lambda]$ can be viewed as a right R-module under the action

$$\left(\sum r_i\lambda^i\right) r = \sum r_i\sigma^i(r)\lambda^i,$$

and is free with base $B = \{\lambda^i : i \in \mathbb{N}\}$, so we appeal to Remark 13.50, taking $W = R$ and viewing $R[\lambda;\sigma]$ as the F-subalgebra of $\operatorname{End} M_W$ generated by ℓ_λ and $\{\ell_r : r \in R\}$, all defined naturally by left multiplication.

$R[\lambda;\sigma]$ shares many of the ring-theoretic properties of $R[\lambda]$. For example, when R is a domain and σ is an injection, the usual degree function (in λ) satisfies

$$\deg(fg) = \deg f + \deg g$$

for all $0 \neq f, g \in R[\lambda;\sigma]$, implying that $R[\lambda;\sigma]$ is also a domain. (The proof, which mimics the standard proof in undergraduate algebra, is left to the reader.) This example arises in various contexts and is generalized in Exercise A4.

One important case is when σ is an automorphism of finite order n. Often, one is interested in $C = \operatorname{Cent}(R[\lambda;\sigma])$. Clearly σ acts as an automorphism on $\operatorname{Cent}(R)$, and C contains the fixed subring $\operatorname{Cent}(R)^{\langle\sigma\rangle}$. Furthermore, $\lambda^n \in C$, proving that

$$\operatorname{Cent}(R)^{\langle\sigma\rangle}[\lambda^n] \subseteq C.$$

To get equality, we need an extra assumption.

Let us say that an automorphism τ of a domain R is **inner** if there is an element $a \in R$ such that $\tau(r)a = ar$ for all $r \in R$. We claim that $\operatorname{Cent}(R)^{\langle\sigma\rangle}[\lambda^n] = C$ when σ^m is not inner for any $m < n$. Indeed, we have just proved ($\subseteq$). On the other hand, if $c = \sum c_i\lambda^i \in C$, then

$$0 = \lambda c - c\lambda = \sum \left(\sigma(c_i) - c_i\right)\lambda^{i+1},$$

implying that $\sigma(c_i) = c_i$ for each i, and likewise, for any $r \in R$,

$$0 = rc - cr = \sum \left(rc_i - c_i\sigma^i(r)\right)\lambda^i,$$

implying that $rc_i = c_i\sigma^i(r)$ for each i; by hypothesis, we conclude that each $c_i = 0$ unless $n \mid i$, yielding the claim.

This argument is generalized in Exercise A9.

Chapter 14

Semisimple Modules and Rings and the Wedderburn-Artin Theorem

In Theorem 13.48 we determined the structure of any ring R which is the direct sum of minimal left ideals. In this chapter, this result is used to prove the celebrated Wedderburn-Artin Theorem, which characterizes simple rings with a minimal left ideal as those of the form $M_n(D)$ (Theorem 14.24). A faster proof of this fact is given in Exercise 6. But the straightforward ideas from the text also lead to a much deeper understanding of the underlying module theory, and yield an even stronger theorem in Chapter 15. For later reference, we also introduce involutions on rings and algebras.

Semisimple modules

Since a minimal left ideal of R is a simple submodule, it is natural to start with the following definition.

Definition 14.1. The **socle** $\operatorname{soc}(M)$ of a module M is the sum of the simple submodules of M. The module M is **semisimple** if $\operatorname{soc}(M) = M$.

For a ring R, $\operatorname{soc}(R)$ is the socle of R viewed as a module over itself.

Remark 14.2. M may fail to have nonzero simple submodules, in which case we just define $\operatorname{soc}(M) = 0$. In fact, any integral domain that is not a field has socle 0; cf. Example 3.3 of Volume 1.

Example 14.3. The situation is straightforward when the underlying ring is a field F; then any module is a vector space V, which has a base B. For each b in B, Fb has dimension 1, and thus must be a simple module. Furthermore, $V = \bigoplus_{b\in B} Fb$; so, in particular, $\operatorname{soc}(V) = V$, i.e., V is semisimple. Our first goal is to see that much of this vector space theory carries over to semisimple modules in general.

Remark 14.4. (i) The sum of semisimple modules is semisimple.

(ii) Any homomorphic image of a simple module S is simple (isomorphic to S) or 0; hence for any map $f\colon M \to N$ of modules, $f(\operatorname{soc} M) \subseteq \operatorname{soc}(N)$. In particular, any homomorphic image of a semisimple module is semisimple.

Proposition 14.5.
(i) *If* $N \subseteq M$, *then* $\operatorname{soc}(N) \subseteq \operatorname{soc}(M)$.
(ii) $\operatorname{soc}(\operatorname{soc}(M)) = \operatorname{soc}(M)$.
(iii) *If* $M = \oplus M_i$ *are* R*-modules, then* $\operatorname{soc}(M) = \oplus \operatorname{soc}(M_i)$.

Proof. (i) By Remark 14.4(ii).

(ii) Every simple submodule of M is a simple submodule of $\operatorname{soc}(M)$.

(iii) $\operatorname{soc}(M_i) \leq \operatorname{soc}(M)$ by (i), so $\bigoplus \operatorname{soc}(M_i) \leq \operatorname{soc}(M)$. On the other hand, $\operatorname{soc}(M)_i \leq \operatorname{soc}(M_i)$ by Remark 14.4, so

$$\operatorname{soc}(M) = \oplus \operatorname{soc}(M)_i \leq \oplus \operatorname{soc}(M_i). \qquad \square$$

Thus, $\operatorname{soc}(M)$ is the unique largest semisimple submodule of M.

Complements of modules.

To proceed further, we take another analogy from vector spaces, which ties into direct sums.

Definition 14.6. A **complement** of a submodule $N \leq M$ is a submodule $N' \leq M$ such that $N + N' = M$ and $N \cap N' = 0$. Thus $M = N \oplus N'$ by Remark 2.7 of Volume 1. The module M is **complemented** if every submodule of M has a complement.

The complement (if it exists) need *not* be unique. Indeed, the x-axis is a one-dimensional subspace of $\mathbb{R}^{(2)}$; any other line passing through the origin is a complement.

Remark 14.7. For submodules N, K, and K' of M, if $K \leq N \leq K+K'$, then $N = K + (K' \cap N)$. (This is an instant application of Proposition 1.20 of Volume 1, taking $N_1 = N$ and $N_2 = K$.)

PROPOSITION 14.8. *Any submodule of a complemented module is complemented. More precisely, if $K \leq N \leq M$ and K has a complement K' in M, then K has the complement $K' \cap N$ in N.*

Proof. By Remark 14.7. □

Unfortunately, complements need not exist. Indeed, for $M = \mathbb{Z}$, every two nonzero submodules intersect nontrivially, so no submodule has a complement. Let us formalize this difficulty.

Definition 14.9. $N \leq M$ is a **large**, or **essential**, submodule of M if $N \cap K \neq 0$ for all $0 \neq K \leq M$.

Obviously, a large submodule cannot have a complement. Thus we need a weaker notion to bypass this difficulty. An **essential complement** for N (in M) is a submodule N' of M, for which $N \cap N' = 0$ and $N + N'$ is a large submodule of M.

PROPOSITION 14.10. *Every submodule N of M has an essential complement.*

Proof. By Zorn's lemma we can take $K \leq M$ maximal with respect to $N \cap K = 0$. We claim that $N + K$ is large in M. Indeed, suppose $T \leq M$ with $(N+K) \cap T = 0$. Then $N \cap (K+T) = 0$. (If $a = b + c \in N \cap (K+T)$, then $a - b = c \in (N+K) \cap T = 0$, implying $a = b \in N \cap K = 0$.) Hence, by our maximality assumption, we see $K + T = K$; i.e., $T \subseteq K$, so $T = 0$. □

Now we connect back to simple modules.

Remark 14.11. Suppose $S \neq 0$ is a simple submodule of M and $N \leq M$. Then $S \cap N \neq 0$ iff $S \subseteq N$. (Indeed $S \cap N \leq S$, and thus must be S by simplicity.)

LEMMA 14.12. *$\mathrm{soc}(M)$ is contained in every large submodule of M.*

Proof. It suffices to show that every large submodule N of M contains every simple submodule S of M, which is true by Remark 14.11. □

We are ready to show the equivalence of three very important module-theoretic properties.

THEOREM 14.13. *The following conditions are equivalent for a module M:*

(i) *M is semisimple.*
(ii) *M is complemented.*
(iii) *M has no proper large submodules.*

Proof. (i) ⇒ (iii) By Lemma 14.12, any large submodule of M contains $\mathrm{soc}(M) = M$.

(iii) ⇒ (ii) Any essential complement (which exists by Proposition 14.10) must be a complement, by (iii).

(ii) ⇒ (i) This is the tricky part. Suppose $\mathrm{soc}(M) \neq M$. By (ii), $\mathrm{soc}(M)$ has a complement $N \neq 0$. We want to find a simple submodule of N, since this would be a simple submodule of M not in $\mathrm{soc}(M)$, a contradiction. Take $0 \neq a \in N$, and take $K \leq N$ maximal with respect to $a \notin K$. By Proposition 14.8, K has a complement S in N.

We claim that S is the desired simple submodule of N. Otherwise, S has a proper nonzero submodule P. P has a complement $P' \neq 0$ in S, so $S = P \oplus P'$. Since $S \cap K = 0$, our assumption on K implies that both $a \in K + P$ and $a \in K + P'$. Write $a = k_1 + p = k_2 + p'$ for suitable $k_i \in K$, $p \in P$, and $p' \in P'$. Then

$$k_1 - k_2 = p' - p \in K \cap S = 0,$$

so $p' = p \in P \cap P' = 0$, implying $a = k_1 \in K$, a contradiction. □

(See Exercise 8 for a stronger version of this theorem.) In particular, a module is semisimple iff it is complemented; thus, every submodule of a semisimple module is semisimple, by Proposition 14.8.

Now let us bring in another extremely important concept, that of finite (composition) length, introduced in Chapter 3 of Volume 1.

Example 14.14. (i) Any infinite-dimensional vector space over a field is semisimple but has infinite length.

(ii) For any $n > 0$, the $\mathbb{Z}$-module $\mathbb{Z}/n$ is a finite set, and thus has finite length, but is not semisimple unless n is a product of distinct primes; cf. Exercise 1.

We need a way of computing the complement within a chain.

LEMMA 14.15. *If $N_1 \subset N_2$ in a semisimple module M, then any given complement N_1' of N_1 (in M) contains a complement of N_2 .*

Proof. Let N_2' be a complement of $N_2 \cap N_1'$ in N_1'. Then

$$N_2 + N_2' \supseteq (N_1 + (N_2 \cap N_1')) + N_2' = N_1 + N_1' = M$$

and

$$N_2 \cap N_2' = N_2 \cap (N_1' \cap N_2') = (N_2 \cap N_1') \cap N_2' = 0.$$

Hence N_2' is a complement of N_2 in M. □

We are ready to bring back direct sums.

THEOREM 14.16. *Suppose M is a semisimple module. The following conditions are equivalent:*

(1) *M has finite composition length.*
(2) *M is Artinian.*
(3) *M is Noetherian.*
(4) *M is a finite direct sum of simple submodules.*

Proof. (4) ⇒ (1) ⇒ (2) By definition.

(2) ⇒ (3) On the contrary, given $N_1 \subset N_2 \subset \cdots$, we inductively take N_i' to be a complement of N_i contained in N_{i-1}', obtained via the lemma. Then $N_1' \supset N_2' \supset \cdots$, contradicting (2).

(3) ⇒ (4) Take a submodule N of M maximal with respect to being a finite direct sum of simple submodules; if $N < M$, then its complement N' contains a simple submodule $S \neq 0$, and $N \oplus S \supset N$, a contradiction. Thus $N = M$. □

More generally, a module is semisimple iff it is the direct sum of simple modules, but the proof requires a bit more care; cf. Exercise 4 (or 5).

Semisimple rings

We are ready to introduce perhaps the most important class of noncommutative rings.

Definition 14.17. A ring R is a **semisimple ring** if R is semisimple as an R-module.

We have a serious linguistic difficulty here: A simple ring need not be semisimple! Indeed, in Example 13A.1 we saw that the Weyl algebra $\mathcal{A}_1(F)$ is simple but has no minimal left ideals. However, it is too late to correct the terminology. (Jacobson advocated a more general definition of semisimple ring, but it did not catch on.)

PROPOSITION 14.18. *A ring R is semisimple iff it is a finite direct sum of minimal left ideals.*

Proof. (⇒) By definition, 1 is contained in a sum of minimal left ideals, so we have minimal left ideals that we name $L_1, \ldots, L_t$, for which $1 = \sum_{i=1}^t a_i$ for suitable a_i in L_i. But then for any r in R we have

$$r = \sum_{i=1}^t r a_i \in L_1 + \cdots + L_t.$$

Hence $\ell(R) \leq t$. We conclude by Theorem 14.16.

(⇐) Obvious, by the definition. □

THEOREM 14.19. *A ring R is semisimple iff $R \cong \prod_{i=1}^t M_{n_i}(D_i)$ for suitable t, suitable $n_i \in \mathbb{N}$, and suitable division rings D_i.*

Proof. By Theorem 13.48 and Proposition 14.18. □

Theorem 14.19 and its consequences raise semisimple rings and modules to a central position in ring theory, and show that the ideal structure of semisimple rings is particularly nice. (The application of Theorem 13.47 via Theorem 13.48 was so successful that one may ask what the structure of $\mathrm{End}_R M$ is for an arbitrary semisimple module M, but the result is not as decisive.)

Remark 14.20. Given a semisimple ring

$$R = R_1 \times \cdots \times R_t,$$

where each $R_i = M_{n_i}(D_i)$, the left ideals of R are precisely of the form $L = L_1 \times \cdots \times L_t$, where each $L_i < R_i$, by Proposition 13.31(iii). Thus the maximal left ideals have the form

$$R_1 \times \cdots \times R_{i-1} \times L_i \times R_{i+1} \times \cdots \times R_t, \tag{14.1}$$

where L_i is a maximal ideal of R_i.

Likewise, each ideal A of R is of the form $A_1 \times \cdots \times A_t$, where each $A_i = 0$ or $A_i = R_i$ (since R_i is simple). In particular, R/A is isomorphic to a direct product of some of the R_i, and thus also is a semisimple ring.

The maximal ideals of R thus have the form

$$P_i = R_1 \times \cdots \times R_{i-1} \times 0 \times R_{i+1} \times \cdots \times R_t, \tag{14.2}$$

so R has precisely t maximal ideals $P_1, \ldots, P_t$.

Digression 14.21: Idempotents in semisimple rings.

Here is another way of viewing left ideals in a semisimple ring. We do not need it here, but it plays an important role later in representation theory, to wit, in computing representations of the symmetric group S_n.

(i) N has a complement iff there is a projection π of M with $\pi(M) = N$. Since projections are just idempotents of $\mathrm{End}_R M$, we see that when M is complemented, $\mathrm{End}_R M$ must have a wealth of idempotents.

(ii) Let us apply (i) to the case $M = R$; we identify $\mathrm{End}_R R$ with R^{op} via Example 13.41(ii). Noting that R^{op} and R have the same idempotents, we can identify any projection $R \to R$ with an idempotent. We conclude that any left ideal L of a semisimple ring R can be written in the form Re for a suitable idempotent e. Thus $ae = a$ for all $a \in L$.

(iii) Continuing (ii), if $L = L_1 \oplus L_2$, then $L_i = Re_i$ for orthogonal idempotents e_i such that $e_1 + e_2 = e$. (This follows from (i), but could be seen directly by writing $e = e_1 + e_2$ for $e_i \in L_i$ and noting that $e_1 = e_1 e = e_1^2 + e_1 e_2$, implying $e_1^2 = e_1$ and $e_1 e_2 = 0$; likewise $e_2^2 = e_2$ and $e_2 e_1 = 0$.)

It follows that a left ideal $L = Re$ of R is minimal iff the idempotent e cannot be written as a sum of two nonzero orthogonal idempotents. (Indeed, any smaller left ideal would have a complement in L.)

(iv) We also have a converse to Schur's Lemma, for any ring R and nonzero idempotent e: eRe is a division ring iff Re is a minimal left ideal. (The proof of ($\Rightarrow$) could be seen as a consequence of Proposition 13.43, but here is a direct proof using Proposition 13.2(iv): For any $ae, be \neq 0$ in Re, we have $be = b(eae^{-1})eae \in Rae$.)

Modules over semisimple rings.

The next remark provides two important structural results.

Remark 14.22. If M is an R-module and L is a minimal left ideal of R, then, for any a in M, either $La = 0$ or La is a simple submodule of M. (Indeed, apply Remark 14.2 to the map $L \to La$ given by right multiplication by a.)

Proposition 14.23. *Any module M over a semisimple ring R is semisimple.*

Proof. Writing $R = \sum L$ as a sum of minimal left ideals, we have

$$M = \sum_{a \in M} Ra = \sum_{L} \sum_{a \in M} La,$$

a sum of simple modules by Remark 14.22. □

The Wedderburn-Artin Theorem

We bring in left Artinian rings, which clearly have minimal left ideals, as seen by applying the minimum condition to the class of nonzero left ideals. As an important special case of Theorem 14.19, we have

Theorem 14.24 (Wedderburn-Artin). *The following properties are equivalent for a ring R:*

(1) $R \cong M_n(D)$.
(2) *R is simple and left Artinian.*
(3) *R is simple with a minimal left ideal.*
(4) *R is simple and semisimple.*

Proof. (1) $\Rightarrow$ (2) By Corollary 13.18.

(2) $\Rightarrow$ (3) Clear.

(3) $\Rightarrow$ (4) $0 \neq \mathrm{soc}(R) \lhd R$, so $\mathrm{soc}(R) = R$.

(4) $\Rightarrow$ (1) By Theorem 14.19 we know that $R \cong \prod_{i=1}^{t} M_{n_i}(D_i)$ for suitable t. But if $t > 1$, then R cannot be simple, so $t = 1$. □

We strengthen this even further in Theorem 15.19; also, the underlying division ring D is unique, to be seen in Corollary 15.9.

Preliminaries about division algebras.

Having proved the Wedderburn-Artin Theorem, we become more interested in the structure of division rings. Although we defer a systematic study until Chapter 24, some basic observations are in order here.

Remark 14.25. (i) Suppose D is a division ring with center F. F is a field (cf. Remark 13.24′), and we often view D as an algebra over F; to emphasize this perspective, we call D a **division algebra**. For any d in D, the subalgebra $F[d]$ is an integral domain.

(ii) Suppose R is a domain algebraic over a field F. (In particular, this holds if R is finite-dimensional over F.) Then, for any $0 \neq a \in R$, the subalgebra $F[a]$ is an integral domain which, by Remark 4.7′ of Volume 1, is a field and thus contains a^{-1}. It follows that R is a division ring.

In one important case, the theory of noncommutative division algebras is nonexistent. We write "f.d." for "finite-dimensional." As in Volume 1, we also write $[D:F]$ for the dimension $\dim_F D$.

Proposition 14.26. *The only f.d. division algebra D over an algebraically closed F is F itself.*

Proof. For each $d \in D$, the integral domain $F[d]$ is f.d. over F and thus is a finite field extension of F, so $F[d] = F$. This shows that $d \in F$ for each $d \in D$, i.e., $D = F$. □

Combining this with Theorem 14.19, we have

THEOREM 14.27. *If R is a f.d. semisimple algebra over an algebraically closed field F, then R is isomorphic to a direct product of matrix algebras, i.e.,*

$$R \cong M_{n_1}(F) \times \cdots \times M_{n_t}(F).$$

Proof. In Theorem 14.19, each $D_j = F$ by Proposition 14.26. □

There is also an appropriate formulation of Schur's Lemma in this case:

PROPOSITION 14.28 (SCHUR'S LEMMA). *Suppose, for F an algebraically closed field, $R \subseteq M_n(F)$ such that $M = F^{(n)}$, viewed naturally as an R-module, is simple. If $f \in \mathrm{End}_R M$, then f is given by scalar multiplication.*

Proof. $\mathrm{End}_R M$ is a f.d. division algebra over F, and thus is isomorphic to F. □

Example 14.29 (A noncommutative f.d. division algebra over $\mathbb{R}$). We define Hamilton's algebra of **quaternions** $\mathbb{H} = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}j \oplus \mathbb{R}k$ where $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $ki = -ik = j$, and $jk = -kj = i$. These rules can be written more concisely as

$$i^2 = j^2 = k^2 = ijk = -1,$$

and can be extended via distributivity to all of $\mathbb{H}$; a computation by means of Exercise 13.1 shows that $\mathbb{H}$ is an $\mathbb{R}$-algebra, and in fact is a division algebra since

$$(\alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k)^{-1} = \frac{1}{\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2}(\alpha_1 - \alpha_2 i - \alpha_3 j - \alpha_4 k),$$

as seen by direct computation.

Let us demonstrate these basic properties of $\mathbb{H}$ by means of the regular representation. $[\mathbb{H} : \mathbb{R}] = 4$, so we could display $\mathbb{H} \subset M_4(\mathbb{R})$. However, we can make the computations easier by taking $C = \mathbb{R} + \mathbb{R}i \subset \mathbb{H}$. C is a field, isomorphic to $\mathbb{C}$, so $\mathbb{H} = C + jC$ is a right vector space over C with base $\{1, j\}$. We inject $\mathbb{H}$ into $\mathrm{End}_C \mathbb{H} \cong M_2(\mathbb{C})$ by determining the

matrices $\hat{1}, \hat{\imath}, \hat{\jmath}, \hat{k}$ corresponding respectively to left multiplication by $1, i, j, k$ with respect to the base $\{1, j\}$. Thus

$$\hat{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad \hat{\imath} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & \bar{i} \end{pmatrix};$$

$$\hat{\jmath} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad \hat{\jmath}\hat{\imath} = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} = \begin{pmatrix} 0 & \bar{i} \\ -i & 0 \end{pmatrix},$$

where $\overline{\ }$ denotes complex conjugation. One easily sees that these matrices indeed satisfy the defining equation of $\mathbb{H}$, and $z_1 + jz_2$ (for $z_1, z_2 \in C$) corresponds to the matrix

$$\begin{pmatrix} z_1 & \overline{z_2} \\ -z_2 & \overline{z_1} \end{pmatrix}. \tag{14.3}$$

Thus $\mathbb{H}$ is indeed a ring, isomorphic to the subring of $M_2(\mathbb{C})$ consisting of matrices of the form (14.3). But we know that the determinant of the matrix $A = \left(\begin{smallmatrix} z_1 & \overline{z_2} \\ -z_2 & \overline{z_1} \end{smallmatrix}\right)$ is $|z_1|^2 + |z_2|^2$, which is positive unless $z_1 = z_2 = 0$, i.e., unless $A = 0$. Thus (by the adjoint formula, for example) for $A \neq 0$,

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} \overline{z_1} & -\overline{z_2} \\ z_2 & z_1 \end{pmatrix},$$

which also has form (14.3). Hence every nonzero element of $\mathbb{H}$ is invertible in $\mathbb{H}$, i.e., $\mathbb{H}$ is a division algebra.

According to the Hollywood version, the division algebra $\mathbb{H}$ was discovered in 1843 by William Rowan Hamilton during a stroll along Brougham Bridge in Ireland, after a fruitless 10-year attempt to construct a division algebra of dimension 3 over $\mathbb{R}$; Hamilton thereupon carved his equations into the stone. Although quaternions were anticipated by work of Gauss and Rodriguez, Hamilton was the one to view them as an algebraic entity, thereby paving the way to the theory of noncommutative algebra.

In fact, it is fairly easy to see why there are no noncommutative division algebras of dimension 3.

Remark 14.30. If D is any f.d. division algebra with center F, then, for any d in D, $F[d]$ is a subfield, by Remark 14.25(ii). Thus, Theorem 4.4 of Volume 1 says that

$$[D : F] = [D : F[d]][F[d] : F]. \tag{14.4}$$

In particular, if D is noncommutative, then both factors on the right side of Equation (14.4) are ≥ 2, implying $[D : F] \geq 4$. (This fact is strengthened in Corollary 24.26 and Theorem 24.32.)

Remark 14.31. Frobenius showed that any noncommutative f.d. division algebra D over $\mathbb{R}$ is isomorphic to $\mathbb{H}$. Indeed, by Remark 14.30, for each $d \in D$, $\mathbb{R}[d]$ is a field extension of $\mathbb{R}$ and thus is isomorphic to $\mathbb{C}$, and in particular contains an element whose square is -1. Playing around with these elements enables one to construct a copy of $\mathbb{H}$ inside D, and prove that it is all of D. Details are given in Exercise 24.1, or the reader can wait until Example 24.45, at which stage the result becomes almost trivial.

Supplement: Rings with involution

We can refine our investigation of semisimple rings by introducing one more piece of structure that is pervasive in examples of rings and algebras.

Definition 14.32. An **involution** of a ring R is an anti-automorphism of order 2, i.e., a map $*\colon R \to R$ satisfying

$$\begin{aligned}(r_1 + r_2)^* &= r_1^* + r_2^*;\\ (r_1 r_2)^* &= r_2^* r_1^*;\\ (r^*)^* &= r\end{aligned}$$

for all r_i in R.

An element r of a ring with involution $(R, *)$ is called **symmetric** (resp. **skew-symmetric** or **anti-symmetric**), if $r^* = r$ (resp. $r^* = -r$).

Remark 14.32′. Let S denote the symmetric elements under a given involution $(*)$, and let K denote the skew-symmetric elements. If $a, b \in S$, then $ab + ba \in S$ since $(ab + ba)^* = b^*a^* + a^*b^* = ba + ab = ab + ba$, and likewise $aba \in S$. If $a, b \in K$, then $ab - ba \in K$ and $aba \in K$. These operations on S and K are important objects of study, carrying their own structures that are studied in Chapter 21.

Example 14.33. (i) Complex conjugation in $\mathbb{C}$ is an involution.

(ii) $M_n(F)$ has the **transpose involution** given by the $A \mapsto A^t$.

(iii) Suppose R is any ring with involution $(*)$, and $a \in R$ is invertible with $a^* = \pm a$. Then one can define a new involution by $r \mapsto ar^*a^{-1}$, as seen by easy verification. For example, taking $n = 2m$ even and $R = M_{2m}(F)$, with $(*)$ the transpose, and $a = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$, where I denotes the $m \times m$ identity matrix, we get the **canonical symplectic involution**, given by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \mapsto \begin{pmatrix} D^t & -B^t \\ -C^t & A^t \end{pmatrix}.$$

In particular, for $n = 2$, the canonical symplectic involution is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

This involution is quite different in structure from (ii); for $n = 2$, the only symmetric matrices are scalars.

(iv) $M_n(\mathbb{C})$ has the **Hermitian transpose involution** $(c_{ij})^* = (\overline{c_{ji}})$.

(v) $\mathbb{H}$ (cf. Example 14.29) has the **standard involution**, given by

$$(\alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k)^* = \alpha_1 - \alpha_2 i - \alpha_3 j - \alpha_4 k.$$

This involution is closely tied in with the structure of $\mathbb{H}$, since

$$aa^* = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 \in \mathbb{R}.$$

These examples can be unified through bilinear forms on vector spaces; cf. Exercises 11 and 12. One can develop an interesting structure theory for rings with involution in the framework of universal algebra: The objects are $(R, *)$ where R is a ring with involution $(*)$, and the homomorphisms $(R, *) \to (T, J)$ are ring homomorphisms $f\colon R \to T$ that "preserve" the involution, i.e., $f(r^*) = f(r)^J$. One could go through much of the structure theory; we leave this for Exercises 17ff. On the other hand, Exercise 18 shows that the theory of rings with involution actually "contains" the theory of rings.

Chapter 15

The Jacobson Program Applied to Left Artinian Rings

In this chapter, we introduce the structure theory of rings, in order to yield decisive results about Artinian rings that strengthen parts of Theorem 14.24. In particular, we show that any Artinian ring R has a nilpotent ideal N such that R/N is semisimple Artinian. The way N and R/N fit together can be quite complicated, even for finite-dimensional algebras over fields, and only is considered later in Appendix 25C.

Let us start by indicating the difficulty in building a general ideal-theoretic structure theory for rings. Presumably one should start with simple rings, i.e., without nonzero ideals, so let us consider the basic question of how to generate ideals in a ring. For subsets L_1, L_2 of a ring R, we recall the definition of the product $L_1L_2 = \{\sum_{\text{finite}} a_ib_i : a_i \in L_1, b_i \in L_2\}$. By associativity, $(L_1L_2)L_3 = L_1(L_2L_3)$, which we denote as $L_1L_2L_3$, and so forth. In particular, if $a \in R$, then RaR denotes $\{\sum_{\text{finite}} r_ias_i : r_i, s_i \in R\}$, and this is the smallest ideal of R containing a. (Clearly $a = 1a1 \in RaR$.) Thus, R is a simple ring iff $RaR = R$ for all $0 \neq a \in R$.

Unfortunately, whereas $Ra = \{ra : r \in R\}$, there is no elementary way to describe RaR without introducing infinite sums. Consequently, it is difficult to obtain information about simple rings in general. For example, the Weyl algebra of Example 13A.1 is an example of a simple Noetherian ring that is not Artinian.

Jacobson's first great contribution to the structure theory lay in utilizing modules to introduce a different elementary definition that encompasses simple rings. In this chapter, we use Jacobson's theory to describe Artinian rings. In Appendix 15A, we continue with some of the main parts of Jacobson's general structure theory, together with results of Amitsur that tie it to other notions that we encounter here.

Primitive rings and ideals

Definition 15.1. An ideal P is **primitive** if $P = \operatorname{Ann}_R M$ for some simple module M. A ring R is **primitive** if 0 is a primitive ideal, i.e., if R has a faithful simple module.

Since this definition involves simple modules rather than simple rings, it is easier to formulate in terms of elements, and primitivity becomes the foundation stone of Jacobson's theory.

Example 15.1′. Every simple ring R is primitive. (By Proposition 13.1, R has nonzero simple modules, which are faithful by Remark 13.36(ii).)

A commutative ring is primitive iff it is a field. (This can be seen directly, or as an immediate consequence of Exercise 2.) However, there are examples of noncommutative primitive rings that are not simple; cf. Exercises 8, 10, 11, and 12. Simple non-Artinian rings (such as the Weyl algebra $\mathcal{A}_1(F)$) must have socle 0, in view of Theorem 14.24.

Remark 15.2. An ideal P is primitive iff the ring R/P is primitive. Indeed, each direction follows easily from Remark 1.21 of Volume 1. If $P = \operatorname{Ann}_R M$ for a simple R-module M, we can view M naturally as a simple R/P-module, and then

$$\operatorname{Ann}_{R/P} M = (\operatorname{Ann}_R M)/P = P/P = 0,$$

proving that R/P is primitive.

On the other hand, if M is a faithful simple R/P-module, then, viewing M as a simple R-module, we see that $P = \operatorname{Ann}_R M$.

Here is a cute way of generating primitive ideals, which strengthens Nakayama's Lemma (of Volume 1).

Remark 15.3. Suppose M is any nonzero f.g. R-module. Writing $M = \sum_{i=1}^{t} Ra_i$ with t minimal, take a submodule $N \supseteq \sum_{i=1}^{t-1} Ra_i$ maximal with respect to $a_t \notin N$. (This is possible, by Zorn's Lemma.) Clearly N is a maximal submodule of N; hence, M/N is a simple module, and its annihilator P is a primitive ideal satisfying $PM \subseteq N \neq M$.

One amazing feature of primitive ideals is their connection to maximal left ideals. Indeed, $\operatorname{Ann}_R M = \bigcap_{a\in M} \operatorname{Ann}_R a$. But, by Remark 13.0, $\operatorname{Ann}_R a$ is a maximal left ideal of R when the module M is simple, so we have:

PROPOSITION 15.4. *Every primitive ideal is the intersection of maximal left ideals.*

The same idea yields

Example 15.5. Every maximal left ideal L of a ring R contains a primitive ideal. Indeed, R/L is a simple R-module, so $\operatorname{Ann}_R R/L$ is a primitive ideal contained in $\operatorname{Ann}_R(1+L) = L$.

Our main interest here lies in the following example.

Example 15.6. Suppose a ring $R = R_1 \times \cdots \times R_t$ is semisimple, where each R_i is simple Artinian. Any primitive ideal of R has the form $\operatorname{Ann} R/L$ for a suitable maximal left ideal L. Since, by Remark 14.20, the maximal left ideal L has the form

$$R_1 \times \cdots \times R_{i-1} \times L_i \times R_{i+1} \times \cdots \times R_t,$$

we see that $R/L \cong 0 \times \cdots \times 0 \times R_i/L_i \times 0 \times \cdots \times 0$. But $\operatorname{Ann}_{R_i}(R_i/L_i) = 0$ since R_i is simple, so

$$\operatorname{Ann}_R(R/L) = R_1 \times \cdots \times R_{i-1} \times 0 \times R_{i+1} \times \cdots \times R_t = P_i$$

as defined in (14.2), which describes the maximal ideals. Thus, every primitive ideal of R is one of $P_1, \ldots, P_t$, and is thus a maximal ideal.

In particular, if the ring R is primitive semisimple, then some $P_i = 0$, i.e., $R = R_i$ is simple Artinian.

One objective of this chapter is to strengthen this result.

Isomorphism classes of simple modules.

A very fundamental question in module theory is to classify the isomorphism classes of simple modules of a given ring. The following fact gives us considerable mileage.

PROPOSITION 15.7. *Suppose a primitive ring R has a minimal nonzero left ideal L. Then every faithful simple R-module M is isomorphic to L.*

Proof. $LM \neq 0$, so for some a in M we have $0 \neq La \leq M$, implying that $La = M$. Hence the map $f\colon L \to M$ given by $r \mapsto ra$ is an isomorphism, by Schur's Lemma. □

COROLLARY 15.8. *If R is a simple ring with nonzero socle, then any two simple R-modules are isomorphic.*

Proof. Any nonzero R-module is faithful, by Remark 13.36′(ii). □

COROLLARY 15.9. *The Wedderburn-Artin decomposition $R = M_n(D)$ of a simple Artinian ring R is unique, in the sense that for $R' = M_u(D')$, we have $R \cong R'$ iff $D' \cong D$ and $n = u$.*

Proof. ($\Leftarrow$) Any isomorphism $D \to D'$ can be extended to an isomorphism $M_n(D) \to M_n(D')$ by fixing the matrix units.

($\Rightarrow$) By Proposition 13.43, $D \cong \operatorname{End}_R(L)^{\mathrm{op}}$, where $L = Re_{11}$ is a minimal left ideal and thus is unique up to isomorphism as an R-module. Hence D is determined up to isomorphism. As we saw in Corollary 13.18, n is the composition length of R as an R-module, which is unique by the Jordan-Hölder theorem. □

Remark 15.9′. (i) If R is an algebra over a field F, then the isomorphisms of Corollary 15.9 are isomorphisms over F, via the same proofs.

(ii) Suppose $R = M_n(D)$ and e is an idempotent of R. Then, as a ring with multiplicative unit e, $eRe \cong M_u(D)$ for suitable u. (Indeed, letting L be a minimal left ideal of R, one has $Re = L^u$ for some u, so, in view of Proposition 13.43 and Theorem 13.47,

$$eRe \cong (\operatorname{End}_R Re)^{op} \cong M_u(\operatorname{End}_R L)^{op} \cong M_u(D).)$$

Here is an application of the utmost importance in the representation theory of finite groups (Chapter 19).

COROLLARY 15.10. *Suppose $R = R_1 \times R_2 \times \cdots \times R_t$ is semisimple, with each R_i simple. Then R has precisely t nonisomorphic simple modules, and each one is isomorphic to a minimal left ideal L_i of R_i, for suitable $1 \leq i \leq t$.*

Proof. By Proposition 13.31, the simple R-modules correspond to the simple R_i-modules, and by Corollary 15.8 each of these is isomorphic to L_i. On the other hand, for $i \neq j$, L_i is not isomorphic to L_j as R-modules, by Proposition 13.34. □

Prime and semiprime rings: first glance.

It is convenient to introduce here a notion that, although of importance in the general theory, quickly becomes superfluous for Artinian rings.

Definition 15.11. A ring R is **prime** if the product of any two nonzero ideals is nonzero.

Thus a commutative prime ring is just an integral domain. Prime rings often are the "correct" generalization of integral domains in the noncommutative theory.

Remark 15.12. (i) In a prime ring R, the product $A_1 \cdots A_m$ of any finite number of nonzero ideals is nonzero. (This is true by definition for $m = 2$; in general, by induction, $0 \neq A_1 \cdots A_{m-1} \lhd R$, so we see that $A_1 \cdots A_m = (A_1 \cdots A_{m-1})A_m \neq 0$.)

(ii) If L is a left ideal and I is a right ideal of R, then $LI \lhd R$. In particular, $LR \lhd R$. It follows at once for R prime that $\operatorname{Ann}_R L = 0$ for any $L < R$, since $(\operatorname{Ann}_R L)(LR) = 0$.

$\mathbb{Z}$ is an example of a prime ring that is not primitive (since it is not a field). A noncommutative example is $M_2(\mathbb{Z})$. Nevertheless, we do have a general connection between primitive and prime.

Remark 15.13. (i) Any primitive ring R is prime. (Indeed, take a faithful simple module M, and suppose $A, B \lhd R$ with $AB = 0$ but $B \neq 0$. Then $0 \neq BM \leq M$, implying that $BM = M$. But then

$$0 = (AB)M = A(BM) = AM,$$

implying that $A = 0$.)

(ii) Conversely, if a prime ring R has a minimal nonzero left ideal L, then $\operatorname{Ann}_R L = 0$; thus, L is a faithful simple module, so R is primitive.

Given $A \subseteq R$, define $A^2 = AA$, and inductively $A^k = A^{k-1}A$. We say that $A \subseteq R$ is **nilpotent** if $A^k = 0$ for some $k \geq 1$. For example, the ideal $A = 2\mathbb{Z}/16$ of the ring $\mathbb{Z}/16$ is nilpotent, since $A^4 = 0$. Also, the ring $\left(\begin{smallmatrix} F & F \\ 0 & F \end{smallmatrix}\right)$ of upper triangular matrices has the nilpotent ideal $\left(\begin{smallmatrix} 0 & F \\ 0 & 0 \end{smallmatrix}\right)$.

Definition 15.14. A ring R is **semiprime** if $A^2 \neq 0$ for any $0 \neq A \lhd R$.

A commutative ring is semiprime iff it is reduced.

Remark 15.15. A semiprime ring R has no nonzero nilpotent ideals. (We need show that $A^k \neq 0$ for any $0 \neq A \lhd R$. This is true by definition for $k = 2$; in general for $k > 2$, by induction, $A^{k-1} \neq 0$, so

$$0 \neq (A^{k-1})^2 = A^{2k-2} \subseteq A^k,$$

since $k \leq 2k - 2$.)

Semiprime rings with involution are considered in Exercises 15–18.

The Jacobson radical

Definition 15.16. The **Jacobson radical** $\operatorname{Jac}(R)$ is the intersection of the primitive ideals of R.

This is not to be confused with the notion of a radical ideal in commutative algebra (given in Definition 9.16 of Volume 1).

PROPOSITION 15.17. *Let $J = \operatorname{Jac}(R)$. Then*
(i) *J is also the intersection of the maximal left ideals of R.*
(ii) *If $a \in J$, then $1 - a$ is left invertible.*

Proof. (i) By Proposition 15.4 and Example 15.5.

(ii) Otherwise $R(1-a)$ is contained in a maximal left ideal L. But $a \in L$ by (i), so $1 \in L$, which is impossible. □

Other basic properties of $\operatorname{Jac}(R)$ are given in Exercises 24ff.

Digression: Jacobson's program. We call a ring R **semiprimitive** if $\operatorname{Jac}(R) = 0$; then R is a subdirect product of primitive rings (cf. Digression 13.30). The Jacobson program in the structure of rings studies R in terms of $\operatorname{Jac}(R)$ and the semiprimitive ring $R/\operatorname{Jac}(R)$. This program is particularly useful for those rings whose primitive homomorphic images are simple. In this chapter we see how it works in the structure of Artinian rings, and then in Appendix 15A we present the bulwark of Jacobson's program, the Density Theorem, together with some key structure theorems of Amitsur.

The structure of left Artinian rings

We apply these results to develop the structure theory for left Artinian rings, which we recall are rings that satisfy the descending chain condition (DCC) on left ideals. Since any left Artinian ring must have minimal nonzero left ideals, Remark 15.13 implies that "prime" and "primitive" coincide in the class of left Artinian rings.

THEOREM 15.18. *Suppose R is left Artinian, and let $J = \operatorname{Jac}(R)$. Then*
(i) *J is the intersection of finitely many maximal left ideals of R.*
(ii) *R/J is a semisimple ring.*
(iii) *There are only finitely many primitive ideals, and each primitive ideal of R is maximal.*
(iv) *J is a nilpotent ideal.*

Proof. (i) Let $\mathcal{S}$ be the set of finite intersections of maximal left ideals of R. Since R is left Artinian, $\mathcal{S}$ has some minimal member A. By definition, for

each maximal left ideal L, we have $A \cap L \in \mathcal{S}$ so $A \cap L = A$, implying that $A \subseteq L$; hence

$$J \supseteq A = L_1 \cap \cdots \cap L_k \supseteq J,$$

for suitable maximal left ideals L_i of R, implying that $J = L_1 \cap \cdots \cap L_k$.

(ii) Continuing the notation of (i), let

$$M = \bigoplus_{i=1}^{k} R/L_i,$$

a semisimple module. The natural map $R \to M$ given by $r \mapsto (r + L_i)$ has kernel $\bigcap_{i=1}^{k} L_i = J$, so Noether I (Theorem 0.1 of Volume 1) implies $R/J \hookrightarrow M$; hence R/J is semisimple as an R-module (cf. Proposition 14.8 and Theorem 14.13), and thus as an R/J-module.

(iii) Any primitive ideal P contains $J = \mathrm{Jac}(R)$. Thus we could pass to R/J, which is semisimple, and we are done by Example 15.6.

(iv) This time take $\mathcal{S} = \{$finite products of primitive ideals, permitting duplication$\}$, and take some minimal A in $\mathcal{S}$. By construction, $PA = A$ for every primitive ideal P, implying that $A = 0$ by Remark 15.3; thus $P_1 \cdots P_k = 0$ for suitable primitive ideals P_i, implying that $J^k = 0$. □

One immediate consequence of Theorem 15.18(iii) is the following improvement of the Wedderburn-Artin Theorem:

THEOREM 15.19. *Any prime left Artinian ring is simple (and thus satisfies the conditions in Theorem 14.24).*

Let us summarize our results characterizing semisimple rings:

THEOREM 15.20. *The following are equivalent for a ring R:*

(1) *R is semisimple.*
(2) *R is a finite direct sum of simple right R-modules.*
(3) *R is left Artinian and semiprime.*
(4) *R is right Artinian and semiprime.*
(5) *R is a finite direct product of matrix rings over division rings.*

Proof. (1) $\Leftrightarrow$ (5) by Theorem 14.19.

(5) $\Rightarrow$ (3) By Remark 14.20, any nonzero ideal A of $R = R_1 \times \cdots \times R_t$ has the form $A_1 \times \cdots \times A_t$, where some $A_i = R_i$. But then $A_i^2 = R_i$ since $1 \in R_i$, implying that $A^2 \neq 0$.

(3) $\Rightarrow$ (1) By Theorem 15.18(iv), J is nilpotent and thus 0, so $R = R/J$ is semisimple, by Theorem 15.18(ii).

Likewise (2), (4), and (5) are equivalent by passing to R^{op}. □

Here is one more striking application.

THEOREM 15.21 (HOPKINS-LEVITZKI). *Any left Artinian ring R is also left Noetherian.*

Proof. We show that R has finite composition length. Let $J = \mathrm{Jac}(R)$. By Theorem 15.18, $J^t = 0$ for some t. Consider the chain of modules

$$R \supseteq J \supseteq J^2 \supseteq \cdots \supseteq J^t = 0.$$

Each factor module $M_i = J^{i-1}/J^i$ is Artinian, and is naturally a module over the semisimple ring R/J (cf. Remark 1.21 of Volume 1) and thus is semisimple, by Proposition 14.23. By Theorem 14.16, any semisimple Artinian module has finite composition length, so $\ell(R) = \sum_{i=1}^{t} \ell(M_i)$, as desired. □

Thus, the Artinian theory is stronger than the Noetherian theory, which will be discussed in Chapter 16. Some other useful facts are given in Exercises 22 and 23.

Nil subsets of left Artinian rings.

Having introduced nil subsets in Definition 13.11, we would like to know when they are nilpotent. We must proceed with care; the subset $\{e_{12}, e_{21}\}$ of $M_2(F)$ is nil but not nilpotent. Nevertheless, there are useful theorems of Burnside, Engel, and Wedderburn, that were unified by Jacobson by means of a closure condition. First, we bound the nilpotence index.

LEMMA 15.22. *Suppose L is a f.g. right module over a division ring D. If $N \subset \mathrm{End}\, L_D$ is nilpotent, then there is a base of L with respect to which each element of N acts as a strictly upper triangular matrix. In particular, $N^n = 0$, where $n = [L : D]$.*

Proof. $[NL : D] \geq [N^2L : D] \geq \cdots$, so there is some $i \leq n$ such that $[N^iL : D] = [N^{i+1}L : D]$. But then $[N^iL : D] = [N^mL : D]$ for all $m > i$, and thus must be 0 (taking m large enough), implying that $[N^iL : D] = 0$, and thus $N^i = 0$, implying that $N^n = 0$.

We get the desired base by letting $V_i = N^{n-i}L$, extending a base of V_1 to a base of V_2, and continuing inductively until we arrive at a base of V_n. Then $NV_i \subseteq V_{i-1}$, as desired. □

THEOREM 15.23. *Suppose R is left Artinian, and N is a nil subset satisfying the condition that for any a_1, a_2 in N there is $\nu = \nu(a_1, a_2) \in \mathbb{Z}$ such that $a_1a_2 + \nu a_2a_1 \in N$. Then N is nilpotent.*

Proof. We do this in several steps.

CASE I. R is simple Artinian, i.e., $R = M_n(D)$. Write $R = \operatorname{End} L_D$ for some minimal left ideal L, viewed as a right vector space of dimension n over D. We induct on n. Note that $\operatorname{Ann}_R L = 0$.

Any nilpotent subset of N has nilpotence index n, by Lemma 15.22. Thus, {nilpotent subsets of N} is Zorn and nonempty (since any single element of N comprises a nilpotent subset), and so has a maximal member $N' \neq 0$. In particular, $(N')^n = 0$, so $(N')^n L = 0$. Taking $V = N'L$, it follows that $0 \neq V < L$.

Let $m = [V : D] < [L : D] = n$, and let $N_1 = \{a \in N : aV \subseteq V\} \supseteq N'$. Note that $a_1a_2 + \nu a_2a_1 \in N_1$ for all $a_1, a_2 \in N_1$. Furthermore, N_1 acts as a nil set of linear transformations on V. Hence, by induction, $N_1^m V = 0$. Likewise, N_1 acts as a nil set of linear transformations on L/V, which has dimension $n - m$ over D, so, by induction, $N_1^{n-m} L \subseteq V$. Hence $N_1^n L = 0$, implying that $N_1^n = 0$, so $N_1 = N'$.

We are done if $N_1 = N$, so assume $N_1 \subset N$. Write $a_1 \cdot a_2$ for $a_1a_2 + \nu a_2a_1$. Thus $a_1a_2 = a_1 \cdot a_2 - \nu a_2a_1$. It follows that $a \cdot N_1 \not\subseteq N_1$, for each $a \in N \setminus N_1$. (Indeed, if $a \cdot N_1 \subseteq N_1$, then

$$aV = aN_1L \subseteq (a \cdot N_1)L + N_1aL \subseteq N_1L + N_1L = N_1L = V,$$

so $a \in N_1$.) Thus, taking $a_0 \in N \setminus N_1$, we can pick $a_1 \in (a_0 \cdot N_1) \setminus N_1$; iterating, given a_{i-1}, we pick $a_i \in (a_{i-1} \cdot N_1) \setminus N_1$. But a_{2n+1} is a sum of products in R (under the usual multiplication), each of which contains one occurrence of a_0 and $2n$ occurrences from N_1. In other words, each summand is a product in which at least n elements of $N_1 = N'$ are multiplied consecutively, and thus a_{2n-1} must be 0, a contradiction. Hence $N_1 = N$.

CASE II. R is semisimple Artinian. Then writing R as a finite product of simple Artinian rings, we apply Case I to each component of N.

CASE III. R is left Artinian. Then $J = \operatorname{Jac}(R)$ is nilpotent, say $J^k = 0$. By Case II, the image of N in R/J is nilpotent, i.e., $N^n \subseteq J$; hence $N^{kn} \subseteq J^k = 0$. □

Remark 15.23′. In Case I, where $R = M_n(D) = \operatorname{End} V_D$, we can pick a base of V over D with respect to which N is strictly upper triangular. (Indeed, we need to find a base $v_1, \dots, v_n$ in V for which $Nv_i \subseteq \sum_{j>i} v_j D$ for each i. Take k minimal such that $N^k V = 0$, and pick any $v \neq 0$ in $N^{k-1}V$. Then $Nv = 0$, so N acts nilpotently on $\bar{V} = V/vD$, and by induction we have $\bar{v}_1, \dots, \bar{v}_{n-1}$ with $N\bar{v}_i \subseteq \sum_{j>i} \bar{v}_j D$ for $1 \leq i \leq n-1$. We conclude by taking v_i with $\bar{v}_i = v_i + vD$ for $1 \leq i \leq n-1$, and $v_n = v$.)

Theorem 15.23 is generalized to left Noetherian rings in Theorem 16.31. Another variant is given in Theorem 15B.4. If we drop all restrictions on R, then there exists a commutative counterexample; cf. Exercise 31.

Finite-dimensional algebras.

Let us see what these results yield when R is a finite-dimensional (f.d.) algebra over a field F. Suppose $m = [R : F]$. Since every left ideal is a subspace of R, any chain of left ideals can have length at most n; likewise for right ideals. Hence R is Artinian and Noetherian.

Definition 15.24. A semisimple F-algebra R is **split** if

$$R \cong M_{n_1}(F) \times \cdots \times M_{n_t}(F)$$

for suitable t and suitable n_i, $1 \leq i \leq t$.

THEOREM 15.25. *Suppose R is a f.d. algebra over a field F. Then $\operatorname{Jac}(R)$ is a nilpotent ideal and $R/\operatorname{Jac}(R)$ is a direct product $R_1 \times \cdots \times R_t$ of f.d. simple F-algebras. If F is algebraically closed, then $R/\operatorname{Jac}(R)$ is split.*

Proof. Theorems 15.20 and 15.23 dispose of everything except the last assertion, which is Theorem 14.27. □

One can push this a bit farther.

PROPOSITION 15.25′. *Suppose R is a f.d. algebra over a field F, with $J = \operatorname{Jac}(R)$ of nilpotence index m. Let V be a complementary F-subspace of J^2 inside J. Then $J = \sum_{i=1}^m V^i$.*

Proof. Let $\hat{J} = \sum_{i=1}^m V^i$. By reverse induction, we show that $J^k \subseteq \hat{J}$ for each k. Indeed, $J^m = 0 \subseteq \hat{J}$, and given $J^k \subseteq \hat{J}$ for $k \leq m$, we have $J^{k-1} = (V + J^2)^{k-1} \subseteq V^{k-1} + J^k \subseteq \hat{J}$ (since $V \subseteq J$). □

Supplement: The classical theory of finite-dimensional algebras

Theorem 15.25 is far from the end of the story for f.d. algebras. Wedderburn proved the amazing result that R/J is actually isomorphic to a subalgebra of R.

THEOREM 15.26 (WEDDERBURN'S PRINCIPAL THEOREM). *Suppose R is a f.d. algebra over a field F and R/J is split, where $J = \operatorname{Jac}(R)$. (In particular, this is the case when F is algebraically closed.) Then the semisimple algebra*

$S = R/J$ is isomorphic to a subalgebra $\tilde{S}$ of R, which is a vector space complement of J in R; i.e., $R = \tilde{S} \oplus J$.

Proof. J is nilpotent, by Theorem 15.18. Once we have the subalgebra $\tilde{S}$, we know that $J \cap \tilde{S}$ is a nilpotent ideal of a semisimple algebra and is thus 0, and then

$$\dim_F \tilde{S} + \dim_F J = \dim_F R/J + \dim_F J = \dim_F R,$$

yielding $\tilde{S} \oplus J = R$. Thus it suffices to prove the first assertion.

We write S as $\bar{R}_1 \times \cdots \times \bar{R}_t$ where $\bar{R}_k \cong M_{n_k}(F)$ for $1 \leq k \leq t$. Let $e_k + J$ denote the unit element of $\bar{R}_k$. By Proposition 13.13, we may assume that $e_1, \dots, e_t$ are orthogonal idempotents of R. Let $\hat{R}_k = e_k R e_k \subset R$. Thus, $\hat{R}_k \hat{R}_\ell \subseteq R e_k e_\ell R = 0$ for each $k \neq \ell$. But $\bar{R}_k$ is clearly a homomorphic image of $\hat{R}_k$, so again appealing to Proposition 13.13, we may lift the matrix units of $\bar{R}_k$ to a set of matrix units $\{e_{ij}^{(k)}, 1 \leq i, j \leq n_k\}$ in R_k. Then $\sum F e_{ij}^{(k)}$ is a subalgebra of $\hat{R}_k$, which we denote as R_k; clearly $R_k \cong M_{n_k}(F) \cong \bar{R}_k$, and the subalgebra $\tilde{S}$ of R generated by $R_1, \dots, R_t$ is clearly isomorphic to

$$R_1 \times \cdots \times R_t \cong \bar{R}_1 \times \cdots \times \bar{R}_t = S. \qquad \square$$

We call $R = \tilde{S} \oplus J$ the **Wedderburn decomposition** of R. The Wedderburn decomposition holds more generally whenever the base field F is perfect; cf. Exercise 18.18. In more modern terms, Wedderburn's Theorem is best understood in the context of separable algebras; cf. Appendix 25B.

Together with Proposition 15.25′, Wedderburn's Theorem shows that any f.d. algebra R over an algebraically closed field F is "generated" by R/J and J/J^2; thus, at times, the theory of a f.d. algebra often reduces to the interplay of the (nilpotent) radical with matrix algebras. This issue is far from trivial, as evidenced by the fact that R is local iff $R/J \cong F$, yet there is great flexibility in the combinatorics of J, given J/J^2. We continue the investigation of f.d. algebras in Appendix 25C, bringing in techniques from the theory of Lie algebras.

Appendix 15A: Structure theorems for rings and algebras

In this appendix, we turn to the general structure theory, featuring Jacobson's general theory of primitive rings and the Jacobson radical, for rings not necessarily satisfying chain conditions. This theory works well for algebras over large fields, via a result of Amitsur.

The Chevalley-Jacobson Density Theorem.

Suppose M is a given R-module, and let $W = (\mathrm{End}_R M)^{\mathrm{op}}$. Clearly M is also a *right* W-module, under the scalar multiplication $af = f(a)$ for $a \in M$ and $f \in W$, since $a(fg) = fg(a) = (ag)f$ for all $a \in M$ and $f, g \in W$. M thereby becomes an R, W-bimodule, which is faithful as a right W-module (since $f(M) = 0$ iff $f = 0$).

Taking $\hat{R} = \mathrm{End}\, M_W$ and using Proposition 13.37, we have a natural homomorphism $\Phi : R \to \hat{R}$ given by $r \mapsto \ell_r$; Φ is an injection whenever the R-module M is faithful. Although it is too much to expect that Φ is onto, one has the celebrated **Density Theorem** that, for M semisimple, the image $\Phi(R)$ is dense in $\hat{R}$ with respect to the following topology:

Given a finite set $a_1, \dots, a_n \in M$ and $f_0 \in \hat{R}$, define

$$B(a_1, \dots, a_n; f_0) = \{f \in \hat{R} : f(a_i) = f_0(a_i),\ 1 \leq i \leq n\}.$$

The $\{B(a_1, \dots, a_n; f_0) : n \in \mathbb{N},\ a_1, \dots a_n \in M,\ f_0 \in \hat{R}\}$ comprise a sub-base of a topology on $\hat{R}$, called the **finite topology**.

Theorem 15A.1 (General Density Theorem). *If M is a semisimple R-module, then $\Phi(R)$ is dense in $\hat{R}$ (under the finite topology).*

Proof. We need to show that given $a_1, \dots, a_n \in R$ and $f \in \hat{R}$, there is r in R with $ra_i = f(a_i)$ for each i.

SPECIAL CASE. Assume that $n = 1$. We take a complement N' of $N = Ra_1$ in M and define the projection $\pi \colon M = N \oplus N' \to N$; i.e., $\pi(a + b) = a$, where $a \in N$ and $b \in N'$. Viewing π in W, we have

$$f(a_1) = f(a_1\pi) = (f(a_1))\pi \in N = Ra_1,$$

as desired.

GENERAL CASE. We reduce to the special case by considering $M^{(n)}$, which also is semisimple, and defining $\widehat{f} = (f, \dots, f) \colon M^{(n)} \to M^{(n)}$ diagonally. We claim that $\widehat{f} \in \mathrm{End}\, M^{(n)}{}_{W'}$, for $W' = \mathrm{End}_R(M^{(n)})^{\mathrm{op}}$. Indeed, any $h \in W'$ is given by a matrix (h_{ij}) of maps $h_{ij} \colon M \to M$, i.e.,

$$(a_1, \dots, a_n)h = \left(\sum_i a_i h_{i1}, \dots, \sum_i a_i h_{in}\right),$$

so

$$\begin{aligned}\widehat{f}((a_1,\dots,a_n)h) &= \widehat{f}\Big(\sum_i a_i h_{i1},\dots,\sum_i a_i h_{in}\Big)\\ &= \Big(f\Big(\sum a_i h_{i1}\Big),\dots,f\Big(\sum a_i h_{in}\Big)\Big)\\ &= \Big(\sum f(a_i)h_{i1},\dots,\sum f(a_i)h_{in}\Big)\\ &= (f(a_1),\dots,f(a_n))\,h\\ &= \widehat{f}(a_1,\dots,a_n)h\end{aligned}$$

for any $(a_1,\dots,a_n)$ in M and h in W'.

But we also have $\Phi'\colon R \to \operatorname{End} M^{(n)}{}_{W'}$ given by $\Phi(r) = \hat{r}$, where

$$\hat{r}(a_1,\dots,a_n) = (ra_1,\dots,ra_n),$$

and thus, by the special case applied to $M^{(n)}$, we have $\hat{r}$ such that

$$(a_1 f,\dots,a_n f) = \widehat{f}(a_1,\dots,a_n) = \hat{r}(a_1,\dots,a_n) = (ra_1,\dots,ra_n). \qquad \square$$

Now suppose the module M is simple. By Schur's Lemma (Proposition 13.40), $(\operatorname{End}_R M)^{\mathrm{op}}$ is a division ring D over which M is viewed as a vector space (on the right). Thus, any endomorphism is determined uniquely by its action on a base of M over D. In particular, any set of D-independent elements $a_1,\dots,a_n \in M$ together with any elements $b_1,\dots,b_n$ of M give rise to some $f \in \operatorname{End} M_D$ such that $f(a_i) = b_i$ for $1 \le i \le n$. The Density Theorem now says:

THEOREM 15A.2 (JACOBSON DENSITY THEOREM FOR SIMPLE MODULES). *Suppose M is a simple R-module, and $D = \operatorname{End}_R M$. For any $n \in \mathbb{N}$, any D-independent elements $a_1,\dots,a_n \in M$, and any elements $b_1,\dots,b_n$ of M, there is r in R such that $ra_i = b_i$ for $1 \le i \le n$.*

For example, let us take R to be any primitive ring. Then R has a simple faithful module M. Note that $\hat{R} = \operatorname{End} M_D$ is itself a primitive ring. Indeed, M is a faithful $\hat{R}$-module under the multiplication $fa = f(a)$ for $f \in \hat{R}$ and $a \in M$, and also is simple (under the criterion of Proposition 13.2) since for any a, b in M there is $f \in \hat{R}$ such that $f(a) = b$. (Simply take $n = 1$ in the discussion above.) Thus R sits densely inside the larger primitive ring $\hat{R}$, and in one important situation they are equal.

COROLLARY 15A.3. *If M is a faithful simple R-module and M has dimension $n < \infty$ as a right vector space over $D = \operatorname{End}_R M$, then R has the form $M_n(D)$ and thus is simple Artinian.*

Proof. In this case $\operatorname{End} M_D \cong M_n(D)$, so we need to show that Φ is onto. Take a base $b_1,\dots,b_n$ of M over D. Any endomorphism f is given by its action on the base. But by Jacobson's Density Theorem there is $r \in R$ such that $rb_i = f(b_i)$, $1 \le i \le n$, and thus $f \in \Phi(R)$. $\square$

Even when there is no extra restriction on a primitive ring R, the Density Theorem often enables us to display R precisely enough to employ techniques of linear algebra.

Perhaps surprisingly, the Density Theorem can be used to study finite-dimensional algebras over a field. If $\dim_F R = n$, then $R \hookrightarrow M_n(F)$ via the regular representation, and we would like to describe R explicitly. The following basic result of Burnside is an immediate consequence of the Density Theorem.

COROLLARY 15A.4. *Suppose A is a subalgebra of $M_n(F) = \operatorname{End} F^{(n)}$, for F an algebraically closed field, and $F^{(n)}$ is simple as an A-module. Then $A = M_n(F)$.*

Proof. By Proposition 14.28, $\operatorname{End}_A F^{(n)} \cong F$. Hence Corollary 15A.3 shows that $A = \operatorname{End} F^{(n)}{}_F = M_n(F)$. $\square$

Nil ideals and the Jacobson radical.

It is rather easy to see that any nil left or right ideal is contained in the Jacobson radical (cf. Exercise 15.27), and one of the basic questions concerning the Jacobson radical is when it is nil. Although the answer is negative, even in the commutative case (since local domains are counterexamples!), we have seen that the Jacobson radical of an Artinian ring is nilpotent, and other instances are given in Exercises A7 and A10. The following theorem of Amitsur provides an important general link between nil ideals and the Jacobson radical. Let $R[\lambda]$ denote the polynomial ring (cf. Definition 13.51).

THEOREM 15A.5 (AMITSUR). *If a ring R has no nonzero nil ideals, then* $\operatorname{Jac}(R[\lambda]) = 0$.

Proof. We start with a placid enough lemma:

LEMMA 15A.6. *If C is a commutative ring and $f = \sum_{i=0}^{t} c_i \lambda^i \in C[\lambda]$ with c_t not nilpotent, then $1 - \lambda f$ is not invertible.*

Proof. If $(1-\lambda f)g = 1$, then taking a prime ideal P of C disjoint from the powers of c_t (cf. Lemma 6.30 of Volume 1) and passing to $(C/P)[\lambda] \cong C[\lambda]/P[\lambda]$, we may assume that C is an integral domain with $c_t \neq 0$. But the leading term of $(1-\lambda f)g$ then has degree ≥ 1, contrary to $(1-\lambda f)g = 1$.

$\square$

Proof of Theorem 15A.5. Otherwise, take nonzero

$$f = \sum_{i=0}^{t} r_i \lambda^i \in J = \mathrm{Jac}(R[\lambda])$$

having a minimal possible number of nonzero coefficients r_i. By hypothesis Rr_tR contains some non-nilpotent element r_t', which we write as $\sum_j r_{j1} r_t r_{j2}$ for suitable r_{j1}, r_{j2} in R. Then

$$\sum_{i=0}^{t} \sum_j r_{j1} r_i r_{j2} \lambda^i = \sum_j r_{j1} f r_{j2} \in J, \tag{15A.1}$$

so replacing f by $\sum_j r_{j1} f r_{j2}$ we may assume that r_t is not nilpotent. Furthermore, for each nonzero r_i, $r_i f - f r_i \in J$ has fewer nonzero coefficients (since the i-th coefficient becomes 0), and thus must be 0; i.e., $r_i f = f r_i$ for each i. This means $r_0, \dots, r_t$ all commute. Let C be the (commutative) subring of R generated by 1 and $r_0, \dots, r_t$.

Since $f \in J$ by Proposition 15.17, we have $g = \sum s_j \lambda^j \in R[\lambda]$ such that

$$(1 - \lambda f) g = 1. \tag{15A.2}$$

We claim by induction on j that each $s_j \in C$. Indeed, $s_0 = 1$, comparing constant terms in (15A.2). In general, taking the coefficient of λ^j yields

$$0 = s_j + \sum_{k=0}^{j-1} c_k s_{j-k},$$

so $s_j = -\sum_{k=0}^{j-1} c_k s_{j-k} \in C$, by induction. But now (15A.2) contradicts Lemma 15A.6. □

There are many proofs of Amitsur's theorem, and a very different approach using graded algebras is given in Exercises 16.39–16.41. Of course this result fails for power series, since $F[[\lambda]]$ is a local ring for any field F. Other useful results relating the Jacobson radicals of a ring and a subring are given in Exercises A4ff.

Algebras over large fields.

We close this appendix with an observation of Amitsur that describes the structure of algebras over a large field.

Proposition 15A.7 (Amitsur). *If R is an algebra over a field F and $r \in R$ such that $\dim_F F[r] \geq t$, then for any distinct $\alpha_1, \dots, \alpha_t \in F$ such that $r - \alpha_i$ are each invertible, the elements $\{(r - \alpha_i)^{-1} : 1 \leq i \leq t\}$ are linearly independent over F.*

Proof. Assume $\sum_{j=1}^{t} \beta_j (r - \alpha_j)^{-1} = 0$. Let $q_i = \prod_{j \neq i} (\lambda - \alpha_j)$. Then r is a root of the polynomial $\sum \beta_i q_i$, which has degree $\leq t - 1$ and thus is 0. Hence for each $1 \leq k \leq t$,

$$0 = \sum_i \beta_i q_i(\alpha_k) = \beta_k q_k(\alpha_k) = \beta_k \prod_{j \neq k} (\alpha_k - \alpha_j),$$

implying $\beta_k = 0$ for each k. □

Corollary 15A.8. *If R is a division algebra over a field F such that $\dim_F R < |F|$ (i.e., the cardinality of a base of R over F is less than the cardinality of F), then R is algebraic over F.*

Proof. Otherwise, take $r \in R$ transcendental. The set $\{(r - \alpha)^{-1} : \alpha \in F\}$ is linearly independent, by the proposition (cf. Example 6.3 of Volume 1). Hence $\dim_F R \geq |F|$, contrary to hypothesis. □

Another application of Proposition 15A.7 involves the Jacobson radical; cf. Exercises A10 and A11.

Appendix 15B. Kolchin's Theorem and the Kolchin Problem

Kolchin proved the following analog to Theorem 15.23 and Lemma 15.22 (also to be compared with Theorem 21.32):

Recall that a matrix s is **unipotent** if $s - I$ is nilpotent. (Here I is the identity matrix.) A multiplicative subgroup G of $M_n(D)$ is called **unipotent** if g is unipotent for each $g \in G$. Unipotent matrices arise in the multiplicative Jordan decomposition (Proposition 2.76 of Volume 1).

Remark 15B.1. A matrix is unipotent iff all of its characteristic values are 1, in view of the Jordan canonical form (Corollary 2.69 of Volume 1).

Remark 15B.2. In characteristic 0, the multiplicative group generated by any upper triangular unipotent matrix s has infinite order. (Indeed, write $s = I + r$ for r nilpotent; $s^n = I + nr + \binom{n}{2} r^2 + \cdots$. Clearly, nr cannot be cancelled by the remaining terms; hence $s^n \neq I$, for each $n > 1$.)

However, when $\mathrm{char}(F) = p$, the matrix $s = \left(\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right)$ satisfies $s^p = I$ for arbitrary $\alpha \in F$.

Proposition 15B.3. *Suppose a matrix $s \in \mathrm{GL}(n, F)$, with $\mathrm{char}(F) = 0$. If s has finite order and all of the eigenvalues of s are n-th roots of unity, then $s^n = I$.*

Proof. Passing to the algebraic closure of F, we may assume that s is upper triangular. But s^n is unipotent, of finite order, and thus must be I, by Remark 15B.2. □

Having some feel for unipotent matrices, let us now consider unipotent groups.

THEOREM 15B.4 (KOLCHIN'S THEOREM). *For any algebraically closed field F, if S is a monoid of unipotent matrices of $M_n(F)$, then all the elements of S can be simultaneously triangularized via a suitable change of base; i.e., S is isomorphic to a submonoid of the group of upper triangular matrices having 1 on each diagonal entry.*

Proof. Let $\hat{S}$ be the F-vector space spanned by S. Clearly $\hat{S}$ is a subalgebra of $M_n(F) = \mathrm{End}_F F^{(n)}$, and it suffices to put $\hat{S}$ in upper triangular form. This means it is enough to find a chain of $\hat{S}$-modules

$$0 \subset V_1 \subset V_2 \subset \cdots \subset V_n = F^{(n)},$$

by the argument given in the last paragraph of the proof of Lemma 15.22: If we inductively expand a base of V_i to V_{i+1} for each i, then clearly $\hat{S}$ is in triangular form with respect to such a base.

It suffices to show that V has an $\hat{S}$-submodule V_1 that is 1-dimensional over F, since then we pass to the action of S on V/V_1 and conclude by induction.

We prove in fact that any simple $\hat{S}$-submodule M of V has dimension 1 over F. Indeed, take $m = \dim_F M$. By Corollary 15A.4, $\hat{S}$ acts on M as the full $m \times m$ matrix algebra. In other words, S contains m^2 linearly F-independent elements $a_1, \ldots, a_m$. But for any $s \in S$, the $m \times m$ matrix $\tilde{s}$ corresponding to s, being unipotent, has trace m, so the m^2 equations $\mathrm{tr}(a_i \tilde{s}) = m$ have a unique solution by Lemma 00.8. This means S has a unique element; hence, $m^2 = 1$, implying $m = 1$ (and $s = 1$). □

It is not difficult to pass to the algebraic closure, so one gets the analog of Kolchin's Theorem when F is any field; cf. Exercise B5. More generally, for any division ring D, we write $\mathrm{GL}(n, D)$ for the multiplicative group of units of $M_n(D)$. Since Kolchin's Theorem is an analog to Theorem 15.23, which holds for arbitrary simple Artinian rings, B. Plotkin and I. Kaplansky asked whether Kolchin's Theorem can be extended to unipotent subgroups (or, better yet, submonoids) of $\mathrm{GL}(n, D)$. This question is known as "the Kolchin Problem" and is particularly intriguing since it is hard to tell whether the best method of attack is group-theoretical or ring-theoretical. For example, here is an alternate formulation.

General form of the Kolchin Problem. *Suppose S is a nil subset of $M_n(D)$ satisfying the condition that for any s_1, s_2 in S, $s_1 s_2 + s_1 + s_2 \in S$. Then $S^n = 0$.*

(Indeed, one takes $G = \{1 + s : s \in S\}$.) This formulation is so close to Theorem 15.23 that one would expect an elementary proof, but so far none is forthcoming. Using Zelmanov's solution of the Restricted Burnside Problem (!) (to be discussed in Appendix 17C), Heinicken proved:

Suppose $\mathrm{char}(F) = p$ where $p = 0$ or $p > (n-1)\left(n - \left[\frac{n}{2}\right]\right)$. If G is a f.g. subgroup of $\mathrm{GL}(n, D)$ such that $(h-1)^n = 0$ for all $h \in H$, then there is a number k such that $(h_1 - 1) \cdots (h_k - 1) = 0$, $\forall h_1, \ldots, h_k \in G$.

Mochizuki showed that since one can take $k = n$, this result works for arbitrary subgroups and thus solves the Kolchin Problem for groups in characteristic 0 or for large enough p. However, this is an exceedingly complicated solution of a problem whose proof over a field is straightforward enough. Also, the monoid formulation of the Kolchin Problem remains open. The Kolchin Problem is treated further in Exercises B3ff, including Derakshen's solution in characteristic 2 (Exercise B2).

Chapter 16

Noetherian Rings and the Role of Prime Rings

In Chapter 15 we introduced structure theory in order to study left Artinian rings. In this (optional) chapter, we consider the structure of left Noetherian rings, which by Theorem 15.21 are more inclusive than left Artinian rings. We have already seen some important examples, including commutative affine algebras, the Weyl algebras, skew polynomial rings, and, more generally, Ore extensions; other examples are to follow. The theory of left Noetherian rings has been one of the main areas of research in noncommutative ring theory in the second half of the twentieth century, involving sophisticated techniques; our object here is only to lay the foundation and to glance at how it interfaces with the structure of rings.

We encounter some serious obstacles at the very outset. First of all, our major theorems about Artinian rings fail for Noetherian rings. For example, $\mathbb{Z}$ has an infinite number of maximal ideals. On the other hand, the first Weyl algebra $\mathcal{A}_1(F)$ is a simple Noetherian domain. Although the Jacobson program remains very important, we need additional techniques.

From a different viewpoint, just as the ring of fractions of an integral domain is a field, the ring of fractions of a left Noetherian domain turns out to be a division ring. However, "domain" is too strong a notion for a general role in the noncommutative structure theory, since matrices over a domain are not a domain. Thus, we want a more general class of left Noetherian rings whose rings of fractions are simple Artinian. We introduced prime rings briefly in Definition 15.11, but proved quickly that prime Artinian

rings are simple. It is in the Noetherian theory that prime rings are given the feature role that they deserve.

Prime rings

The hierarchy of simple, primitive, and prime rings is given in:

Remark 16.0. (i) Any simple ring is primitive, by Example 15.5.

(ii) Any primitive ring R is prime, by Remark 15.13(i).

We have seen counterexamples to the converses.

Example 16.1. Here are some major examples of prime Noetherian rings that are not Artinian:

(i) $M_n(C)$ for any Noetherian ring C that is not Artinian (and modifications such as in Exercise 2).

(ii) The Weyl algebras of Examples 13A.1 and 13A.2 (which are actually simple domains but not division rings).

(iii) Skew polynomial rings (and, more generally, Ore extensions) over Noetherian domains (cf. Exercise 13A.7).

(iv) In Appendix 21C we obtain another important example of a Noetherian domain — the enveloping algebra of a finite-dimensional Lie algebra.

Influenced by the prominence of integral domains in the commutative theory, we are led at times to build a theory using prime rings instead of primitive rings at the foundation. We start by reviewing when a ring is prime.

PROPOSITION 16.2. *The following conditions are equivalent for a ring R:*

(i) *R is a prime ring.*

(ii) *$\operatorname{Ann}_R L = 0$ for any nonzero left ideal L of R.*

(iii) *If $aRb = 0$, then $a = 0$ or $b = 0$.*

Proof. (i) $\Rightarrow$ (ii) By Remark 15.13′(ii).

(ii) $\Rightarrow$ (iii) Take $L = Rb$.

(iii) $\Rightarrow$ (i) Suppose $A, B \lhd R$ with $AB = 0$ but $B \neq 0$. Take $b \neq 0$ in B. Then, for each $a \in A$, $aRb = 0$, implying that each $a = 0$. □

Sometimes it is easiest to verify that a ring is prime via the following easy idea, which yields surprisingly many examples.

Definition 16.3. Suppose R is a subring of W. Another subring T of W **centralizes** R if $ar = ra$, $\forall a \in T$, $r \in R$. W is a **centralizing extension** of R if $W = RT$ for some subring T centralizing R.

W is a **central extension** of R when $W = R\,\mathrm{Cent}(W)$.

(A more general concept was considered briefly in Exercise 15A.3.) Although we focus on central extensions in this chapter, there are important examples of centralizing extensions that are not central extensions, such as $M_n(R)$ for any F-algebra R (taking T to be $M_n(F)$, which contains all of the matrix units).

Remark 16.3′. If a prime ring W is a centralizing extension of R, then R is also a prime ring. (Indeed, notation as in Definition 16.3, if $0 \neq A, B \lhd R$, then $0 \neq AT, BT \lhd W$, implying that $0 \neq (AT)(BT) = (AB)T$, so $AB \neq 0$.)

Prime ideals.

Definition 16.4. An ideal P of R is **prime** if R/P is a prime ring.

Our next observations about a prime ideal P are all proved easily, by passing to the ring R/P, in analogy to Lemma 6.15 of Volume 1.

Remark 16.5. The following conditions are equivalent for an ideal P of R:

(i) P is a prime ideal.

(ii) For any ideals A, B of R properly containing P, we have $AB \not\subseteq P$.

(iii) For any ideals A, B of R not contained in P, we have $AB \not\subseteq P$.

(Indeed, (ii) $\Rightarrow$ (i) $\Rightarrow$ (iii) is seen by passing to R/P, and (iii) $\Rightarrow$ (ii) is formal.)

Remark 16.6. If P is a prime ideal and $A_1 \cdots A_t \subseteq P$ for suitable $A_i \lhd R$, then some $A_i \subseteq P$. (This is seen by induction applied to Remark 16.5(ii).)

The lower nilradical.

Prime ideals lead to an alternative to the Jacobson radical.

Definition 16.7. The **lower nilradical** $N(R)$ of a ring R is the intersection of all the prime ideals in R.

Since every primitive ideal is prime, clearly $N(R) \subseteq \mathrm{Jac}(R)$.

Proposition 16.8. *$N(R)$ is a nil ideal of R and contains all nilpotent ideals of R.*

Proof. If $A^t = 0$ for $A \lhd R$, then for every prime ideal P of R we have $A^t \subseteq P$, implying that $A \subseteq P$; hence

$$A \subseteq \cap\{\text{Prime ideals of } R\} = N(R).$$

We show that $N(R)$ is nil by proving for any non-nilpotent element r of R that R has a prime ideal P not containing r. Indeed, the set of ideals disjoint from $S = \{r^i : i \in \mathbb{N}\}$ is Zorn and nonempty since $0 \notin S$, and thus contains a maximal member P. We claim that P is a prime ideal. Indeed, if B_1, B_2 are ideals containing P, then B_1, B_2 contain respective powers r^i, r^j of r, so $r^{i+j} \in B_1B_2$, proving that $B_1B_2 \not\subseteq P$. □

Corollary 16.8′. *If $N(R) = 0$, then R is a semiprime ring.*

The converse is true but harder; cf. Exercise 4. This leads to another description of $N(R)$ in Exercise 5.

The same kind of argument gives us another way of looking at algebras over uncountable fields; cf. Exercises 7 and 8. In analogy to Jacobson's program, the **Baer-Levitzki-Goldie-Herstein program** studies a ring R in terms of $N(R)$ and the semiprime ring $R/N(R)$. The basic questions to be addressed are:

1. What can one say about prime rings? How far are they from being simple Artinian?
2. What is the structure of semiprime rings?
3. When is $N(R)$ nilpotent?

For the remainder of this chapter we shall analyze these questions, mostly for left Noetherian rings. In general, the sum of all nil ideals is a nil ideal called the **upper nilradical** $\mathrm{Nil}(R)$; cf. Exercise 6. $N(R) \subseteq \mathrm{Nil}(R)$ by Proposition 16.8, and one also asks for which classes of rings do we have $N(R) = \mathrm{Nil}(R)$? (This can fail even for commutative rings; cf. Exercise 15.31.) We obtain some satisfactory results in Proposition 16.26 and Theorem 16.31.

Even when $\mathrm{Nil}(R) = 0$, it is unknown in general whether R can have a nonzero nil left ideal; this is one of the equivalent formulations of the famous **Koethe question**. Recently Koethe's question has been chipped at by means of ingenious examples, by A. Smoktunovicz and others, but it remains unanswered.

A third program would involve the intersection of the *maximal* ideals. This ideal, called the **Brown-McCoy radical**, would then be 0 iff R is the subdirect product of simple rings. Although the Brown-McCoy radical is usually bigger than the Jacobson radical, one finds it almost as difficult to study simple rings as primitive rings. Thus, this approach has faded from use.

A fourth structural program is started in Exercises 9 and 10. Although more technical in nature, it has some pretty applications, such as in Exercise 23.3.

Rings of fractions and Goldie's Theorems

We turn to Goldie's contribution, the noncommutative analog of the basic fact that every integral domain has a field of fractions.

Noncommutative rings of fractions.

Motivated by the results of Chapter 8 of Volume 1, we want to generalize the notion of "ring of fractions" to rings that need not necessarily be domains. Throughout this discussion, S is a multiplicative submonoid of R, not containing 0, and s always denotes an element of S. We want to view S as the set of denominators in a suitable ring containing R. More precisely we want the *ring of fractions* $S^{-1}R$ of R *with respect to* S (if it exists) to be a ring together with a homomorphism $\nu\colon R \to S^{-1}R$ satisfying the following properties:

(i) $\nu(s)$ is invertible in $S^{-1}R$ for all s in S.

(ii) Every element of $S^{-1}R$ has the form $\nu(s)^{-1}\nu(r)$ for suitable r in R, s in S.

(iii) (Universality) Suppose $\varphi\colon R \to T$ is any ring homomorphism such that $\varphi(s)$ is invertible in T for every s in S. Then φ extends to a unique homomorphism $\hat{\varphi}\colon S^{-1}R \to T$ satisfying $\varphi = \hat{\varphi}\circ\nu$; i.e., the following diagram commutes:

(16.1)
$$\begin{array}{ccc} R & & \\ {\scriptstyle\nu}\downarrow & \searrow^{\varphi} & \\ S^{-1}R & \xrightarrow[\hat{\varphi}]{} & T \end{array}$$

Before trying to construct $S^{-1}R$ and ν, we consider a consequence of their existence, in order to gain intuition.

Remark 16.9. The homomorphism $\hat{\varphi}$ of Condition (iii) must satisfy the formula
$$\hat{\varphi}(\nu(s)^{-1}\nu(r)) = \hat{\varphi}(\nu(s))^{-1}\hat{\varphi}(\nu(r)) = \varphi(s)^{-1}\varphi(r),$$
proving that $\hat{\varphi}$ is uniquely defined.

By abstract nonsense (as in Proposition 8.6 of Volume 1), our construction $S^{-1}R$ and ν (if they exists) must be unique up to isomorphism. We are

particularly interested in knowing when ν is an injection. Thus, as in the commutative case, we consider the situation where $R \subseteq Q = S^{-1}R$, and we write $s^{-1}r$ instead of $\nu(s)^{-1}\nu(r)$, suppressing ν in the notation. In this setup, the elements of S must satisfy another property.

Definition 16.10. An element $s \in R$ is **left regular** if $rs \neq 0$ for all $r \neq 0$ in R; s is **regular** if $rs \neq 0$ and $sr \neq 0$ for all $r \neq 0$ in R. Clearly, all nonzero elements of a domain are regular.

Of course, if $R \subseteq Q$ and $s \in S$ is invertible in Q, then s must be regular in R, for if $rs = 0$ in R, then, computing in Q, we get $0 = rss^{-1} = r$.

Looking for inspiration from the commutative case, we have a difficulty at the very outset — should a fraction $\frac{r}{s}$ be treated as $s^{-1}r$ or rs^{-1}? (These could be different!) We cope with this ambiguity as follows:

Remark 16.11. We require $rs^{-1} = (1^{-1}r)(s^{-1}1) \in S^{-1}R$ for any $r \in R$ and $s \in S$, so by Condition (ii),

$$rs^{-1} = s'^{-1}r' \tag{16.2}$$

for suitable r', s' with $s' \in S$. Solving (16.2) yields the (left) **Ore condition**

$$s'r = r's \quad \text{for suitable} \quad r', s' \quad \text{with} \quad s' \in S. \tag{16.3}$$

From now on, we only consider left fractions $s^{-1}r$.

Remark 16.12. The Ore condition implies in particular that any two elements r, s of S have a common left multiple, namely $(s'r = r's)$, which belongs to S. By induction, any finite number of elements of S have a common left multiple in S, so any finite number of (left) fractions can be written with a common left denominator. Indeed, given $s_i^{-1}r_i$, $1 \leq i \leq t$, we take a common left multiple $s = r_i's_i \in S$ of the s_i, and then note that $s_i^{-1} = s^{-1}r_i'$, implying

$$s_i^{-1}r_i = s^{-1}r_i'r_i, \quad 1 \leq i \leq t. \tag{16.4}$$

In the commutative theory, an integral domain has a field of fractions in which any element has the form $\frac{r}{s}$. In the noncommutative setting, one would like a domain R to have a ring of fractions D that is a division ring, and more generally, one would like a prime ring to have a simple Artinian ring of fractions.

Note that the subset of *all* regular elements of R is a monoid; indeed, if s_1, s_2 are regular and $0 = r(s_1s_2) = (rs_1)s_2$, then $rs_1 = 0$, implying $r = 0$.

Definition 16.13. A ring $Q = Q(R)$ is a **left ring of fractions** of R if Q is the ring of (left) fractions with respect to the set S of all regular elements, for which also the homomorphism ν is an injection.

The ring of fractions Q need not exist for arbitrary R, but, by universality, if Q does exist, then Q is unique up to isomorphism.

Note that Equation (16.4) (for $t = 2$) tells us how to add elements of Q:

$$(s_1^{-1}r_1) + (s_2^{-1}r_2) = s^{-1}(r_1'r_1 + r_2'r_2).$$

Multiplication in Q likewise can be defined via the Ore condition; cf. Exercise 11. Here is an important special case.

Definition 16.14. A domain R is called an **Ore domain** if it satisfies the Ore condition with respect to $S = R \setminus \{0\}$, or in other words, if $Rr \cap Rs \neq 0$ for any $0 \neq r, s$ in R.

Remark 16.15. Given an Ore domain R, we can build the desired left ring of fractions, taking $S = R \setminus \{0\}$. First we define an equivalence on $S \times R$ by saying that $(s_1, r_1) \sim (s_2, r_2)$ if there exist $a_i \in R$ such that $a_1s_1 = a_2s_2 \in S$ and $a_1r_1 = a_2r_2$. (Intuitively this means

$$s_1^{-1}r_1 = (a_1s_1)^{-1}(a_1r_1) = (a_2s_2)^{-1}(a_2r_2) = s_2^{-1}r_2.)$$

Writing $s^{-1}r$ for the equivalence class of (s, r), we define addition and multiplication of $s_1^{-1}r_1$ and $s_2^{-1}r_2$ as follows:

Taking $0 \neq a_i \in R$ such that $s = a_1s_1 = a_2s_2$, define

$$s_1^{-1}r_1 + s_2^{-1}r_2 = s^{-1}(a_1r_1 + a_2r_2).$$

Taking $0 \neq a_i \in R$ such that $a_1r_1 = a_2s_2$, define

$$(s_1^{-1}r_1)(s_2^{-1}r_2) = (a_1s_1)^{-1}(a_2r_2).$$

Checking that D is an associative ring under these operations, one concludes that D is a division ring since $(s^{-1}r)^{-1} = r^{-1}s$. (We leave the proof to the reader, since soon we obtain a more general verification.)

The analogous construction yields a left ring of fractions of any ring R with respect to any submonoid S satisfying appropriate conditions given in Exercise 12. Unfortunately, this method does not always interface so easily with other ring-theoretic properties. We do get the following useful fact from this approach.

LEMMA 16.16. *Suppose R has a left ring of fractions Q. For any left ideal L of R, the left ideal QL of Q has the form $\{s^{-1}a : s \in S,\ a \in L\}$.*

Proof. Any element of QL can be written as $\sum {s_i}^{-1}r_ia_i$ where $a_i \in L$. Applying Equation (16.4) (with r_ia_i replacing r_i), we see that this has the form $s^{-1}a$, where $a = \sum r_i'r_ia_i \in L$. □

A major breakthrough came with Goldie's discovery that left Noetherian domains are Ore, via the following argument, generalized to arbitrary left Noetherian rings:

PROPOSITION 16.17. *If $L < R$ and $Rs \cap L = 0$ with $s \in R$ left regular, then the left ideals $L, Ls, Ls^2, \ldots$ are independent (as submodules of R; cf. Definition 2.15 of Volume 1).*

Proof. Otherwise $\sum_{i=m}^{n} a_is^i = 0$ for suitable $a_i \in L$ with $a_m \neq 0$, implying that

$$\sum_{i>m} a_is^{i-m} = -a_m \in Rs \cap L = 0,$$

implying $a_m = 0$, a contradiction. □

COROLLARY 16.18. *If R is left Noetherian, then $Rs \cap L \neq 0$ for every left regular $s \in R$ and each $0 \neq L < R$. In particular, any left Noetherian domain R is Ore.*

Proof. The first assertion is the contrapositive of Proposition 16.17, since R, being left Noetherian, cannot have an infinite set of independent left ideals. The second assertion follows at once, taking $L = Rr$, since every nonzero element of a domain is regular. □

A good example is the Weyl algebra, which we recalled in Example 16.1 is a simple Noetherian domain, but *not* Artinian; yet it does have a ring of fractions that is a division ring (and thus simple Artinian).

The role of large left ideals.

Let us write $N <_{\mathrm{e}\ R} M$ to denote that N is a large R-submodule of M, which we recall from Definition 14.9 means $N \cap K \neq 0$ for all $0 \neq K < M$. Since every submodule contains a cyclic submodule, it suffices to check that $N \cap Ra \neq 0$ for every $0 \neq a \in M$. We delete $_R$, if R is understood. Thus, $L <_{\mathrm{e}} R$ means L is a large left ideal of R.

Note that the intersection of a finite number of large submodules is large.

LEMMA 16.19.
(i) *If $f: M \to N$ is a map of R-modules and $L <_e N$, then $f^{-1}(L) <_e M$.*
(ii) *If $L <_e R$ and $r \in R$, then $\{a \in R : ar \in L\} <_e R$.*

Proof. (i) For any $0 \neq K \leq M$ we want $K \cap f^{-1}(L) \neq 0$. This is obvious if $f(K) = 0$, so assume that $f(K) \neq 0$. Then $f(K) \cap L \neq 0$, so for any $0 \neq f(k) \in f(K) \cap L$, we have $k \in K \cap f^{-1}(L)$.

(ii) By (i), taking $f: R \to Rr$ to be right multiplication by r. □

Now we link large R-submodules to large Q-submodules when Q is a left ring of fractions of R.

Remark 16.20. For any R-submodule K of Q, if $s^{-1}r \in K \leq Q$, then $r \in R \cap K$.

LEMMA 16.21. *Suppose R has a left ring of fractions Q.*
(i) *$R <_{e\ R} Q$.*
(ii) *If $L <_e R$, then $QL <_{e\ Q} Q$.*
(iii) *If $L <_{e\ Q} Q$, then $L <_{e\ R} Q$, and thus $L \cap R <_e R$.*

Proof. (i) By Remark 16.20.

(ii) $L <_{e\ R} Q$ in view of (i), so a fortiori $QL <_{e\ R} Q$, and thus $QL <_{e\ Q} Q$.

(iii) Suppose $0 \neq K < {}_R Q$. By Remark 16.20, we have some $0 \neq r \in K \cap R$. Then $Qr \cap L \neq 0$, so there exists $0 \neq s'^{-1}r'r \in L$, implying that $0 \neq r'r \in K \cap L$. This shows that $L <_{e\ R} Q$. Thus $L \cap R <_{e\ R} Q$, implying that $L \cap R <_e R$. □

From now on, we take S to be the set of all regular elements of R. We are ready to characterize when R has a semisimple left ring of fractions, focusing on the following two properties:

(G1) $Rs <_e R$ for each $s \in S$.

(G2) $L \cap S \neq \emptyset$, for every large left ideal L of R.

PROPOSITION 16.22. *Suppose R has a left ring of fractions Q. The ring Q is semisimple iff (G2) holds in R.*

Proof. We use the characterization from Theorem 14.13 that the ring Q is semisimple, iff the only large left ideal of Q is Q itself. Also we appeal repeatedly to Lemma 16.21.

($\Rightarrow$) Suppose $L <_e R$. Then $QL <_{e\ Q} Q$, so $QL = Q$. Hence $1 = s^{-1}a$ for some a in L and s in S. But then $s = a \in L$.

($\Leftarrow$) Suppose $L <_{e\ Q} Q$. Then $L \cap R <_e R$, and thus contains some $s \in S$. Hence $1 = s^{-1}s \in L$. □

THEOREM 16.23 (GOLDIE). *A ring R has a semisimple left ring of fractions iff R satisfies properties* (G1) *and* (G2).

Proof. ($\Rightarrow$) To prove (G1), we appeal to Proposition 16.17. Assume on the contrary that $Rs \cap Ra = 0$ with $a \neq 0$. We claim that the left ideals

$$Qa, \quad Qas, \quad Qas^2, \quad \dots$$

of Q are independent. Otherwise we have some relation $\sum_{i=m}^{n} q_i a s^i = 0$ for $q_i \in Q$, not all 0; but writing $q_i = s_0^{-1} r_i$ we have $\sum_{i=m}^{n} r_i a s^i = 0$, contrary to Proposition 16.17.

We have contradicted the finite composition length of Q. Thus, (G1) must hold, after all.

(G2) holds, by Proposition 16.22.

($\Leftarrow$) At last we construct the left ring of fractions, which then is semisimple by Proposition 16.22. Our method is to formalize the fact that right multiplication by s^{-1} yields a map $f_s: Rs \to R$ sending $rs \mapsto r$. Let

$$\mathcal{S} = \{(L, f) \mid L <_e R,\ f: L \to R \quad \text{is a map of } R\text{-modules}\}.$$

We define an equivalence relation on $\mathcal{S}$ by:

$(L_1, f_1) \sim (L_2, f_2)$, iff f_1 and f_2 are equal on a large left ideal contained in $L_1 \cap L_2$.

We denote the equivalence class of (L, f) as $[L, f]$. Let $Q(R)$ denote the set of equivalence classes, endowed with addition

$$[L_1, f_1] + [L_2, f_2] = [L_1 \cap L_2, f_1 + f_2]$$

and multiplication

$$[L_1, f_1][L_2, f_2] = [L, f_2 \circ f_1]$$

where $L = f_1^{-1}(L_2)$, a large left ideal by Lemma 16.19. These operations are easily seen to be well-defined, yielding a ring structure whose zero element is $[R, 0]$; furthermore, we have a homomorphism $\nu: R \to Q(R)$ sending r to $[R, g_r]$, where g_r is right multiplication by r. We claim that ν is an injection. Indeed, $[R, g_r] = [R, 0]$ iff g_r vanishes on a large left ideal L of R. But $L \cap S$ must contain some element s, so $sr = g_r(s) = 0$, implying $r = 0$.

Thus, we may view R as a subring of $Q(R)$. Next, we show that every element $[L, f]$ of $Q(R)$ has the form $\nu(s)^{-1}\nu(r)$ where $r \in R$ and $s \in S$. Taking $s \in L \cap S$, let $r = f(s)$. Define $f_s: Rs \to R$ by $rs \mapsto r$. Clearly, $(Rs, f_s) = \nu(s)^{-1}$. For any $r's \in Rs$, we have

$$f(r's) = r'f(s) = r'r,$$

so the restriction of f to the large left ideal Rs corresponds in $Q(R)$ to right multiplication of $\nu(r')\nu(s)$ by $\nu(s)^{-1}\nu(r)$.

Finally, we verify universality: Given a ring homomorphism $\varphi\colon R \to T$ such that $\varphi(s)$ is invertible for all $s \in S$, we must define $\hat{\varphi}\colon Q(R) \to T$ satisfying $f = \widehat{\varphi} \circ \nu$.

First note that $\varphi(s)$ invertible implies, for any $a \in R$, that

$$\varphi(a) = \varphi(as)\varphi(s)^{-1},$$

and thus, when $as \in S$, $\varphi(as)^{-1}\varphi(a) = \varphi(s)^{-1}$.

To define $\hat{\varphi}$, take any $[L, f] \in Q(R)$, and picking $s \in L \cap S$, define

$$\hat{\varphi}([L, f]) = \varphi(s)^{-1}\varphi(f(s)).$$

To see that this is independent of the choice of $s \in L$, take any $s' \in L$; then $Rs \cap Rs'$ contains a regular element s'', and writing $s'' = as = a's'$, we have

$$\begin{aligned}\varphi(s'')^{-1}\varphi(f(s'')) &= \varphi(as)^{-1}\varphi(f(as)) \\ &= \varphi(as)^{-1}\varphi(a)\varphi(f(s)) \\ &= \varphi(s)^{-1}\varphi(f(s)),\end{aligned}$$

and likewise $\varphi(s'')^{-1}\varphi(f(s'')) = \varphi(s')^{-1}\varphi(f(s'))$.

It is easy to see now that $\hat{\varphi}$ is well-defined, and it remains to show that $\hat{\varphi}$ is a homomorphism. Given $[L_1, f_1]$ and $[L_2, f_2]$ in $Q(R)$, we take $s_2 \in L_2 \cap S$ and $s \in f_1^{-1}(Rs_2) \cap S$; writing $f_1(s) = as_2$, we have

$$\begin{aligned}\varphi([L, f_2 \circ f_1]) &= \varphi(s)^{-1}\varphi(f_2 f_1(s)) \\ &= \varphi(s)^{-1}\varphi(f_2(as_2)) \\ &= \varphi(s)^{-1}\varphi(a)\varphi(f_2(s_2)) \\ &= \varphi(s)^{-1}\varphi(f_1(s))\varphi(s_2)^{-1}\varphi(f_2(s_2)) \\ &= \varphi([L_1, f_1])\varphi([L_2, f_2]),\end{aligned}$$

as desired. □

The proof of Theorem 16.23 can be viewed in a much broader context, some instances of which are given in Exercises 23 and 24; the full categorical generality is outside the scope of this text.

Rings with ACC(ideals).

Conditions (G1) and (G2), discovered by Goldie, are *not* called the Goldie conditions. The real Goldie conditions, defined in Exercise 15, are

chain conditions equivalent to (G1) and (G2), and thus to R having a semisimple left ring of fractions. Goldie's conditions are formally weaker than R being left Noetherian, and thus imply that left Noetherian rings have a semisimple left ring of fractions. We prove the latter fact here and leave the finer tuning for Exercises 14–16. Of course, not every prime Goldie ring is left Noetherian — any non-Noetherian integral domain is an example.

Recall the definition of ACC from Definition 7.1 of Volume 1. Thus, the left Noetherian condition is precisely ACC(left ideals); consequently, a ring satisfying the weaker condition ACC(ideals) is sometimes called **weakly Noetherian**.

THEOREM 16.24. *Suppose R is a ring satisfying ACC(ideals). Then*

(i) *Every ideal A contains a finite product $P_1 \cdots P_t$ for prime ideals $P_i \supset A$.*
(ii) *There is a finite product of prime ideals that is 0.*
(iii) *R has only finitely many minimal prime ideals.*
(iv) *The intersection N of the prime ideals of R is nilpotent.*

Proof. (i) is by Theorem 9.2 of Volume 1 (applied to R/A), which did not require commutativity of R.

(ii) is the case where the ideal A is 0.

(iii) is by Corollary 9.3 of Volume 1, which also did not require commutativity of R.

(iv) If $P_1 \cdots P_t = 0$, then $N^t \subseteq P_1 \cdots P_t = 0$. □

(Parts (ii), (iii), and (iv) hold under a weaker hypothesis; cf. Exercise 14.) The following technical result is quite useful. Given $r \in R$, we write $\ell(r)$ for $\operatorname{Ann}_R r = \{r' \in R : r'r = 0\}$.

LEMMA 16.25. *Suppose R satisfies ACC($\ell(r) : r \in R$). Then any nonzero nil right ideal T contains an element $a \neq 0$ such that $aRa = 0$.*

Proof. Take $0 \neq a \in T$ with $\ell(a)$ maximal possible. We claim that $ara = 0$ for all $r \in R$. Indeed, take $r \in R$ arbitrary with $ar \neq 0$, and take n maximal such that $(ar)^n \neq 0$ (noting that $ar \in T$ is nilpotent). Since $\ell(aw) \geq \ell(a)$ for all w in R, we have $\ell(aw) = \ell(a)$ for all w such that $aw \neq 0$, and in particular,

$$\ell((ar)^n) = \ell(a(rar \cdots ar)) = \ell(a).$$

But this implies $ar \in \ell((ar)^n) = \ell(a)$, i.e., $ara = 0$, as desired. □

Of course, $aRa = 0$ implies $(RaR)^2 = 0$. Thus we have

PROPOSITION 16.26 (LEVITZKI). *If R is semiprime and satisfies ACC on left ideals of the form $\{\ell(r) : r \in R\}$, then R has no nonzero nil right ideals, and also R has no nonzero nil left ideals.*

Proof. R has no nonzero nil right ideals (since R would then have a nonzero nilpotent ideal RaR, by the lemma). We conclude the proof by means of the next remark. □

Remark 16.26′. A ring R has a nonzero nil left ideal L iff R has a nonzero right ideal. (Indeed, by symmetry, it suffices to prove ($\Rightarrow$). For any nonzero a in L and any r in R we have $n = n(r)$ such that $(ra)^n = 0$. But then $(ar)^{n+1} = a(ra)^n r = 0$, so we see that aR is a nonzero nil right ideal.)

LEMMA 16.27. *In any ring R, if $\ell(a) = \ell(a^2)$, then $\ell(a) \cap Ra = 0$.*

Proof. If $ra \in \ell(a) \cap Ra$, then $ra^2 = 0$, implying $ra = 0$. □

Prime and semiprime left Noetherian rings.

We are finally ready for the major structure theorem of left Noetherian rings. First we show that left regular implies right regular in left Noetherian rings.

PROPOSITION 16.28. *If R is semiprime left Noetherian and $s \in R$ with $\ell(s) = 0$, then s is regular.*

Proof. Otherwise, let $I = \{r \in R : sr = 0\} \neq 0$, a right ideal of R which thus is non-nil, by Proposition 16.26. Thus, I contains a non-nilpotent element b with $\ell(b)$ maximal possible. Then $\ell(b) = \ell(b^2)$ so, by Lemma 16.27, $0 = \ell(b) \cap Rb \supseteq Rs \cap Rb$, contrary to Corollary 16.18. □

THEOREM 16.29 (GOLDIE). *Any semiprime left Noetherian ring R satisfies properties (G1) and (G2), and thus has a semisimple left ring of fractions.*

Proof. (G1) follows from Corollary 16.18, so it remains to prove (G2). Suppose $L <_e R$. By Proposition 16.26, L has a non-nilpotent element a_1, which we choose with $\ell(a_1)$ maximal possible. Let

$$A_1 = L \cap \ell(a_1) \neq 0.$$

Inductively, for each $i \geq 1$, if $A_i \neq 0$, take non-nilpotent $a_{i+1} \in A_i$ with $\ell(a_{i+1})$ maximal possible, and let $A_{i+1} = A_i \cap \ell(a_{i+1})$.

By choice of each a_j we have $\ell(a_j) = \ell(a_j^2)$. It follows for each k that

$$Ra_k \cap \sum_{j=1}^{k-1} Ra_j = 0,$$

since otherwise, taking $0 \neq r_k a_k = \sum_{j=u}^{k-1} r_j a_j$ with $r_u a_u \neq 0$, we get

$$0 = r_k a_k a_u = \sum_{j=u}^{k-1} r_j a_j a_u = r_u a_u^2,$$

so $r_u a_u \in \ell(a_u) \cap Ra_u = 0$ by Lemma 16.27, a contradiction.

In particular, $Ra_k \not\subseteq \sum_{j=1}^{k-1} Ra_j$, so $Ra_1 \subset Ra_1 + Ra_2 \subset \cdots$. Since R is left Noetherian, this procedure must terminate at some stage k, i.e.,

$$0 = A_k = L \cap \ell(a_1) \cap \cdots \cap \ell(a_k),$$

implying that $\ell(a_1) \cap \cdots \cap \ell(a_k) = 0$.

Let $s = a_1 + \cdots + a_k \in L$. We claim that $\ell(s) = 0$. Indeed if $rs = 0$, then

$$ra_k = -\sum_{j=1}^{k-1} ra_j \in Ra_k \cap \sum_{j=1}^{k-1} Ra_j = 0;$$

inductively we see that $ra_i = 0$ for each $i \leq k$, implying that $r \in \bigcap_{i=1}^{k} \ell(a_i) = 0$.

Hence, s is regular by Proposition 16.28. □

COROLLARY 16.29′. *Any prime left Noetherian ring R has a simple Artinian left ring of fractions.*

Proof. By the theorem, R has a semisimple left ring of fractions Q, which we claim is prime. Indeed if $A, B \lhd Q$ with $AB = 0$, then $(A \cap R)(B \cap R) = 0$, implying that $A \cap R = 0$ or $B \cap R = 0$. Since $R <_{e\ R} Q$ by Lemma 16.21(i), we get $A = 0$ or $B = 0$, as desired.

Hence Q is prime semisimple and thus simple Artinian. □

Goldie's Theorem was certainly a theorem whose time had come. Within five years it was applied in a myriad of ways, and in retrospect could be seen in broader contexts. For example, Johnson defined the **singular ideal** of a ring, which is 0 for most interesting classes of rings (including left Noetherian rings, of course), and for these rings Utumi constructed a generalized left ring of fractions, given in Exercise 23. Gabriel [**Ga1**] discovered a general theory of localization, formulated via general category theory, which led to a noncommutative generalization of the foundation of algebraic geometry.

Unfortunately, although one can say a fair amount about the structure of quotient rings (for example, Utumi's quotient ring is "von Neumann regular," described in Exercises 19–21), general quotient rings lack certain basic ring-theoretic properties of semisimple rings.

Applications to left Noetherian rings

Now that we have Goldie's lovely theorems in hand, what can we use them for? We pause to provide two quick but striking applications to the theory of Noetherian rings. The idea is to use Goldie's Theorem to transfer information from left Artinian rings to semiprime left Noetherian rings, and then to generalize to arbitrary left Noetherian rings.

Nil subsets of Noetherian rings.

First we show that nil subsets often are nilpotent. We start with the lower nilradical $N(R)$, cf. Definition 16.7, which we saw earlier is a nil ideal containing every nilpotent ideal. Thus, letting $N_1(R)$ denote the sum of all nilpotent ideals, we have $N_1(R) \subseteq N(R)$.

Remark 16.30. (i) Any finite sum of nilpotent ideals is a nilpotent ideal (although $N_1(R)$ need not be nilpotent). More precisely, if $I_j^{m_j} = 0$ for $1 \leq j \leq t$, then $(I_1 + \ldots + I_t)^{m_1+\ldots+m_t} = 0$, since any string of $m_1 + \ldots + m_t$ elements in the product must contain at least m_j elements from I_j for some j.

(ii) For R left Noetherian, $N(R) = N_1(R)$ and is nilpotent. (Indeed, first we note that since R is left Noetherian, $N_1(R)$ is a *finite* sum of nilpotent ideals and thus is nilpotent by (i). Next we claim that $R/N_1(R)$ is semiprime. Indeed, if an ideal $A \supseteq N_1(R)$ and $A^2 \subseteq N_1(R)$, then A is nilpotent, implying that $A = N_1(R)$ since $N_1(R)$ is nilpotent. Now, by Proposition 16.26, $R/N_1(R)$ has no nonzero nil ideals, so we conclude that $N(R)/N_1(R) = 0$.)

Now we can improve Theorem 15.23.

THEOREM 16.31. *Suppose R is left Noetherian and S is a nil subset satisfying the condition that for any s_1, s_2 in S there is $\nu = \nu(s_1, s_2) \in \mathbb{Z}$ such that $s_1 s_2 + \nu s_2 s_1 \in S$. Then S is nilpotent.*

Proof. Since $N(R)$ is nilpotent, we may pass to $R/N(R)$ and assume that R is semiprime and left Noetherian. But then R is a subset of its semisimple left ring of fractions, so we are done by Theorem 15.23. □

When R is not left Noetherian, the situation can be much more intricate. The sum of two nil ideals still is a nil ideal, by Exercise 6, but we still do not know whether the sum of two nil left ideals must be nil. (This is one of the formulations of Koethe's question, noted earlier.)

Invariant base number.

A ring R has **invariant base number**, or IBN, if any isomorphism $R^{(m)} \cong R^{(n)}$ as R-modules implies $m = n$. Any commutative ring has IBN by Theorem 2.31 of Volume 1; its proof relied on the following observation.

Remark 16.32. R lacks IBN iff there is an $n \times m$ matrix A and an $m \times n$ matrix B, for $m < n$, such that $AB = I_n$ and $BA = I_m$.

PROPOSITION 16.33. *If there is a ring homomorphism $\varphi: R \to T$ where T has IBN, then R has IBN.*

Proof. Otherwise, there are matrices A, B over R as in Remark 16.32, and applying φ yields matrices over T having the same properties, contrary to T satisfying IBN. □

PROPOSITION 16.34. *Any semisimple ring R has IBN.*

Proof. R has some composition length t, which then implies $\ell(R^{(n)}) = nt$. If $R^{(n)} \cong R^{(m)}$, then $nt = \ell(R^{(n)}) = \ell(R^{(m)}) = mt$, implying $m = n$. □

THEOREM 16.35. *Any left Noetherian ring R has IBN.*

Proof. The semiprime left Noetherian ring $R/N(R)$ is a subring of a semisimple ring, so has IBN; hence R has IBN, by Proposition 16.33. □

On the other hand, Exercise 21 gives an example of a ring lacking IBN.

The representation theory of rings and algebras: An introduction

We say a ring R is **embedded** into a ring T if there is an injection from R into T. In the proofs of Theorems 16.31 and 16.35, we did not use the full force of Goldie's Theorem, but only the fact that a semiprime left Noetherian ring can be embedded into a semisimple ring. (Another application of such embedding arguments is given in Exercise 17.) This raises the question in general of studying rings in terms of their homomorphisms to nicer rings, especially simple Artinian rings, leading us back to representations (Definition 13.38), one of the most sublime concepts in algebra.

Normally, when considering a representation $\Phi: R \to \operatorname{End} M_W$, we would like M to be a free module and W to be "nice." If W is a field F and M is a vector space over F of some finite dimension n, then $\operatorname{End} M_F \cong M_n(F)$, and we can study R in terms of properties of matrices (such as trace and determinant).

Example 16.36. The regular representation of an F-algebra R is an embedding of R into $\operatorname{End}_F R$.

Finite-dimensional representations of various algebraic structures comprise the main theme of Part V; algebras with faithful f.d. representations are to be studied in depth in Chapter 23.

One useful way of obtaining useful representations is by means of Goldie's Theorem, which shows that a prime left Noetherian ring R is embedded into its simple Artinian ring of fractions $Q(R)$. Accordingly, any prime ideal P of R gives us the representation $R \to Q(R/P)$ with kernel P, and left Noetherian rings often are studied in terms of their prime ideals (again, motivated by the commutative case, in the geometric language of the prime spectrum). This approach has been generalized to categorical techniques in localization.

In general, representations are closely tied in with module theory via the following basic observation, to be utilized repeatedly in Chapter 19.

Remark 16.37. If $\rho\colon R \to \operatorname{End} M_W$ is a representation of the ring R, then M becomes an R-module under the action

$$ra = \rho(r)(a), \quad \forall r \in R,\ a \in M.$$

Conversely, for any C-algebra R, any R-module structure on M gives rise to a representation $\hat{\rho}\colon R \to \operatorname{End}_C M$, defined by

$$\hat{\rho}(r)(a) = ra.$$

Thus, in one sense, the goal of representation theory might be to characterize all R-modules.

Remark 16.38. Suppose the ring R is semisimple. By Proposition 14.23, any R-module is semisimple, and thus a direct sum of simple modules, by Exercise 14.4, which are described precisely in Corollary 15.10.

From this point of view, for left Artinian rings, the obstruction to understanding the representation theory is the Jacobson radical.

Digression 16.39. For arbitrary rings, the starting point is Schur's Lemma (Proposition 13.40), which shows that the endomorphism ring of a simple module is a division ring. (More generally, Theorem 13.47 shows that the endomorphism ring of a semisimple Artinian module is a direct product of matrix rings.) Thus, each simple R-module gives rise to a representation of R, and the determination of the simple R-modules is the first step in representation theory. Since the simple R-modules have the form R/L, where L is a maximal left ideal, we could study instead the maximal left ideals of a ring.

Again, we are confronted with the Jacobson radical, which is the intersection of the maximal left ideals and lies in the kernel of each such representation.

Even when $\operatorname{Jac}(R) = 0$, we need to know how to piece together simple modules, a very delicate issue which is outside the scope of this book.

Indecomposable modules.

One natural strategy to classify modules was already outlined in Remark 2.9 of Volume 1. Recall that a module is **indecomposable** if it cannot be written as a direct sum of two nonzero submodules; we want to classify the indecomposable submodules of a ring, and write a module "uniquely" as a direct sum of indecomposable submodules.

LEMMA 16.40. *Every Noetherian module M is a finite direct sum of indecomposable modules.*

Proof. By Noetherian induction, we may assume that the assertion holds for M/N for every nonzero submodule N of M. But the assertion is immediate unless M is decomposable, i.e., $M = K \oplus N$. The assertion holds for $N \cong M/K$ and for $K \cong M/N$, and thus for $N \oplus K$. □

By Proposition 7.10 of Volume 1, the assertion of Lemma 16.40 holds for any f.g. module over a left Noetherian ring R, so in this case our quest reduces to finding the f.g. indecomposable modules (up to isomorphism). First we want to know how many we need to find.

Definition 16.41. A ring R has **finite representation type** (f.r.t.) if R has only finitely many f.g. indecomposable modules, up to isomorphism.

Example 16.42. Every semisimple ring has f.r.t., by Remark 16.38.

One important source of examples is to be provided by Maschke's Theorem (Theorem 19.26), which says that any group algebra of a finite group over a field of characteristic 0 is semisimple and thus has finite representation type.

When R is not semisimple, the theory becomes much more subtle, even for f.g. modules. For example, in characteristic 2, the group algebra of the Klein group is an example of a four-dimensional algebra lacking f.r.t., as to be seen in Exercise 19.8. This crucial issue of f.r.t. is discussed further in Appendix 25C.

We are also interested in the uniqueness of the decomposition into indecomposables, and some information can be obtained by means of general ring theory.

Remark 16.43. Any direct summand of a module M has the form $\pi(M)$, where the projection π is an idempotent of $\operatorname{End}_R M$. Thus, decomposability of a module $M = M_1 \oplus M_2$ is equivalent to finding a nontrivial idempotent in $\operatorname{End}_R M$.

On the other hand, the Jacobson radical cannot contain nontrivial idempotents (cf. Exercise 15.29). We call a ring R **local** if $R/\operatorname{Jac}(R)$ is a division ring; cf. Exercises 15.32–15.35. It follows at once that a local ring cannot contain nontrivial idempotents. Consequently, if $\operatorname{End}_R M$ is local, then M is indecomposable.

Accordingly, we say that a module M is an **LE-module** if $\operatorname{End}_R M$ is a local ring. Direct sum decompositions into LE-modules are unique in a rather strong sense; cf. Exercise 29, usually known as the **Krull-Schmidt Theorem**. Thus, we would like to know which indecomposable modules are LE. Fortunately, this is true for all modules of finite composition length. Let us see this by means of **Fitting's Lemma**, a more general version of Remark 2.67 of Volume 1.

Remark 16.44. Given a map $f\colon M \to N$ and submodule $N \leq M$ with $f(N) = f(M)$, we have $M = N + \ker f$. (Indeed, for any $a \in M$, we write $f(a) = f(b)$ for $b \in N$, and note that $a - b \in \ker f$.)

LEMMA 16.45. *For any Noetherian module M, every onto map $f\colon M \to M$ is an isomorphism.*

Proof. $\ker f^n = \ker f^{n+1}$ for some n. If $a \in \ker f$, then $a = f^n(b)$ for some $b \in M$. $f^{n+1}(b) = f(a) = 0$, implying that $b \in \ker f^{n+1} = \ker f^n$, so $a = f^n(b) = 0$. □

PROPOSITION 16.46 (FITTING'S LEMMA). *If M has finite composition length n, then $M = f^n(M) \oplus \ker f^n$ for any map $f\colon M \to M$; furthermore, f restricts to an isomorphism on $f^n(M)$ and a nilpotent map on $\ker f^n$.*

Proof. Since the chain $M \geq f(M) \geq f^2(M) \geq \cdots$ has length $\leq n$, we have $f^i(M) = f^{i+1}(M)$ for some $i \leq n$, implying that $f^n(M) = f^{n+1}(M)$. Hence $f^n(M) = f^n(f^n(M))$, so $M = f^n(M) + \ker f^n$ by Remark 16.44.

But likewise the chain $0 \leq \ker f \leq \ker f^2 \leq \cdots$ has length $\leq n$, implying that $\ker f^{n+1} = \ker f^n$ and thus $\ker f^{2n} = \ker f^n$. If $a \in f^n(M) \cap \ker f^n$, then writing $a = f^n(b)$ we get $b \in \ker f^{2n} = \ker f^n$, implying $a = 0$. Thus, $M = f^n(M) \oplus \ker f^n$. Then $f\colon f^n(M) \to f^{n+1}(M) = f^n(M)$ is onto, and thus is an isomorphism by Lemma 16.45. The last assertion now is obvious. □

COROLLARY 16.47. *If M is indecomposable of finite composition length, then every map $f\colon M \to M$ is either an isomorphism or nilpotent. In particular, M is an LE-module.*

Proof. The first assertion is obvious and implies that the nilpotent elements of $\operatorname{End}_R M$ form the unique maximal ideal (cf. Exercise 34); hence, $\operatorname{End}_R M$ is local. □

Fitting's Lemma was extended by Harada and Sai; cf. Pierce [**Pie**].

Example 16.48. Suppose R is any left Artinian ring. Then R is also Noetherian, by Theorem 15.21, so every f.g. module over R is both Artinian and Noetherian and thus has a composition series.

In particular, suppose R is a f.d. algebra over a field F. Then we can write R as a direct sum of indecomposable R-modules $M_1 \oplus \cdots \oplus M_t$, and one can consider the directed graph whose vertices are indexed by the M_i and whose edges from i to j correspond to module homomorphisms from M_i to M_j. The precise definition is given in Definition 25C.5 and is one of the basic tools in the representation theory of algebras. The last generation of research has seen startling advances from the perspective of category theory, which we discuss briefly in Appendix 25C.

Representations into left Artinian rings.

Having been sidetracked somewhat by the difficult question of classifying all representations of a ring, let us return briefly to ask about "good" representations, say into simple Artinian rings. Mal′cev provided a domain that cannot be embedded into a division ring (cf. Exercise 18), leading P. M. Cohn and his school to characterize domains embeddible into division rings. This line of investigation culminated in an elegant theory of Schofield [**Scho, Theorem 7.11**], characterizing representations into Artinian algebras; using his methods, for example, one can find a domain not embeddible into a division ring but that can be embedded into $M_2(D)$ over a division ring D.

Supplement: Graded and filtered algebras

Gradings and filtrations of algebras (not necessarily commutative) were introduced in Volume 1 in Definition 7.20ff., and commutative graded algebras were seen in Chapter 10 (preceding Definition 10.33) to be the "correct" algebraic structure for studying projective geometry. Since commutative graded algebras play such an important role, researchers have been motivated to generalize the noncommutative structure theory to the graded case.

We assume in this discussion that R is a given algebra graded over an Abelian group $(\mathcal{G}, +)$. (This is a reasonably general situation, since if R is graded over a submonoid $\mathcal{M}$ of $\mathcal{G}$, one could extend the grade to all of $\mathcal{G}$ by putting $M_g = 0$ for all $g \in \mathcal{G} \setminus \mathcal{M}$.)

Remark 16.49. If R is graded by an Abelian group $\mathcal{G}$ and $N < \mathcal{G}$, then R is also $\mathcal{G}/N$-graded by putting $R_{Ng} = \bigoplus_{a \in N} R_{ag}$. For example, if R is $\mathbb{Z}$-graded, then R is $\mathbb{Z}/2$-graded; and if I is any ideal generated by even elements, then R/I is also $\mathbb{Z}/2$-graded.

Physicists, who often have a better sense of drama than mathematicians, have renamed $\mathbb{Z}/2$-graded algebras "superalgebras"; i.e., $R = R_0 \oplus R_1$ where R_0 is a subring of R and R_1 is an R_0-module such that $R_1^2 \subseteq R_0$. Superalgebras arise when physicists combine two related phenomena, such as symmetry and antisymmetry.

Example 16.50. $W = M_n(R)$ is $\mathbb{Z}$-graded, where $W_n = \sum_{i+j=n} Re_{ij}$. This induces a $\mathbb{Z}/2$ -grading, where e_{ij} is even (resp. odd) if $i + j$ is even (resp. odd).

Example 16.51. In studying graded rings, one gets guidance by working in the appropriate category, i.e., introducing the grading into every appropriate part of the theory. Having the basic concepts at our disposal, we could embark along the path of Chapters 14, 15, and 16 to obtain graded versions of the basic structure theorems. Since this is rather straightforward for the most part, we leave it for Exercises 31–41, albeit one encounters a surprise when extending Goldie's Theorem; cf. Exercises 36 and 38. The grade has useful applications even in the nongraded theory; cf. Exercises 39–41.

Arbitrary (nongraded) algebras often have interesting natural filtrations (cf. Definition 7.29 of Volume 1) and thereby can be studied in terms of their associated graded algebras.

Appendix 16A: Deformations and quantum algebras

One way in which filtered algebras arise quite naturally is in the theory of deformations. This subject has evolved considerably since its inception by Gerstenhaber.

Definition 16A.0. A **deformation** of an associative F-algebra R was originally a new multiplication on the power series algebra $R[[t]]$ such that t remains a central indeterminate, and specializing $t \mapsto 0$ yields the original

algebra R. (Later, one only required that the new algebra is also filtered from below by the powers of t.)

We write the multiplication as

$$ab = \sum \mu_i(a,b)t^i, \quad \text{for} \quad a, b \in R, \tag{16A.1}$$

where $\mu_i \colon R \times R \to R$ is a bilinear map. The fact that $R[[t]]$ is associative translates to the formulas

$$\sum_{i+j=m} \mu_j(\mu_i(a,b),c) = \sum_{i+j=m} \mu_j(a,\mu_i(b,c)) \tag{16A.2}$$

for all $a, b, c \in R$. Since by hypothesis μ_0 is multiplication in R, Equation (16A.2) becomes

$$\begin{aligned} \sum_{\substack{i+j=m \\ i,j>0}} \mu_j(\mu_i(a,b),c) - \mu_j(a,\mu_i(b,c)) \\ = a\mu_m(b,c) - \mu_m(ab,c) + \mu_m(a,bc) - \mu_m(a,b)c. \end{aligned} \tag{16A.3}$$

This is interpreted homologically in Chapter 25; also see Exercise A1.

It has become more convenient to consider instead a new multiplication on $\hat{R} = R[t, t^{-1}]$, the localization of $R[t]$ at the powers of t, where specializing $t \mapsto 1$ yields the original algebra R. Although this algebra loses the natural $\mathbb{N}$-filtration, it permits more flexibility in computation.

Definition 16A.1. The **quantization** R_q of an algebra R is a specialization of t in $\hat{R}$ to some other element q of F. (The notation "q" has become standard, although occasionally q itself is used to denote the indeterminate.) There are two major cases:

(i) The **degenerate** case, for which q is a root of 1; then the deformed algebra is finite-dimensional when R is f.d.

(ii) The **nondegenerate** case, for which q is not a root of 1.

Often R is taken to be the (commutative) coordinate algebra of an affine variety V, which we designate here as $\mathcal{O}(V)$. (In Volume 1 we used the notation $F[V]$, but we switch notation here in order to avoid clashing with the group algebra notation.) For example, $\mathcal{O}(F^{(2)})$ denotes the coordinate algebra of the plane, which is the polynomial algebra $F[\lambda_1, \lambda_2]$.

Example 16A.2. We can deform $\mathcal{O}(F^{(2)}) = F[\lambda_1, \lambda_2]$ to $F[\lambda_1, \lambda_2, t, t^{-1}]$, where $\lambda_2\lambda_1 = t\lambda_1\lambda_2$ (and thus $\lambda_1\lambda_2 = t^{-1}\lambda_2\lambda_1$). Specializing $t \mapsto 1$ gives the usual polynomial algebra.

The **quantum coordinate algebra** $\mathcal{O}_q(F^{(2)})$ of the plane (called the **quantum plane** for short) is defined by specializing $t \mapsto q$ for $q \notin \{0, \pm 1\}$. Note in the terminology of Appendix 13A that this construction is always a skew polynomial algebra $T[\lambda_2; \sigma]$, where $T = F[\lambda_1]$ and $\sigma(\lambda_1) = q\lambda_1$; in the degenerate case, the automorphism σ has finite order.

Example 16A.3. Identifying $M_2(F)$ with $F^{(4)}$, we have the **quantized matrix algebra**

$$\mathcal{O}_q(M_2(F)) = F\{x_{11}, x_{12}, x_{21}, x_{22}\}/\langle I \rangle,$$

where I is the ideal generated by

$$x_{11}x_{12} - qx_{12}x_{11},\ x_{11}x_{21} - qx_{21}x_{11},\ x_{12}x_{21} - x_{21}x_{12},\ x_{12}x_{22} - qx_{22}x_{12},$$
$$x_{21}x_{22} - qx_{22}x_{21}, \quad \text{and} \quad x_{11}x_{22} - x_{22}x_{11} - (q - q^{-1})x_{12}x_{21}.$$

The **quantum determinant** $\delta_q = x_{11}x_{22} - qx_{12}x_{21} = x_{22}x_{11} - q^{-1}x_{12}x_{21}$.

It is easy to see that $\delta_q \in \mathrm{Cent}(\mathcal{O}_q(M_2(F)))$. We define

$$\mathcal{O}_q(\mathrm{GL}(2, F)) = \mathcal{O}_q(M_2(F))[\delta_q^{-1}];$$

$$\mathcal{O}_q(\mathrm{SL}(2, F)) = \mathcal{O}_q(M_2(F))/\langle \delta_q - 1 \rangle.$$

Noting that

$$\mathcal{O}_q(\mathrm{SL}(2, F)) \cong \mathcal{O}_q(M_2(F))/\langle x_{11}x_{22} - qx_{12}x_{21} - 1 \rangle,$$

we see that $\mathcal{O}_q(\mathrm{GL}(2, F))$ (resp. $\mathcal{O}_q(\mathrm{SL}(2, F))$) is indeed a deformation of $\mathcal{O}(\mathrm{SL}(2, F))$ (resp. $\mathcal{O}(\mathrm{GL}(2, F))$).

More examples are given in Exercises A2–A6 and in Definition 21C.7 and Exercises 21C.17–21C.21. When the algebra R of Definition 16A.1 is commutative, as in Examples 16A.2 and 16A.3, one can study its quantizations by means of extra structure, as indicated in Exercise 21.86. Deformation theory, especially with reference to quantizing, has become a very active area of research, and we return to it in Chapter 26.

Chapter 17

Algebras in Terms of Generators and Relations

In Volume 1, our main interest in commutative algebra was in affine algebras, to prepare the groundwork for affine geometry. Here we consider the noncommutative analog.

Definition 17.1. An **affine algebra over a field** F is an F-algebra R, generated by some finite set $S = \{a_1, \dots, a_n\}$ in the sense that the only F-subalgebra of R containing S is R itself; in this situation, we denote R as $F\{a_1, \dots, a_n\}$.

(In the literature, affine algebras often are defined more generally over a commutative Noetherian ring, since much of the theory remains valid in this more general situation.)

The structure theory of commutative affine algebras (involving the prime spectrum) has been spectacularly successful, leading to the following results proved in Volume 1:

1. Each (commutative) affine algebra is Noetherian, by the Hilbert Basis Theorem. Thus, the powerful theory of commutative Noetherian rings is applicable to affine algebras.
2. Every prime ideal is the intersection of maximal ideals, by Proposition 6.37. It follows that the intersection of all maximal ideals is a nilpotent ideal.

3. The (classical) Krull dimension exists, and equals the transcendence degree, by Theorem 6.34.
4. (Catenarity) Any saturated chains between two given prime ideals have the same length, by Theorem 9.11.

We also recall that any commutative affine algebra R is a homomorphic image of a (commutative) polynomial algebra; i.e., $R = F[\lambda_1, \dots, \lambda_n]/\mathcal{I}$, where $\mathcal{I} \lhd F[\lambda_1, \dots, \lambda_n]$, and the ideal $\mathcal{I}$ is finitely generated by Theorem 7.18. Thus, R can be described in terms of finitely many generators and relations, and the structure of the polynomial algebra "contains" all of the theory of affine algebras. This approach led us in Volume 1 to Gröbner bases and the ensuing combinatoric theory.

Unfortunately, all of the results listed above can fail for noncommutative affine algebras. Affine algebras need not be Noetherian, but rather can be just about as general as one could imagine, as we shall see. Perhaps it is precisely their initial intractability that has challenged many excellent mathematicians to study them. Over the past half-century, techniques have been developed to deal with noncommutative affine (as well as nonaffine) algebras — some of them are presented here, whereas others are beyond the scope of this book.

In this chapter, we describe algebraic structures in terms of generators and relations, and consider some basic issues arising from this description including Cayley graphs and growth of algebras and groups. Finer aspects of the theory are dealt with in the appendices and their exercises.

Groups had been studied from this vantage point even earlier than algebras. Burnside posed his celebrated problem (considered in Appendix 17C) over 100 years ago, followed by the work of Dehn [**Deh**] described in Appendix 17B. Soon thereafter, Nielson [**Nie2**] and Schreier obtained deep results concerning presentations of infinite groups, including the surprising theorem that any subgroup of a free group is free; these ideas were developed in an important paper of M. Hall and Tibor [**HalmT**]. Advanced combinatorial techniques were developed in the 1960's and 1970's to provide counterexamples to a host of questions about infinite groups, including the Burnside Problem. More recently, it has been seen that graphs and topology provide a way of formulating and analyzing the combinatoric issues involved.

Free algebraic structures

To understand algebraic structures in terms of generators and relations, we must first define an algebraic structure without any relations at all; such a structure is called **free**. This was described in Chapter 0 of Volume 1 as being that structure $\mathcal{F}$ (depending on a given index set I) possessing a base

$B = \{b_i : i \in I\}$ such that, for any structure A of the same type and for any subset $S = \{s_i : i \in I\}$ of A, there is a unique homomorphism $\phi: \mathcal{F} \mapsto A$ sending $b_i \mapsto s_i$ for each i. Taking I large enough to enable S to contain a set of generators of A, we may assume that $\phi(\mathcal{F})$ generates A (so ϕ is onto); ϕ is called a **presentation** of A, and $\ker \phi$ is called the set of **relations** (in this presentation).

A general construction of free objects in universal algebra was given in Exercise 0.11 of Volume 1, which could be used in any particular case that comes up, but it is instructive to construct explicit free objects for the most important categories. Of course, the specific construction used in a given category is not crucial, since the free object (having base of a given cardinality) is unique up to isomorphism by "abstract nonsense"; cf. Exercise 5. Here are some examples.

In Chapter 2 of Volume 1, we constructed the **free R-module** with base $\mathcal{M}$ to be the direct sum of copies of R, comprised of formal sums

$$\Big\{ \sum_{w \in \mathcal{M}} r_w w : \text{ almost all } r_w = 0 \Big\},$$

where addition and scalar multiplication are defined componentwise.

Since Abelian groups are just $\mathbb{Z}$-modules, the **free Abelian group** thus is a direct sum of copies of $(\mathbb{Z}, +)$.

Also, the **free commutative algebra** is the polynomial algebra, by Proposition 5.7 of Volume 1.

The free monoid.

We turn to the free monoid, which is needed in other constructions.

Definition 17.2. Let $X = \{x_i : i \in I\}$. A **word** is a string of the x_i, such as $x_1x_2x_1x_1x_4$. The **free (associative) monoid** $\mathcal{M}$ on X is the set of words in the x_i, including the "blank word" denoted 1; multiplication on $\mathcal{M}$ is given by juxtaposition of words; i.e.,

$$x_1x_2x_1 \cdot x_1x_1x_4x_2 = x_1x_2x_1x_1x_1x_4x_2.$$

We replace contiguous occurrences of x_i by the appropriate power, so this word is rewritten as $x_1x_2x_1^3x_4x_2$. Obviously, multiplication by juxtaposition is associative, and 1 is the neutral element.

The **length** $|w|$ of a word $w = x_{i_1} \cdots x_{i_t}$ equals t, and formally $|1| = 0$.

Remark 17.3. The length defines a monoid homomorphism $\mathcal{M} \to (\mathbb{N}, +)$; i.e.,

$$|wv| = |w| + |v|, \quad \forall w, v \in \mathcal{M}. \tag{17.1}$$

Thus, the only invertible element of $\mathcal{M}$ is the blank word 1.

Remark 17.4. For any monoid M and any subset $\{a_i : i \in I\}$ of M, the map $x_i \mapsto a_i$ extends naturally to a unique monoid homomorphism $\mathcal{M} \to M$ given by $x_{i_1} \cdots x_{i_t} \mapsto a_{i_1} \cdots a_{i_t}$; hence, $\mathcal{M}$ is indeed the free monoid.

The free algebra.

We turn to the free objects of primary interest for us, starting with the free associative algebra over a commutative ring C. We must combine the free module with the free monoid.

Definition 17.5. The **free (associative) C-algebra** $C\{X\}$ with respect to a generating set $X = \{x_i : i \in I\}$ is the free C-module with base $\mathcal{M}$ (of Definition 17.2), together with multiplication

$$\Big(\sum_{w \in \mathcal{M}} c_w w \Big) \Big(\sum_{v \in \mathcal{M}} c'_v v \Big) = \sum_{u \in \mathcal{M}} c''_u u,$$

where $c''_u = \sum_{wv=u} c_w c'_v$. (Almost all of the c''_u are 0, since by definition almost all of the c_w and c'_v are zero.) The elements of $C\{X\}$ are called **(noncommutative) polynomials**.

$C\{x_1, \ldots, x_n\}$ denotes the free (affine) C-algebra with respect to the set $X = \{x_1, \ldots, x_n\}$.

$C\{X\}$ is readily seen to be a C-algebra; associativity is obtained either by direct verification or by means of the more general construction of Exercises 1 and 2.

Remark 17.6. For any C-algebra R and any function $\psi: X \to R$, we have a unique homomorphism $\phi: C\{X\} \to R$ such that $\phi(x_i) = \psi(x_i)$; namely, letting $r_i = \psi(x_i)$, we define ϕ first on words by

$$\phi(x_{i_1}x_{i_2} \cdots x_{i_t}) = r_{i_1} r_{i_2} \cdots r_{i_t},$$

and then put $\phi(\sum_w c_w w) = \sum c_w \phi(w)$.

This displays $C\{X\}$ as the free algebra with respect to the set X. In particular, when $\{r_i : i \in I\}$ generate R, ϕ is onto, so $R \cong C\{X\}/\ker \phi$.

The downside of all of this is that the study of free algebras (having arbitrarily large generating sets) contains all of the theory of associative algebras.

Remark 17.7. For $I = \{1\}$, $C\{x_1\}$ is isomorphic to the commutative polynomial algebra $C[\lambda]$. However, $C\{x_1, x_2\}$ is already large enough to contain the free algebra on any countable set; cf. Exercise 6.

Nevertheless, the free algebra can be studied, via words.

Definition 17.8. Assume that C is an integral domain.

(i) For any polynomial $f = \sum c_w w$, the $c_w w$ are called the **monomials** of f. The **degree** of a monomial $c_w w$ is defined to be $|w|$; the (total) **degree** of f is the maximum degree of its nonzero monomials. Thus, the free associative algebra $C\{X\}$ is $\mathbb{N}$-graded via the degree. To pinpoint the nonzero monomials of f, we repeat Definition 7B.1 in the noncommutative setting and define the **support**

$$\operatorname{supp}(f) = \{w \in \mathcal{M} : c_w \neq 0\}.$$

A polynomial f is **homogeneous** if each monomial in the support of f has the same degree.

(ii) Refining (i), $C\{x_1, \dots, x_n\}$ is $\mathbb{Z}^{(n)}$-graded via the monomials as follows: For a monomial $h(x_1, \dots, x_n)$, define $d_i = \deg_i h$ to be the number of times x_i appears in h; thus $\sum_{i=1}^n d_i$ is the degree of h.

We want some sort of lexicographic order on the free monoid, but the usual lexicographic order does not preserve multiplication; for example, if we take $x_2 > x_1$, then $x_2 x_1 > x_2$ but $(x_2 x_1)^2 = x_2 x_1 x_2 x_1 < x_2^2$. This leads us to the following modification:

Definition 17.9. The **crossword dictionary order** on $\mathcal{M}$ is defined by induction on length; we say that words $w < v$ if any of the following criteria are met:

1. $|w| < |v|$.
2. $|w| = |v|$ and the first letter of w is less than the first letter of v.
3. $|w| = |v|$ and $w = x_i w'$, $v = x_i v'$, where inductively $w' < v'$.

Obviously, $<$ is a total order on $\mathcal{M}$.

Lemma 17.10.

(i) *The crossword dictionary order* $(<)$ *satisfies the minimum condition.*

(ii) $(\mathcal{M}, <)$ *is an ordered monoid.*

Proof. (i) We want to show that any $S \subset \mathcal{M}$ has a minimal element. Take $w \in S$ with $m = |w|$ minimal. Suppose x_j is the first letter of w, and we pick $w \in S$ such that j is minimal. Consider $S' = \{v \in \mathcal{M} : x_j v \in S\}$. By induction on m, we can find a smallest element $v \in S'$. But then $x_j v$ must be minimal in S.

(ii) We need to verify $ac > bc$ and $ca > cb$ for any words $a > b$ and c. This is obvious if $|a| > |b|$, and for $|a| = |b|$ we apply criterion (2) of the definition and/or induction. □

Since the crossword dictionary order satisfies the minimum condition, one can use it for proofs by induction on the order. Let us apply these considerations to $C\{X\}$. The **leading monomial** of $f \in C\{X\}$ is that monomial $c_w w$ for which w is maximal in $\operatorname{supp}(f)$ with respect to the crossword dictionary order.

Remark 17.11. Suppose C is an integral domain. By Lemma 17.10, the product of the leading monomials of polynomials f and g is the leading monomial of fg. It follows that $C\{X\}$ is a (noncommutative) domain.

At the same time, the structure of the free algebra already tells us something in general, and gives us a nice "dichotomy" result. Obviously, the free algebra $C\{x_1, x_2\}$ on two generators is not an Ore domain, since x_1 and x_2 have no common left multiple. Conversely, we have:

Proposition 17.12. *Any domain R (over C) is either an Ore domain or contains a free algebra on two generators.*

Proof. We are done if $C\{a, b\}$ is free for suitable $a, b \in R$. Thus, for any $0 \neq a, b \in R$ we may assume that we have a relation $f(a, b) = 0$, for some polynomial $f(x_1, x_2) \neq 0$ of smallest possible degree. But writing $f = g(x_1, x_2)x_1 + h(x_1, x_2)x_2$, we see that $h(x_1, x_2) \neq 0$ since otherwise $g(a, b)a = 0$, implying $g(a, b) = 0$, contrary to minimality of $\deg f$. By induction, $h(a, b) \neq 0$, implying that $0 \neq h(a, b)b = -g(a, b)a \in Ra \cap Rb$, verifying the Ore condition. □

Remark 17.13. Perhaps surprisingly, despite Exercise 16.18, the free algebra over any field can be embedded in a division algebra. There are several proofs of this fact, one of which (given in Exercise 24) relies on the celebrated Magnus-Witt Theorem, that the free group is ordered.

In Exercise 7 we see that the free algebra is also primitive as a ring.

Although, as we have seen, any algebra R generated by $|I|$ elements is a homomorphic image of the free algebra on $\{x_i : i \in I\}$, the set of relations may be too unwieldy to yield much information about the algebra. Accordingly, let us define a more manageable class of algebras.

Definition 17.14. An algebra $R = C\{X\}/A$ is a **monomial algebra** if A is generated by monomials.

For example, $F[\lambda]/\langle\lambda^2\rangle$ is a monomial algebra. Monomial algebras might seem very special, but as we shall see, they are varied enough to enable us to study various general properties of algebras.

The free group

The free group is trickier to construct than the free algebra, because it requires multiplicative inverses. Although the free group (on a set of given cardinality) must be unique up to isomorphism, there are several ways of constructing it, each having its own advantages. First we view it as a monoid.

Example 17.15 (The free group). Intuitively, we want to create inverses for the elements of the free monoid. To achieve this goal, we start with the free monoid $\mathcal{M}$ in a larger alphabet $\{x_i, y_i : i \in I\}$. We make $y_i = x_i^{-1}$ by defining an equivalence of words, saying that two words w_1 and w_2 are equivalent if we can pass from one to the other by inserting or deleting various x_iy_i or y_ix_i. (Thus, x_iy_i and y_ix_i are equivalent to 1.) It is easy to see that the equivalence classes comprise a group $\mathcal{G}$, where the class of y_i is the inverse of the class of x_i. So we write x_i^{-1} instead of y_i, and denote the free group $\mathcal{G}$ as the monoid generated by $\{x_i, x_i^{-1} : i \in I\}$, satisfying the relations $\{x_ix_i^{-1} = 1 = x_i^{-1}x_i,\ \forall i \in I\}$.

We call $|I|$ the **rank** of the free group $\mathcal{G}$.

Remark 17.15′. The free group of rank 1 is Abelian and thus is isomorphic to $(\mathbb{Z}, +)$. On the other hand, the free group on the countably infinite set $\{x_i : i \in \mathbb{N}\}$ can be embedded in the free group on two generators; cf. Exercise 8.

The construction of Example 17.15 could be described in terms of the structure of monoids, as indicated in Exercise 4, but unfortunately this theory is somewhat complicated, and it is not at all obvious that the natural map from the free monoid $\mathcal{M}$ to the free group $\mathcal{G}$ given by $x_i \mapsto x_i$ is 1:1. This fact will become clear from the precise description in Example 17B.6, utilizing the theory of reduction procedures, but meanwhile there is a simpler way to see it, by means of injecting the free group into a matrix algebra. An ad hoc injection is given in Exercise 9. Here is a systematic method.

For subgroups A, B of a group G, we define an **alternating product** of **length** m to be $a_1b_1a_2b_2\cdots a_mb_m$, where possibly $a_1 = e$ and/or $b_m = e$, but all other $a_i, b_i \neq e$. We say that A and B **interact freely** if no alternating product of length ≥ 2 equals e. (This also implies $a_1b_1 \neq e$ for $a_1 \in A$ and $b_1 \in B$, for otherwise $e = ea_1b_1e$ is an alternating product of length 2, contrary to the definition.)

Lemma 17.16 (The Pingpong Lemma). *Suppose a group G acts on a set S, and $A, B \leq G$. Suppose moreover that S has disjoint subsets Γ_A and Γ_B such that, for all $a \in A \setminus \{e\}$ and $b \in B \setminus \{e\}$,*

$$a\Gamma_B \subseteq \Gamma_A, \qquad b\Gamma_A \subseteq \Gamma_B, \qquad b\Gamma_B \cap \Gamma_B \neq \emptyset.$$

Then A and B interact freely.

Proof. Otherwise, take an alternating product $a_1b_1\cdots a_mb_m$ that equals e, having minimal possible length m. If $a_1 = e$, then $a_2b_2\cdots a_m(b_mb_1^{-1}) = e$, an alternating product of length $m-1$, so we may assume that $a_1 \neq e$. Likewise, if $b_m = e$, then the relation $(a_m^{-1}a_1)b_1\cdots a_{m-1}b_{m-1} = e$ is shorter, so $b_m \neq e$. Taking $c \in b_m\Gamma_B \cap \Gamma_B$ according to hypothesis, we write $c = b_mc'$ for $c' \in \Gamma_B$ and have

$$c' = ec' = a_1b_1\cdots a_mb_mc' = a_1b_1\cdots a_mc \in \Gamma_A,$$

implying that $c' \in \Gamma_A \cap \Gamma_B$, contrary to hypothesis. □

In particular, under the hypotheses of the pingpong lemma, any non-torsion elements $a \in A$ and $b \in B$ must generate a free subgroup.

Example 17.16′. $G = \mathrm{GL}(2, \mathbb{Q})$ has a free subgroup generated by the matrices $a = \left(\begin{smallmatrix} 1 & 0 \\ 3 & 1 \end{smallmatrix}\right)$ and $b = \left(\begin{smallmatrix} 1 & 3 \\ 0 & 1 \end{smallmatrix}\right)$, as seen by taking $A = \langle a\rangle$, $B = \langle b\rangle$, and

$$\Gamma_A = \left\{ \begin{pmatrix} j \\ k \end{pmatrix} : |j| < |k| \right\}, \qquad \Gamma_B = \left\{ \begin{pmatrix} j \\ k \end{pmatrix} : |j| > |k| \right\}.$$

Indeed, $\left(\begin{smallmatrix} 1 & 3 \\ 0 & 1 \end{smallmatrix}\right)^m \left(\begin{smallmatrix} j \\ k \end{smallmatrix}\right) = \left(\begin{smallmatrix} j+3km \\ k \end{smallmatrix}\right)$, and clearly $|j + 3mk| > |k|$ for any $m \neq 0$ when $|j| < |k|$, proving that $b^m\Gamma_A \subseteq \Gamma_B$; likewise $a^m\Gamma_B \subseteq \Gamma_A$ for each $m \neq 0$. Similarly, one checks that $b^m\Gamma_B \cap \Gamma_B \neq \emptyset$ for each $m \neq 0$.

Combinatorics of the Free Group – P. Hall's collecting process.

Recall that (a, b) denotes the **group commutator** $aba^{-1}b^{-1}$. One very successful combinatoric method of studying the free group $\mathcal{G}$ on $\{x_1, \ldots, x_n\}$ is to rewrite its elements as products of (group) commutators of increasing complexity. To prepare this line of thought, we need a few results from group theory.

Definition 17.17. We write b^a for the conjugate aba^{-1}. Thus, the group commutator $(a, b) = b^ab^{-1}$.

Definition 17.17′. For any subgroups H, K of the group G, we write (H, K) for the subgroup generated by $\{(h, k) : \ h \in H,\ k \in K\}$. Thus, the commutator subgroup $G' = (G, G)$. The **lower central series** $\gamma_i = \gamma_i(G)$ is defined as follows: $\gamma_1 = G$, and inductively $\gamma_k = (\gamma_{k-1}, G)$.

We need the following facts concerning group commutators (for all elements $a, b, c \in G$), the latter discovered by P. Hall:

$$(a, bc) = (a, b)(a, c)^b; \qquad (ab, c) = (b, c)^a(a, c) \tag{17.2}$$

$$(a^c, (b, c))(c^b, (a, b))(b^a, (c, a)) = 1. \tag{17.3}$$

Here is one important application.

LEMMA 17.18. *$(\gamma_i, \gamma_j) \le \gamma_{i+j}$, for all i, j.*

Proof. By induction on j; the assertion is trivial for $j = 1$. For general j, every element of γ_j is a product of elements of (γ_{j-1}, G), so, in view of (17.2) it suffices to check that $(\gamma_i, (\gamma_{j-1}, G)) \subseteq \gamma_{i+j}$; i.e, $(a, (b, c)) \in \gamma_{i+j}$ whenever $a \in \gamma_i$, $b \in \gamma_{j-1}$, and $c \in G$. Let $d = a^{c^{-1}} \in \gamma_i$; we can rewrite (17.3) (replacing a by d) as

$$(a, (b, c)) = ((d, b), c^b)((c, d), b^d). \tag{17.4}$$

By induction, $(d, b) \in \gamma_{i+j-1}$, so $((d, b), c^b) \in \gamma_{i+j}$. Furthermore, $(c, d) \in \gamma_{i+1}$ and $b^d \in \gamma_{j-1}$, so by induction, $((c, d), b^d) \in \gamma_{i+j}$. Substituting in (17.4) shows that $(a, (b, c)) \in \gamma_{i+j}$, as desired. □

For our current purpose, we take $G = \mathcal{G}$, the free group, so $\gamma_i = \gamma_i(\mathcal{G})$. We aim toward the result that γ_t/γ_{t+1} is a free Abelian group of finite rank for each t. This is easy for $t = 1$, since the group $\mathcal{G}/\mathcal{G}'$ is clearly free Abelian on the images of the generators $x_1, \dots, x_n$. However, we want a uniform approach that works for all t.

In general, we work in $\mathcal{G}/\gamma_{t+1}$. Our strategy is to rewrite words by means of the relation

$$ab = (a, b)ba, \tag{17.5}$$

building higher and higher commutators until they fall into γ_{t+1} (and thus can be ignored). To make this precise, we need a definition.

Definition 17.19. Each x_i is called a **higher commutator** of **weight** 1. Inductively, given two higher commutators v and w of respective weights ℓ and m, we say that (v, w) is a **higher commutator** of **weight** $\ell + m$. By Lemma 17.18, each higher commutator of weight t is in $\mathcal{G}^{(t)}$.

We order the higher commutators (v, w) by the **weighted crossword dictionary order** where $x_1 < x_2 < \cdots < x_n$, i.e., first by weight, then by v, and last by w.

We would like to rewrite a word to get the form

$$\cdots v_m^{k_m} v_{m-1}^{k_{m-1}} \cdots v_2^{k_2} v_1^{k_1}, \qquad k_i \in \mathbb{Z},$$

where the v_j are higher commutators arranged in increasing order *from the right*; i.e., $v_1 = x_1$, $v_2 = x_2$, …, $v_n = x_n$, $v_{n+1} = (x_1, x_2)$, and so forth. For example, Relation (17.5) enables us to rearrange the x_i, each time replacing $x_i x_j$ by $(x_i, x_j)x_j x_i$ when $i < j$. Since we need whatever notational convenience we can clutch, we write (a_1, a_2, a_3) for the higher commutator $(a_1, (a_2, a_3))$, and so forth. But now we have new terms to move, namely the (x_i, x_j) for $i < j$, as well as higher commutators of weight ≥ 3. We repeat the procedure with higher commutators of greater weights; for example, $x_i(x_j, x_k)$ is replaced by $(x_i, x_j, x_k)(x_j, x_k)x_i$. More generally,

$$x_i(x_{j_1}, \dots, x_{j_t}) = (x_i, x_{j_1}, \dots, x_{j_t})(x_{j_1}, \dots, x_{j_t})x_i.$$

The situation becomes more complicated when we also deal with negative powers, since we would like our commutator symbols only to involve positive powers of the x_i. (For example, we do not want to deal with the higher commutator (x_1, x_2^{-1}).) Thus, having generated higher commutators $a^{\pm 1}$ and $b^{\pm 1}$ by moving terms to the right, we need formulas for moving $a^{\pm 1}$ to the right of $b^{\pm 1}$; we do not mind if we generate higher commutators of greater weight, since eventually these will be of weight $> t$ and discarded. Thus, we may proceed by reverse induction on the weight of the higher commutators.

Having dealt with ab in (17.5), we also need to consider $a^{-1}b^{-1}$, ab^{-1}, and $a^{-1}b$.

$$a^{-1}b^{-1} = b^{-1}a^{-1}(a, b), \tag{17.6}$$

and we can move a^{-1} and b^{-1} to the right of (a, b) by reverse induction on weight.

For $a^{-1}b$, we first observe that

$$(a^{-1}, b) = a^{-1}(b, a)a = a^{-1}(a, b)^{-1}a = (a, b)^{-1}a^{-1}(a, a, b)a. \tag{17.7}$$

By reverse induction on weight, we can move a^{-1} to the right of (a,a,b) to cancel a and write (a^{-1},b) as $(a,b)^{-1}$ times the product of higher commutators of greater weight, and then move $(a,b)^{-1}$ to the right (again by reverse induction on weight); putting everything together yields

$$(a^{-1},b) = \cdots (a,b)^{-1}, \tag{17.8}$$

where $\cdots$ denotes higher commutators that we have generated of greater weight than (a,b). Applying (17.8) to (17.5) yields

$$a^{-1}b = (a^{-1},b)ba^{-1} = \cdots (a,b)^{-1}ba^{-1}. \tag{17.9}$$

In this way, we have moved the higher commutators of least weight to the right, having generated higher commutators of greater weight, which we continue to rearrange via reverse induction on weight.

Finally, we have

$$ab^{-1} = b^{-1}(a,b)^{-1}a = (a,b)^{-1}b^{-1}(b,a,b)a, \tag{17.10}$$

and again by reverse induction, we can move b^{-1} to the right of (b,a,b).

In short, first we collect all occurrences of $x_1^{\pm 1}$ to the right using (17.5), (17.6), (17.9), and (17.10), getting $x_1^{k_1}$ at the right for some $k_1 \in \mathbb{Z}$; next we collect all x_2 to the immediate left of x_1 and continue until we have collected all single letters, and obtain the new word

$$w x_n^{k_n} \cdots x_1^{k_1},$$

where w only contains higher commutators of weight greater than 1. We move all occurrences (if any) of $(x_1,x_2)^{\pm 1}$ in w to the right (in w), and so on.

Example 17.19′. Let us illustrate this procedure for $t = 3$. Consider

$$(x_2x_1)^3 = x_2x_1x_2x_1x_2x_1,$$

involving only the two letters x_1 and x_2. We begin by moving the occurrences of x_1 to the right, starting from the right. First we get

$$x_2x_1x_2(x_1,x_2)x_2x_1^2,$$

and then

$$x_2(x_1,x_2)x_2(x_1,x_1,x_2)(x_1,x_2)(x_1,x_2)x_2x_1^3.$$

Having collected all single occurrences of x_1 at the right, we start moving x_2 to get

$$x_2(x_1,x_2)(x_2,x_1,x_1,x_2)(x_1,x_1,x_2)(x_2,x_1,x_2)(x_1,x_2)(x_2,x_1,x_2)(x_1,x_2)x_2^2x_1^3.$$

Suppressing all higher commutators of weight 4 or more and moving the leftmost x_2 to the right yields

$$(x_2,x_1,x_2)(x_1,x_2)(x_1,x_1,x_2)(x_2,x_1,x_2)^2(x_1,x_2)(x_2,x_1,x_2)^2(x_1,x_2)x_2^3x_1^3.$$

Now moving the occurrences of (x_1,x_2) and finally of (x_2,x_1,x_2) to the right (and still suppressing higher commutators of weight ≥ 4) yields

$$(x_2,x_1,x_2)^5(x_1,x_1,x_2)(x_1,x_2)^3x_2^3x_1^3. \tag{17.11}$$

Note that in (17.11), all higher commutators of weight < 3 occur with power 3. This key observation is needed later for the Restricted Burnside Problem; cf. Exercise C4.

When moving higher commutators to the right according to their order, starting with the smallest ones, we never encounter a situation $u(v,w)$ where $u < v$ and thus we never create (u,v,w) for $u < v$. (This is because (v,w) is only created when we move v, but by our procedure, all occurrences of u have already been moved to the right.) This observation leads us to the following definition:

Definition 17.19″. Basic (higher) commutators in the free group are defined by induction on the weight:

1. Each x_i is a basic commutator.
2. (x_i,x_j) is a basic commutator iff $i < j$.
3. Suppose u, v, and w are basic commutators with $u \geq v$ and $v < w$. Then (u,v,w) is a basic commutator.

For example, the only basic commutator in x_1 and x_2 having weight 2 is (x_1,x_2). The basic commutators of weight 3 are (x_1,x_1,x_2) and (x_2,x_1,x_2) (but not (x_2,x_2,x_1) since (x_2,x_1) is not a basic commutator). Continuing, the basic commutators of weight 4 are (x_1,x_1,x_1,x_2), (x_2,x_1,x_1,x_2), and (x_2,x_2,x_1,x_2). An example of a basic commutator of weight 5 is $((x_1,x_2),(x_1,x_1,x_2))$. (Here $u = w = (x_1,x_2)$ and $v = x_1$.)

This way of rewriting elements of the free group as products of basic commutators (modulo γ_t) in order of descending weight is called **Hall's collecting process**; cf. Exercise 18. The amazing fact, obtained in Exercises 19–22, is that the basic commutators of weight t yield a base for the free Abelian group γ_t/γ_{t+1}. One concludes in Exercise 23 that the free group can be ordered.

Resolutions of modules

We turn quickly to modules, which have the pleasant property that every submodule is the kernel of a module homomorphism. (This property is studied much more extensively in the context of homology, in Chapter 25.) Here is a preliminary result of independent interest.

PROPOSITION 17.20. *If* $0 \to M_1 \stackrel{f_1}{\to} M_2 \stackrel{f_2}{\to} \cdots \to M_k \to 0$ *is an exact sequence of f.g. modules over a left Artinian ring, then* $\sum_{j=1}^{k}(-1)^j \ell(M_j) = 0$, *where* $\ell(\ )$ *denotes the composition length.*

Proof. By induction on k, the proposition being obvious for $k = 1$ and $k = 2$. By induction, the exact sequence

$$0 \to \ker f_3 \to M_3 \stackrel{f_3}{\to} M_4 \to \cdots \to M_k \to 0$$

satisfies

$$-\ell(\ker f_3) + \sum_{j=3}^{k}(-1)^{j-1}\ell(M_j) = 0.$$

But $\ker f_3 \cong f_2(M_2) \cong M_2/(f_1(M_1))$, and $\ell(f_1(M_1)) = \ell(M_1)$ since f_1 is monic, so $\ell(\ker f_3) = \ell(M_2) - \ell(M_1)$; hence,

$$-\sum_{j=1}^{k}(-1)^j \ell(M_j) = 0,$$

as desired. □

Definition 17.21. Suppose $M \in R$-**Mod**. A **free resolution** of M is a (possibly infinite) exact sequence of free modules

$$\cdots \to F_n \to F_{n-1} \to \cdots \to F_0 \to M \to 0.$$

The resolution is called a **f.g. free resolution** if each F_n is f.g. If $F_{n+1} = 0$ for some n (in which case we could take $F_m = 0$ for all $m > n$), we say that the resolution has **length** n.

Example 17.21′. (i) Any module has a free resolution, obtained as follows: Write $M = F_0/K_0$, where F_0 is free. Thus we have an exact sequence $K_0 \to F_0 \to M \to 0$. But K_0 can be written as $F_1/K_2 1$, yielding the exact sequence

$$K_1 \to F_1 \to F_0 \to M \to 0.$$

This procedure is continued indefinitely.

Note that if K_n is free, then we can stop, with $F_n = K_n$, and have the free resolution $0 \to F_n \to \cdots \to F_0 \to M \to 0$ of length n.

(ii) Suppose that the ring R is left Noetherian. We can obtain a f.g. free resolution of a f.g. module M as in (i), noting at each stage that F_i can be taken to be f.g., and thus Noetherian, so its submodule K_i is also f.g., and so forth.

(iii) In Corollary 2.41 of Volume 1, we proved that every f.g. module over a PID has a f.g. free resolution

$$0 \to F_1 \to F_0 \to M \to 0,$$

thereby paving the way for the lovely theory presented there. The same conclusion holds for f.g. modules over PLID's, by Exercise 13.21.

This example cries out for a generalization, where we would like to say that a left Noetherian ring R has dimension n (in some sense) if every f.g. module has a f.g. free resolution of length $\leq n$. Thus, $n = 0$ for any division algebra and $n = 1$ for any PLID. Unfortunately, this definition falls short for the "best" noncommutative ring.

Example 17.22. Suppose L is a minimal left ideal of $R = M_2(F)$. Then $\dim_F L = 2$, whereas the dimension (as a vector space over F) of any free R-module is a multiple of 4. It follows at once that each K_i in the resolution must have dimension congruent to 2 modulo 4, and thus is not free; hence, L cannot have a f.g. free resolution of finite length.

A similar example holds which is a homomorphic image of a PID; cf. Exercise 26. To correct these failures, we must think more categorically, as to be done in Chapter 25.

Graphs

In recent years, graphs have proven to be a versatile tool in many aspects of combinatoric algebra; we introduce them here to understand some fundamental concepts in group theory.

Definition 17.23. A **graph** Γ is a collection of **vertices**, joined by **edges**. The graph Γ is called **directed** if each edge has a direction (customarily designated by an arrow). Thus, in a directed graph, an edge e is identified with an ordered pair (v, w) of **vertices**, where v is called the **initial vertex** and w the **terminal vertex**. In short, the (directed) graph $\Gamma = \Gamma(E, V)$ can be identified with its set V of vertices and set E of edges, where $E \subseteq V \times V$.

The **degree** of a vertex v is the number of edges to which v belongs. We only consider graphs for which each vertex has finite degree. An edge of the form (v, v) is called a **loop**; we assume that our graphs are loopless. The graph Γ is called **finite** if V (and thus E) is finite.

Note that this definition does not provide for multiple edges; one could do this formally or just designate the multiplicity for each edge. In this chapter, we only consider edges of multiplicity 1; the issue of higher multiplicities is deferred until Chapter 22.

We have a category whose objects are (directed) graphs; for graphs $\Gamma_1 = \Gamma(V_1, E_1)$ and $\Gamma_2 = \Gamma(V_2, E_2)$, a **morphism** $\Gamma_1 \to \Gamma_2$ is a function $f : V_1 \to V_2$ such that $(f(v), f(w)) \in E_2$ for any edge $(v, w) \in E_1$.

Remark 17.23′. Here is one way to treat undirected graphs. Given an edge $e = (v, w)$, we define the **opposite edge** $\bar{e} = (w, v)$ in the opposite direction. We **double** a directed graph by introducing $\bar{e}$ to E for each edge $e \in E$. (In this way, we could formally define an undirected graph as a doubled directed graph in which we identify e and $\bar{e}$, for each edge e.)

Definition 17.24. A **path** from a vertex v to a vertex w is a sequence of edges starting with v and ending with w; the number of such edges needed is called the **length** of the path. In other words, we might denote a path p of length ℓ as $e_1 e_2 \cdots e_\ell$, where

$$e_1 = (v, v_1), \quad e_2 = (v_1, v_2), \quad \ldots, \quad e_\ell = (v_{\ell-1}, w)$$

for suitable intermediate vertices v_i. In the doubled graph, we can also define the **reverse path**

$$\overline{p} = \bar{e}_\ell \cdots \bar{e}_1$$

from w to v. A graph is **connected** if for any two vertices v, w there is a path either from v to w or from w to v.

Note that $\overline{\overline{p}} = p$ for any path p. A path from a vertex v to itself is called a **circuit**, or v-**circuit** if we wish to emphasize the endpoint v. We have already excluded circuits of length 1, i.e., loops. A trivial example of a circuit is $e\bar{e}$, an edge followed by its opposite edge. More generally, we say that any path p is **equivalent** to a path p' if we can obtain p' from p by inserting and/or deleting paths of the form $e\bar{e}$; a circuit equivalent to a trivial circuit is also called **trivial**. (For example, $e_1\bar{e}_2 e_2 e_3 \bar{e}_4 e_4 \bar{e}_3 \bar{e}_1$ is trivial; first we delete $\bar{e}_2 e_2$, then $\bar{e}_4 e_4$, then $e_3\bar{e}_3$, and are left with $e_1\bar{e}_1$.) A connected graph having no nontrivial circuits is called a **tree**.

A vertex v of a directed graph Γ is **initial for** the graph Γ if Γ has no edge of the form (w, v). (Thus, the only edges containing v are of the form $(v, ?)$.) An example of a directed graph without an initial vertex is $\{(n, n+1) : n \in \mathbb{Z}\}$. Any connected, directed graph can have at most one initial vertex, since no path could connect two initial vertices.

The notion of path can be generalized to an **infinite path** with a given initial vertex v; this is an infinite sequence of edges $e_i = (v_{i-1}, v_i)$, $i \geq 1$, where $v_0 = v$. The following observation is one of the basic uses of graph theory in combinatorics.

Proposition 17.25 (König Graph Theorem). *Suppose Γ is an infinite connected, directed graph with initial vertex v_0. Then Γ has an infinite path with initial vertex v_0.*

Proof. If there is some circuit, then the result is immediate; one just takes a path from v_0 to the circuit and continues infinitely many times in the circuit. Thus, we may assume that Γ is a tree. Suppose $\deg v_0 = m_0$. This means there are m_0 edges

$$\{e_{0,i} = (v_0, v_{1,i}) : 1 \leq i \leq m_0\}.$$

If we erase the edges $e_{0,i}$, $1 \leq i \leq m_0$, then we have m_0 connected graphs $\Gamma_1, \ldots, \Gamma_{m_0}$, each with respective initial vertex $v_{1,i}$, $1 \leq i \leq m_0$, so some Γ_i must be infinite; we write v_1 for the corresponding $v_{1,i}$. Continuing, we get $v_2, v_3, \ldots$ in turn, and thus have an infinite path. □

There is an analogous formulation for undirected graphs; cf. Exercise 28. Intuitively, the König graph theorem says that any process with only finitely many decisions to be taken at each stage, either involves finitely many decisions altogether or else contains some sequence of decisions that does not terminate.

The Cayley graph.

Our combinatoric approach utilizes the following graph arising from a f.g. monoid.

Definition 17.26 (The Cayley graph of a monoid). Any monoid M with a given set of generators $\{a_0 = 1, a_1, \ldots, a_n\}$ gives rise to a directed graph, called the **Cayley graph** whose vertices are the elements of M, in which vertices v_1, v_2 are joined by an edge $e = (v_1, v_2)$ iff $v_2 = v_1 a_i$ for some $1 \leq i \leq n$. Any homomorphism of monoids (sending generators to generators) induces a corresponding morphism of the respective graphs.

Example 17.27. The Cayley graph of the free monoid $\mathcal{M}$ on n letters $x_1, \dots, x_n$ has an initial vertex corresponding to 1, and each vertex corresponds to a unique word w. There is a unique path of length t from 1 to any word of length t; namely if $w = x_{j_1} \cdots x_{j_t}$, then we put $v_i = x_{j_1} \cdots x_{j_i}$, and $e_i = (v_{i-1}, v_i)$ for $1 \leq i \leq t$. Each vertex v has degree $n+1$ (being the initial vertex of (v, vx_i) for each $1 \leq i \leq n$, and the terminal vertex of (v', v) where $v = v'x_j$ for some j).

The Cayley graph of any monoid M generated by n elements is an image of the Cayley graph of the free monoid, according to the monoid homomorphism $\mathcal{M} \to M$.

The situation becomes much more complicated for algebras, since the operation of addition does not translate naturally to the graph structure. We can bypass this difficulty in one useful situation.

Definition 17.28. Suppose $R = F\{x_1, \dots, x_n\}/A$ is a monomial algebra, where A is the ideal generated by words $\{w_i : i \in I\}$; i.e., R is defined by the monomial relations $\{w_i = 0 : i \in I\}$. The **Cayley graph** of R (with respect to this presentation) is the image of the Cayley graph of the free monoid $\mathcal{M}$, where we erase any vertex (and edge to which it belongs) corresponding to a word in A.

Thus, the vertices of the graph correspond to the "surviving" words in R, which clearly comprise a base of R. For example, the Cayley graph of $F[x]/\langle x^n \rangle$ would be the path of length $n-1$, from the vertex 1 to the vertex x^{n-1}. Although this definition is stated in terms of algebras, it is really a variant of Example 17.26, as explained in Exercise 30.

Example 17.29 (The Cayley graph of a group). Suppose G is a group with a given finite set of generators $\{g_1, \dots, g_n\}$. As in Example 17.26, the vertices of the Cayley graph Γ are words (this time in the g_i and g_i^{-1}); Γ has an edge $e = (v_1, v_2)$ iff $v_2 = v_1 g_i^{\pm 1}$ for some i, $1 \leq i \leq n$. But $v_2 = v_1 g_i$ iff $v_1 = v_2 g_i^{-1}$, so Γ also has the opposite edge $\bar{e}$. Accordingly, we may view Γ as an undirected graph.

Circuits in the Cayley graph clearly correspond to relations in the group. Thus, the Cayley graph of the free group in n letters is a tree, where each vertex has degree $2n$. The word problem for an arbitrary group G is equivalent to identifying which words map to circuits in the Cayley graph of G.

Remark 17.29′. We follow the convention that the generators $\{g_1, \dots, g_n\}$ of G are chosen to be a minimal generating set. In particular, each $g_i \neq 1$, $g_i \neq g_j^{\pm 1}$, and $g_i \neq g_j^{\pm 1} g_k^{\pm 1}$ for all i, j, k. Consequently, the Cayley graph has no circuits of length ≤ 3.

We have just seen how the Cayley graph enables us to pass from various algebraic structures to graphs. One can also go in the other direction, as to be seen in Chapter 22 and Appendix 25C.

Growth of algebraic structures

When considering algebraic structures in terms of generators and relations, we must deal respectively with the following basic questions:

A. How quickly do the generators generate given elements?
B. How can we reduce an element to some "normal form" in terms of its generators?
C. Can all structural questions be answered by means of algorithms described in terms of generators and relations?

We start with Question A. At this stage, we are interested mostly in monoids, groups, and associative algebras. We give the relevant definitions here for these important examples; they are unified in Exercise 42 in the language of universal algebra. (Nonassociative algebras are considered in Exercise 21B.42.)

Definition 17.30. Suppose G is a monoid with a finite generating set S. Define $S^1 = S$ and, inductively,

$$S^k = S^{k-1}S = \{s_{i_1} \cdots s_{i_k} : s_{i_j} \in S,\ 1 \leq j \leq k\}.$$

We assume that $1 \in S$ so that $S^{k-1} \subseteq S^k$. (Otherwise, the definition is more complicated.) The **growth function** $\mathfrak{g}_S : \mathbb{N} \to \mathbb{N}$ of G (with respect to S) is defined as $\mathfrak{g}_S(k) = |S^k|$.

This information is encoded in the formal power series $\sum_{k=1}^{\infty} d_k \lambda^k$, where $d_k = |S^k \setminus S^{k-1}|$. Although this series is not used much in group theory, its analog for algebras is studied extensively.

When considering algebras over a field, we also take the vector space structure into account.

Definition 17.31. Given a field F, suppose R is an associative affine F-algebra with a finite generating set S containing the element 1. Letting V denote the subspace $\sum_{s \in S} Fs$, put $V^1 = V$ and $V^k = VV^{k-1}$ for each $k \geq 2$. (We write $V^k(R; S)$ when R and S need to be specified.) The **growth function** $g_S : \mathbb{N} \to \mathbb{N}$ of R is defined as $\mathfrak{g}_S(k) = \dim_F V^k$.

Let $d_k = \dim_F(V^k/V^{k-1})$. (Sometimes we write $d_k(R; S)$ to specify R and S.) The **Hilbert series**, also called the **Poincaré series** of R (with respect to S), is defined as the formal power series $\sum_{k=1}^{\infty} d_k \lambda^k$, viewed in $\mathbb{Z}[[\lambda]]$.

Noting that, since $1 \in S$, the $\{V^k : k \in \mathbb{N}\}$ in Definition 17.31 comprise a filtration of R, we generalize the definition as follows:

Definition 17.32. For any $\mathbb{N}$-filtration $\{R_k : k \in \mathbb{N}\}$ of an algebra R, define $d_k = \dim_F(R_k/R_{k-1})$, the corresponding **growth function** $f(k) = d_k$, and the **Hilbert series** $\sum_{i=1}^{\infty} d_k \lambda^k$ **of the filtration**.

Remark 17.33. Given any filtration $\{R_k : k \in \mathbb{N}\}$, take the associated graded algebra

$$\operatorname{gr} R = \bigoplus_{k \in \mathbb{N}} R_k/R_{k-1};$$

cf. Definition 7.30 of Volume 1. By definition, the growth function and thus the Hilbert series of $\operatorname{gr} R$ with respect to its natural filtration is the same as the Hilbert series of R with respect to the original filtration. This reduces much of the theory of growth to the case of graded algebras.

Example 17.34. Suppose $R = F\{a_1, \dots, a_n\}$ is affine.

(i) We filter R according to the length of words in the a_i; in other words, R_k is the space spanned by all words of length $\leq k$. We write $\bar{a}_i$ for $a_i + R_0$ in $\operatorname{gr} R$. Thus, $\operatorname{gr} R = F\{\bar{a}_1, \dots, \bar{a}_n\}$; any element $r \in R$ is sent to its "leading part" in $\operatorname{gr} R$ according to total degree.

(ii) We can refine the filtration by generalizing these considerations to algebras filtered by $\mathbb{N}^{(n)}$, ordered lexicographically. Namely, writing $\mathbf{k} = (k_1, \dots, k_n)$, we now define $R_{\mathbf{k}}$ to be generated by words in which $\deg a_i = k_i$, $1 \leq k \leq n$. The associated graded algebra is

$$\operatorname{gr} R = \bigoplus_{\mathbf{k} \in \mathbb{N}^{(n)}} \left(R_{\mathbf{k}} \Big/ \sum_{\mathbf{j} < \mathbf{k}} R_{\mathbf{j}} \right).$$

(iii) The following variant reduces much of the theory of growth to the case of monomial algebras. We order $a_1, \dots, a_n$ alphabetically and order the words formed from products of $a_1, \dots, a_n$ according to the crossword dictionary order of Definition 17.9. We say that a word is **reducible** if it can be written in R as a linear combination of smaller words. Clearly the set of irreducible words comprises a base of R over F, and any dependence relation in R can be expressed by writing the largest word in the relation as a linear combination of the others. It follows that $\operatorname{gr} R$ is precisely the monomial algebra whose relations are comprised of the reducible words.

But this means that the d_k of Definition 17.32 is just the number of paths in the Cayley graph of $\operatorname{gr} R$ (cf. Definition 17.28) of length k that originate with its initial vertex 1. Thus, the determination of growth is a purely graph-theoretic problem that can be transferred to an appropriate monoid.

Although we have several parallel definitions of growth, Definition 17.31 encompasses Definition 17.30 via the construction of monoid algebras (given in Exercise 1; group algebras are studied in depth in Chapters 19 and 20). Indeed, the growth series of any monoid G with respect to a generating subset S is precisely the same as the Hilbert series of the monoid algebra $F[G]$ with respect to S. On the other hand, in Example 17.34(iii), we reduced considerations about algebra growth to monoid growth. Thus, the theories of growth of algebras and of monoids are the same in many respects; the theory of growth of groups is more special, groups being a special case of monoids.

Rationality of the Hilbert series.

The Hilbert series of an n-dimensional algebra is a polynomial of degree n (since $d_k = 0$ for all $k > n$); conversely, if the Hilbert series of an affine algebra is a polynomial, then the algebra ceases to grow past the degree of the polynomial and thus must be finite-dimensional. So our next goal is to understand the Hilbert series for non-f.d. algebras. In view of Remark 17.33, it is enough to consider $\mathbb{N}$-graded algebras, and at times it is easier to consider, more generally, their graded modules.

Definition 17.35. Suppose R is an $\mathbb{N}$-graded ring. The Poincaré series of a f.g. graded R-module M is the formal power series

$$P(M) = \sum_{i=0}^{\infty} \ell(M_i)\lambda^i \in \mathbb{Z}[[\lambda]],$$

where $\ell(\ )$ denotes the composition length as an R_0-module; cf. Definition 3.7ff. of Volume 1.

(This definition reduces to Definition 17.31 for $M = R$ when $R_0 = F$.) Recalling that $\mathbb{Z}[[\lambda]] \subset \mathbb{Q}((\lambda))$, which contains $\mathbb{Q}(\lambda)$, the **rational functions** in λ, we are led to ask when $P(M)$ is rational. The next result gives a good preliminary answer.

Proposition 17.36. *Suppose $R = F[a_1, \dots, a_n]$ is an $\mathbb{N}$-graded ring with $a_1, \dots, a_n$ homogeneous, and R_0 is left Artinian. Then the Poincaré series $P(M)$ of any f.g. graded R-module M is rational, of the form*

$$\frac{f(\lambda)}{\prod_{i=1}^{n}(1 - \lambda^{\deg a_i})}.$$

Proof. Induction on n, the assertion being clear for $n = 0$. $N = M/a_nM$ is naturally graded. Let $d = \deg a_n$. For each j, left multiplication by a_n induces a map $f\colon M_j \to M_{j+d}$ and thus an exact sequence

$$0 \to K_j \to M_j \to M_{j+d} \to N_{j+d} \to 0,$$

where $K_j = \ker f$ and $N_{j+d} = M_{j+d}/a_nM_j$. By Proposition 17.20,

$$\ell(K_j) + \ell(M_{j+d}) = \ell(M_j) + \ell(N_{j+d});$$

multiplying each equation by λ^{j+d} and adding (for all $j \in \mathbb{N}$) yields

$$\lambda^d P(K) + P(M) - \sum_{j=0}^{d-1} \ell(M_j)\lambda^j = \lambda^d P(M) + P(N) - \sum_{j=0}^{d-1} \ell(N_j)\lambda^j,$$

where $K = \oplus K_j$. Thus,

$$(1-\lambda^d)P(M) = P(N) - \lambda^d P(K) + \sum_{j=0}^{d-1} \ell(M_j)\lambda^j - \sum_{j=0}^{d-1} \ell(N_j)\lambda^j.$$

But K and N both are annihilated by a_n and thus defined over $R/\langle a_n\rangle$, so, by induction, $P(N) - \lambda^d P(K)$ has the form

$$\frac{g(\lambda)}{\prod_{i=1}^{n-1}(1-\lambda^{\deg a_i})},$$

implying that

$$P(M) = \frac{g(\lambda) + \sum_{j=0}^{d-1}(\ell(M_j) - \ell(N_j))\lambda^j \prod_{i=1}^{n-1}(1-\lambda^{\deg a_i})}{\prod_{i=1}^{n}(1-\lambda^{\deg a_i})}. \qquad \square$$

COROLLARY 17.37. *The Hilbert series of an arbitrary commutative graded affine algebra R (with respect to a homogeneous generating set) is rational.*

Proof. View R as a module over itself. $\square$

Putting everything together yields the following basic result, illustrating the power of passing to graded algebras:

THEOREM 17.38. *The Hilbert series of any commutative affine algebra R (with respect to any generating set) is rational.*

Proof. Pass to the associated graded algebra via Remark 17.33 and apply Corollary 17.37. $\square$

Example 17.39 (Polynomial algebras).

(i) If $R = F[\lambda]$ and $S = \{1, \lambda\}$, then $S^k \setminus S^{k-1} = \{\lambda^k\}$, so $d_k = 1$ for each k, and the Hilbert series is $\sum_{k\geq 0} \lambda^k$, which formally equals $\frac{1}{1-\lambda}$.

(ii) If $R = F[\lambda_1, \ldots, \lambda_n]$ and $S = \{1, \lambda_1, \ldots, \lambda_n\}$, then $S^k \setminus S^{k-1}$ is the set of monomials of total degree k. By the same token as (i), the coefficient of the monomial $\lambda_1^{k_1} \cdots \lambda_n^{k_n}$ is the same as in the expansion $\prod_{i=1}^n \frac{1}{1-\lambda_i}$, so specializing all $\lambda_i \mapsto \lambda$ shows that the Hilbert series of R is $\frac{1}{(1-\lambda)^n}$.

The Hilbert series of any commutative affine algebra $R = F[a_1, \ldots, a_n]$ with respect to the generating set $\{a_1, \ldots, a_n\}$ is rational; cf. Exercise 32. This can be stretched a bit further; cf. Exercise 33. Although not all noncommutative affine algebras have rational Hilbert series, many do; one instance is given in Exercise 23.26.

Rates of growth.

Although the Hilbert series is very important, it reflects the particular choice of generating set. A somewhat cruder indicator is independent of the generating set.

Definition 17.40. Define the following partial order on functions $\mathbb{N} \to \mathbb{N}$: $f \preceq g$ iff there are c, m in $\mathbb{N}$ such that $f(k) \leq cg(mk)$ for almost all $k \in \mathbb{N}$. We say that $f \sim g$ if $f \preceq g$ and $g \preceq f$.

$\sim$ is clearly an equivalence relation.

Remark 17.41. The equivalence class $[\mathfrak{g}_S]$ of the growth function of a f.g. monoid G does not depend on the choice of the generating set S of G. Indeed, adding or deleting one element at a time reduces the assertion to the case where the other generating set $\hat{S} = S \cup \{s'\}$. Then $s' \in S^m$ for some m, implying that $\hat{S}^k \subseteq S^{mk} \subseteq \hat{S}^{mk}$ for each k, and thus $\mathfrak{g}_S \sim \mathfrak{g}_{S'}$. Thus, we may define the **growth rate** $\mu(G)$ of G to be the equivalence class $[\mathfrak{g}_S]$. By the same token, for any affine algebra R, the equivalence class $[\mathfrak{g}_S]$ does not depend on the choice of the generating set S of R, and we speak of the **growth rate** $\mu(R)$ of R.

Definition 17.42. Define the **polynomial function** $p_m\colon x \mapsto x^m$ and the **exponential function** $e_t\colon x \mapsto t^x$. We say that an affine algebra R has **polynomially bounded growth** if $\mu(R) \preceq p_m$ for some m; R has **polynomial growth** of **degree m** if $\mu(R) \sim p_m$. Polynomial growth of degree 1 (resp. degree 2) is called **linear** (resp. **quadratic**) **growth**.

R has **exponential growth** if $\mu(R) \sim e_2$. We say that R has **subexponential growth** if $\mu(R) \prec e_2$. (Sometimes in the literature, one also requires that $\mu(R) \succ p_m$ for each m. We call this **intermediate growth**).

Remark 17.42′. e_t is clearly the greatest possible growth rate for any monoid, group, or associative algebra generated by t elements. Also, note that $u^k = t^{(\log_t u)k}$, implying that the e_t have equivalent growth for all numbers $t > 1$, which is why we take $t = 2$.

This theory becomes interesting only for infinite-dimensional algebras.

Remark 17.43. An affine algebra R is finite-dimensional iff $\mu(R) \sim p_0$, since its growth stops. On the other hand, if R is infinite-dimensional, then clearly $\mathfrak{g}_R(k+1) > \mathfrak{g}_R(k)$ for each k, implying that $\mathfrak{g}_R(k) \geq k$, and the growth of R must be at least linear.

Remark 17.44. (i) If $R \subseteq W$, then $\mu(R) \preceq \mu(W)$. (Just take a generating set of W that contains a generating set of R.)

(ii) If $I \lhd R$, then $\mu(R/I) \preceq \mu(R)$.

(iii) Notation as in Definition 16.3, if $W = RT$ is a centralizing extension of R such that T is f.g. over a subring $C \subseteq \mathrm{Cent}(R)$, then $\mu(W) = \mu(R)$. (Indeed, write $T = \sum_{i=1}^m Cb_i$ and let $b_i b_j = \sum_{k=1}^m c_{ijk} b_k$. Taking a generating set S of R that contains all the c_{ijk}, note that $S' = S \cup \{b_1, \dots, b_m\}$ generates W. Then $V^k(W; S') \subseteq \sum_{i=1}^m V^k(R; S) b_i$, so $d_k(W; S') \leq m d_k(R; S)$ and the growth rates are the same.)

(iv) $\mu(M_n(R)) = \mu(R)$ for any affine algebra R and any number n. (A special case of (iii).)

This result can be generalized to arbitrary f.g. extensions of subrings; cf. Exercise 35.

One can create rather erratic growth rates; cf. Exercise 39. In fact, the growth rate could conceivably bounce back and forth between two different real numbers, neither of which need be rational.

Gel′fand-Kirillov dimension

A close study of the growth rate leads at once to difficulties — since there are infinitely many possible growth rates between, say, p_2 and p_3, one finds it difficult to compare these growth rates in a meaningful manner. There is a single number that measures the polynomially bounded growth of an algebra. (See Krause-Lenagan [**KrauL**] for an excellent, much more detailed treatment.)

Definition 17.45. Notation as in Definition 17.31, let $\widetilde{d}_k = \dim_F V^k$. (Thus, $\widetilde{d}_k = \sum_{i=0}^k d_i$.) We define the **Gel′fand-Kirillov dimension**

$$\mathrm{GKdim}(R) = \varlimsup_{k\to\infty} \log_k \widetilde{d}_k.$$

At first glance, this definition may seem technical, if not baffling. However, it tends to be very useful, since it quantifies the growth. First let us see that at least it makes sense.

Remark 17.45′. GKdim is well-defined, i.e., independent of the choice of generating set. Indeed, in view of Remark 17.41, it suffices to show that equivalent growth functions give rise to the same value of GKdim. But

$$\log_k(c\widetilde{d}_{mk}) = \log_k c + \log_k(mk) \log_{mk} \widetilde{d}_{mk} = \log_k c + (1+\log_k m) \log_{mk} \widetilde{d}_{mk},$$

whose limit is again GKdim(R), since $\varlimsup_{k\to\infty} \log_k c = \varlimsup_{k\to\infty} \log_k m = 0$.

Furthermore, the growth rate p_k obviously corresponds to GKdim $= k$. In particular, GKdim$(R) = 0$ for any finite-dimensional algebra R.

The following observation is an instant application of Remark 17.44.

Remark 17.46. (i) If $R_1 \subseteq R_2$, then GKdim$(R_1) \leq$ GKdim(R_2).

(ii) If $I \lhd R$, then GKdim$(R/I) \leq$ GKdim(R).

(iii) GKdim(W) = GKdim(R) for any centralizing extension $W = RT$ of R such that $\dim_F T < \infty$ (notation as in Definition 16.3).

(iv) GKdim$(M_n(R))$ = GKdim(R) for any affine algebra R and any number n.

In view of Remark 17.46(i), one can generalize GKdim to non-affine algebras by taking the supremum of the GKdim of all affine subalgebras, but we do not enter into this extra level of complication. GKdim behaves astonishingly well, mainly because of the following example.

Example 17.47. (i) GKdim$(R[\lambda])$ = GKdim$(R)+1$. Indeed, taking the generating set $S' = S \cup \{\lambda\}$ for $R[\lambda]$, where S is a generating set for R, we see that $\widetilde{d}_k(R[\lambda]; S') = \sum_{j=0}^k \widetilde{d}_j(R; S)$ since we get $d_j(R; S)$ monomials for each term λ^{k-j}. Choosing only $j > \frac{k}{2}$ shows

$$\widetilde{d}_k(R[\lambda]; S') \geq \frac{k}{2} \widetilde{d}_{k/2}(R; S).$$

On the other hand, clearly $\widetilde{d}_k(R[\lambda];S') \leq k\widetilde{d}_k(R;S)$. Together, taking logarithms (base k) yields

$$1 - \log_k 2 + \log_k \widetilde{d}_{k/2}(R;S) \leq \log_k \widetilde{d}_k(R[\lambda];S') \leq 1 + \log_k \widetilde{d}_k(R;S);$$

taking lim sup as $k \mapsto \infty$ yields $1 + \mathrm{GKdim}(R)$ at both ends, proving that $\mathrm{GKdim}(R[\lambda]) = 1 + \mathrm{GKdim}(R)$.

(ii) $\mathrm{GKdim}(F[\lambda_1, \dots, \lambda_n]) = n$, seen by iterating (i).

LEMMA 17.48. *If R is commutative affine and integral over an affine subalgebra C, then $\mu(R) \sim \mu(C)$, implying that* $\mathrm{GKdim}(R) = \mathrm{GKdim}(C)$.

Proof. By induction, we may assume that $R = C[a]$ for a integral; then R is f.g. over C, so we are done by Remark 17.44(iii) and Remark 17.45′. □

PROPOSITION 17.49. *Any commutative affine algebra has integral* GKdim, *equal both to its Krull dimension and to its transcendence degree.*

Proof. This follows from Theorem 17.38, but there is a more direct argument: By the Noether normalization theorem, R is integral over a polynomial algebra R_0. Thus, we are done by Example 17.47(ii) and Lemma 17.48, compared with Theorem 6.35 of Volume 1. □

COROLLARY 17.50. *If R is an algebra with filtration whose associated graded algebra is commutative affine, then* $\mathrm{GKdim}(R)$ *is an integer.*

Proof. Apply Remark 17.33 to Proposition 17.49. □

From this point of view, GKdim is an excellent noncommutative generalization of the various transcendence dimensions used in commutative algebra. It turns out that enveloping algebras of finite-dimensional Lie algebras and prime affine PI-algebras have integral GKdim, as seen in Remark 21A.4 and Exercise 23.26 respectively, although examples exist of noncommutative affine algebras with non-integral GKdim; cf. Exercises 38 and 39.

Hyperwords and Bergman's Gap.

Bergman [**Berg**] proved that any algebra of $\mathrm{GKdim} > 1$ has $\mathrm{GKdim} \geq 2$. This result is really a theorem about words. It is convenient to generalize the definition of word to enable us to study behavior of arbitrarily long words.

Definition 17.51. A **hyperword** is a right infinite sequence of letters. For any word u, the hyperword comprised of u repeated indefinitely is denoted as u^∞.

Remark 17.52. If a hyperword h satisfies $h = uh$, where u is a (finite) word, then $h = u^\infty$. (Indeed, for each k, $u^k h = u^{k+1}h$, implying $h = u^{k+1}h$; now let $k \mapsto \infty$.)

Of course in an alphabet of n letters, there are n^t possible (distinct) words of length t. We write $\nu_t(h)$ to be the number of distinct words of length t that appear as subwords in a given hyperword h. For example, $\nu_1(h)$ is the number of distinct letters appearing in h.

Remark 17.53. For any occurrence of a subword v of length t in a hyperword h, we get a subword of length $t+1$ by tacking on the next letter of h. Hence $\nu_{t+1}(h) \geq \nu_t(h)$ for any t. By the pigeonhole principle, if $\nu_{t+1}(h) = \nu_t(h)$, then for any subword v of h having length t, there is a unique subword of length $t+1$ starting with v.

We are ready for a basic lemma:

LEMMA 17.54. *If $\nu_t(h) = \nu_{t+1}(h) = m$ for suitable t, then h has the form vu^∞, where $|u| \leq m$ and $|v| < m$.*

Proof. Write $h = x_{i_1}x_{i_2}x_{i_3}\cdots$, and, for $1 \leq j \leq m+1$, let

$$v_j = x_{i_j}x_{i_{j+1}}\cdots x_{i_{j+t-1}}$$

denote the subword of h of length t starting in the j position. Then $v_1, \dots, v_{m+1}$ are $m+1$ subwords of h having length t so, by hypothesis, two of these are equal; say $v_j = v_k$ with $1 \leq j < k \leq m+1$.

Let h_j denote the sub-hyperword of h starting in the j position. By Remark 17.53, t consecutive letters of h determine the next letter uniquely, so h_j is determined by its initial subword v_j. Hence

$$h_j = h_k = x_{i_j}\cdots x_{i_{k-1}}h_j,$$

so Remark 17.52 implies $h_j = u^\infty$ where $u = x_{i_j}\cdots x_{i_{k-1}}$, as desired. □

We say that a hyperword is **quasiperiodic** if it has the form vu^∞. Thus, Lemma 17.54 gives a condition for every hyperword to be quasiperiodic.

THEOREM 17.55 (BERGMAN GAP THEOREM). *If* $\mathrm{GKdim}(R) > 1$, *then* $\mathrm{GKdim}(R) \geq 2$.

Proof. In view of Example 17.34(iii), it suffices to prove this for monomial algebras R, in which case the growth function corresponds to the paths of length k from the initial vertex. We consider hyperwords in the generators

of R having the property that any finite subword has nonzero value. (These values are linearly independent, since R is a monomial algebra.)

First assume that all hyperwords are quasiperiodic. For any quasiperiodic hyperword $h = uv^\infty$, there are at most $|u| + |v|$ distinct subwords of any given length, since any subword starting after the $|u| + |v|$ position just duplicates a subword starting earlier.

But all of these quasiperiodic hyperwords have the form uv^∞, where all the $|u|$ and $|v|$ are bounded by some number m, since otherwise, by the König graph theorem (Proposition 17.25, first applied to u and then to v), one would get a non-quasiperiodic hyperword. This means there are at most n^{2m} hyperwords, and thus at most $2mn^{2m}$ monomials of any given length k. It follows that

$$\widetilde{d}_k \leq \sum_{i=1}^{k} 2mn^{2m} = 2mkn^{2m},$$

so $\log_k \widetilde{d}_k \leq 1 + \log_k(2mn^{2m})$ whose limit is 1 as $k \to \infty$.

Thus, we are done unless there is a hyperword h that is not quasiperiodic. By Lemma 17.54, $\nu_{t+1}(h) > \nu_t(h)$ for each t. But $\nu_1(h) > 1$, so

$$\widetilde{d}_k \geq 2 + 3 + \cdots + (k+1) = \frac{k^2 + 3k}{2}, \tag{17.12}$$

which is quadratic, so $\mathrm{GKdim}(R) \geq 2$. □

Remark 17.56. More precisely, we have proved that the growth of an infinite f.g. monoid is either linear or at least quadratic, with a bound given by Equation (17.12).

The next step presumably would be to develop a structure theory of affine algebras based on the Gel′fand-Kirillov dimension, starting with the fact that $\mathrm{GKdim}(R) = 0$ iff R is f.d. Some impressive steps were taken early on, such as the theorem of Small and Warfield [**SmaW**], which states that if R is prime with $\mathrm{GKdim}(R) = 1$, then R is f.g. (as a module) over a central subring isomorphic to $F[\lambda]$. However, further progress has been slow. Smoktunowicz and Vishne [**SmoV**] have found some rather poorly behaved algebras of GKdim 2, and Lenagan and Smoktunowicz[**LenS**] have found infinite dimensional F-algebras of finite GKdim that are algebraic over F. It does follow from Proposition 17.12 that any affine domain of subexponential growth is Ore, but little else is known.

Growth of groups

Since the theory of growth of groups is more restrictive than that of algebras, one hopes to say more about their growth. We recall nilpotent groups from Definition 4.83 of Volume 1 and the subsequent discussion.

Definition 17.57. A group G is **nilpotent of class** $\leq t$ iff $\gamma_{t+1}(G) = \{1\}$; cf. Definition 17.17′. A group is **virtually nilpotent** if it has a nilpotent subgroup of finite index.

It turns out, by work of Wolf, Milnor, and Gromov, that a f.g. group G has polynomially bounded growth iff it is virtually nilpotent, in which case G has polynomial growth (of some integer-valued degree). Furthermore, any f.g. solvable group G of subexponential growth is virtually nilpotent (and thus cannot have intermediate growth). Following Tits' appendix to Gromov [**Grom1**] and [**KrauL**], we present the result for solvable groups. We start by recording special cases of Remarks 17.43 and 17.44.

Remark 17.58. Assume G is a f.g. group.

(i) $\mu(G) \sim p_0$ (the constant function 1) iff G is finite.

(ii) Any infinite cyclic group has linear growth, since each $d_k = 1$.

(iii) If $H \leq G$, then $\mu(H) \preceq \mu(G)$.

(iv) If $N \lhd G$, then $\mu(G/N) \preceq \mu(G)$.

Here is another easy observation.

Remark 17.58′. If $G = G_1 \times G_2$ where $\mathfrak{g}_i$ is the growth function of G_i, then the growth function of G is $\mathfrak{g}_1\mathfrak{g}_2$. In particular, if the groups G_i have polynomial growth rates d_i for $i = 1, 2$, then G has polynomial growth rate $d_1 + d_2$. (Indeed, take generating sets S_i of G_i for $i = 1, 2$, and let $S = S_1 \times S_2$. Then $S^k = S_1^k \times S_2^k$. The second assertion follows since $x^{d_1+d_2} = x^{d_1}x^{d_2}$.)

Proposition 17.59. *Any f.g. Abelian group G has polynomial growth.*

Proof. By the fundamental theorem of f.g. Abelian groups (Theorem 2.77 of Volume 1), $G \cong \mathbb{Z}^{(n)} \times H$ for some finite group H, in which case $\mu(G) \sim p_n$ by Remark 17.58(i), (ii) and Remark 17.58′. □

Since nilpotent groups often behave like Abelian groups, one is not surprised to have the following theorem of Wolf [**Wo**], with the explicit formula due to Bass [**Ba3**]:

THEOREM 17.60. *If a nilpotent group N has lower central series $N = \gamma_1 > \gamma_2 > \cdots > \gamma_t = \{1\}$, then the growth rate of N is polynomially bounded; explicitly, $\mu(N) \preceq p_m$ for $m = \sum_{j=1}^{t-1} jd_j$, where d_j is the rank of the f.g. Abelian group γ_j/γ_{j+1}.*

Proof. (Tits) We follow the rearranging procedure of Definition 17.19ff. and Example 17.19′, taking care to choose a generating set S matching the structure. Namely, we define a function $\nu\colon N \to \{1,\dots,t\}$ by taking $\nu(g)$ to be the maximal j such that $g \in \gamma_j$; i.e., $\nu(1) = t$, and for $g \neq 1$, $\nu(g) = j$ iff $g \in \gamma_j \backslash \gamma_{j+1}$. We take a generating set $S = \bigcup_j S_j$ for which

$$S_j = \{s \in S : \nu(s) = j\} = \{s_{j,1}, \dots, s_{j,n_j}\}$$

is stipulated to satisfy the following properties for each j:

1. $\{\gamma_{j+1}s_{j,1}, \dots, \gamma_{j+1}s_{j,n_j}\}$ generates γ_j/γ_{j+1}.
2. If $s \in S$ and if $s' \in S$, then the group commutator $(s^\varepsilon, s') \in S$ for $\varepsilon = \pm 1$.
3. If $s \in S_j$ and k is the smallest positive integer such that $s^k \in \gamma_{j+1}$ (if such k exists), then $s^k \in S$.

Note that any generating set satisfying (1) can be expanded to one satisfying (2) and (3), since at each stage j there are only a finite number of elements that need to be adjoined. Thus, we can obtain S with these properties.

Also, by Lemma 17.18, if $\nu(s) = i$ and $\nu(s') = j$, then $\nu((s,s')) \geq i + j$.

Given any word $w = s_{j_1,k_1} \cdots s_{j_\ell,k_\ell}$, where each $s_{j_u,k_u} \in S_{j_u}$, we want to keep track of which generators belong to the various j_u. We define the **f-length** of w to be the t-tuple $(\ell_1, \dots, \ell_t)$ such that ℓ_j is the number of letters from w (counting multiplicities) in $\bigcup_{i<j} S_i$. Thus, $\ell_t = \ell = |w|$; by convention, $\ell_j = 0$ for $j \leq 1$.

Take some $s \in S\backslash\{1\}$ with $j_0 = \nu(s)$ minimal possible. We claim, for any number m, that any given word $w = s_{j_1,k_1} \cdots s_{j_\ell,k_\ell}$ of f-length $(\ell_1, \dots, \ell_t)$, in which $s^{\pm 1}$ appears at least m times, can be rearranged to a word w' of f-length $(\ell'_1, \dots, \ell'_t)$, equal to w in N, having the same number of occurrences of $s^{\pm 1}$, such that the last m letters appearing in w' are s or s^{-1}, and

$$\ell'_j \leq \ell_j + m\ell_{j-j_0} + \binom{m}{2}\ell_{j-2j_0} + \binom{m}{3}\ell_{j-3j_0} + \cdots = \sum_{u\geq 0}\binom{m}{u}\ell_{j-uj_0}.$$

(This sum clearly terminates, since $\ell_{j-uj_0} = 0$ whenever $uj_0 \geq j$.)

The proof of the claim is by induction on m, the assertion being vacuous for $m = 0$. In general, we apply induction to get a word w'' of f-length $(\ell''_1, \dots, \ell''_t)$, equal to w, such that the last $m-1$ letters appearing in w'' are $s^{\pm 1}$. This means that to obtain the word of the form w' we need to take another occurrence of $s^{\pm 1}$ and move it to the right to meet the other $m-1$ occurrences. We do this by applying Equation (17.5) up to $|w''| - m$ times; more precisely, for each letter $s'' \in S_{j''}$, we have

$$ss'' = (s, s'')s''s; \qquad s^{-1}s'' = (s^{-1}, s'')s''s^{-1}, \tag{17.13}$$

where, by Lemma 17.18, the new commutator (s^ε, s'') is in $\gamma_{j''+j_0}$. We want $(s^\varepsilon, s'') \in j$; this can only happen when $j'' + j_0 \leq j$ iff $j'' \leq j - j_0$; i.e., the number of new letters from S_j, obtained by adjoining commutators from applying (17.13), is at most ℓ''_{j-j_0}. Hence, obtaining w' from w'' and applying induction yields

$$\begin{aligned}\ell'_j &\leq \ell''_j + \ell''_{j-j_0}\\ &\leq \sum_u \binom{m-1}{u}\ell_{j-uj_0} + \sum_u \binom{m-1}{u}\ell_{j-(u+1)j_0}\\ &= \sum_u \binom{m-1}{u}\ell_{j-uj_0} + \sum_u \binom{m-1}{u-1}\ell_{j-uj_0} = \sum_u \binom{m}{u}\ell_{j-uj_0},\end{aligned}$$

as desired.

Having proved the claim, we take some $\tilde{\ell} \in \mathbb{N}$ such that $\ell_j \leq \tilde{\ell}^j$ for each $1 \leq j \leq t$. Then taking m such that $\ell_{j_0} \leq m \leq \tilde{\ell}^{j_0}$, we see that $\binom{m}{j} \leq m^j \leq \tilde{\ell}^{j_0 j}$ for each j; hence, each element of f-length $(\ell_1, \dots, \ell_t)$ can be rewritten as $\tilde{w}s^{\pm m'}$ where $\tilde{w}$ is a word generated by $S \setminus \{s\}$ of f-length $(\tilde{\ell}_1, \dots, \tilde{\ell}_t)$, with $\tilde{\ell}_j \leq t\tilde{\ell}^{j_0 j} - m'$ and $0 \leq m' \leq m$. (Since the claim permits both s and s^{-1} to appear, one could possibly get cancellation in the power of s at the end.)

To prove the theorem, we open a second induction on $|S_{j_0}|$. Consider the subgroup H generated by $S \setminus \{s\}$. By induction, since we removed an element of degree j_0, the conclusion of the theorem holds for H. We consider two cases.

CASE I. H_{j_0}/H_{j_0+1} is Abelian of rank $\leq d_{j_0} - 1$. Then, there is a constant c for which the number of elements in H of length $\tilde{\ell}$ is at most $c\,\tilde{\ell}^{(\sum_j jd_j) - j_0}$. But the number of choices for $s^{\pm m'}$ is at most $2\tilde{\ell}^{j_0} + 1$, so the number of possible elements of N of length $\tilde{\ell}$ is at most

$$c\,\tilde{\ell}^{(\sum_j jd_j)-j_0}(2\tilde{\ell}^{j_0} + 1) \leq 3c\,\tilde{\ell}^{\sum_j jd_j}.$$

CASE II. H_{j_0}/H_{j_0+1} is Abelian of rank d_{j_0}. Then $[N{:}H] < \infty$, so $s^k \in H$ for some k. Thus, any $g \in N$ can be written in the form hs^m for some $0 \leq m < k$, and the result for N follows at once from the result for H. □

The explicit formula is in fact an equality; cf. Exercise 44. It follows that any virtually nilpotent group has polynomial growth of degree $\sum_j jd_j$. In the other direction, we say a group is **virtually solvable** if it has a solvable group of finite index.

THEOREM 17.61 (MILNOR-WOLF). *Any f.g. virtually solvable group of subexponential growth is virtually nilpotent.*

This theorem is proved after some preparation.

LEMMA 17.62. *For any f.g. group G, $\mu(G) \sim \mu(H)$ for each subgroup H of finite index.*

Proof. By Proposition 00.3, H contains a subgroup N of finite index that is normal in G. Clearly $\mu(N) \preceq \mu(H) \preceq \mu(G)$, so it suffices to prove that $\mu(G) \preceq \mu(N)$. N is f.g., by Remark 00.1.

We borrow the argument of Remark 17.44(iii). Take a generating set $S = \{1, a_2, \dots, a_\ell\}$ of N, and write $G = \bigcup_{j=1}^{t} Ng_j$. Then $g_j g_k = a_{jk} g_u$ for suitable $a_{jk} \in N$ and $u = u(j,k) \in \{1, \dots, t\}$, and likewise $g_j a_i = \tilde{a}_{ij} g_j$ for suitable $\tilde{a}_{ij} \in N$. Let

$$\hat{S} = S \cup \{a_{jk}, \tilde{a}_{ij} : 1 \le i \le \ell,\ 1 \le j,k \le t\},$$

obviously also a generating set of N. Then clearly

$$(\hat{S} \cup \{g_1, \dots, g_t\})^n \subseteq \hat{S}^n \cup \bigcup_{j=1}^{t} \hat{S}^{n-1} g_j,$$

so $|(\hat{S} \cup \{g_1, \dots, g_t\})^n| \le |\hat{S}^n| + t|\hat{S}^{n-1}| \le (t+1)|\hat{S}^n|$. We conclude that $\mu(G) = [\mathfrak{g}_{\hat{S} \cup \{g_1, \dots, g_t\}}] \preceq [\mathfrak{g}_{\hat{S}}] = \mu(N)$. □

(Lemma 17.62 could also be proved by passing to the group algebra; cf. Exercise 36.)

LEMMA 17.63. *Suppose G is a f.g. group of subexponential growth. If N is a normal subgroup, with G/N cyclic, then N is f.g.*

Proof. This is clear by Remark 00.1 if $[G : N] < \infty$, so we assume that G/N is infinite cyclic. Take $a \in G$ such that Na generates G/N. In particular, $a^k \notin N$ for each $k \neq 0$.

Take $a_1, \dots, a_\ell \in N$ such that $S = \{a, a_1, \dots, a_\ell\}$ generates G, and let $\tilde{N}$ be the subgroup of N generated by $\{a^j a_i a^{-j} :\ j \in \mathbb{N},\ 1 \le i \le \ell\}$. We claim that $\tilde{N} = N$. Indeed, if $g \in N$, then writing g as a product of powers of a and the $a^j a_i a^{-j}$, we can always move the powers of a to the right, since

$a^j a_i a^u = (a^j a_i a^{-j}) a^{u+j}$. Thus, we can write $g = wa^k$ where $w \in \tilde{N}$, and get $a^k = w^{-1} g \in N$, implying $k = 0$; hence, $g = w \in \tilde{N}$.

Now define $N_k = \langle a^j a_i a^{-j} :\ j \le k,\ 1 \le i \le \ell \rangle$. We claim that $N = N_k$ for some k. Otherwise, for each k, there is $b_k = a^k a_{i_k} a^{-k} \in N_k \backslash N_{k-1}$. There are 2^k products of the form $b_1^{\epsilon_1} \cdots b_k^{\epsilon_k}$ with each $\epsilon_u \in \{0, 1\}$, all of which are distinct, for if

$$b_1^{\epsilon_1} \cdots b_k^{\epsilon_k} = b_1^{\epsilon'_1} \cdots b_k^{\epsilon'_k},$$

then $b_k^{\epsilon_k - \epsilon'_k} = (b_1^{\epsilon_1} \cdots b_{k-1}^{\epsilon_{k-1}})^{-1} b_1^{\epsilon'_1} \cdots b_{k-1}^{\epsilon'_{k-1}} \in N_{k-1}$. But $b_k^{\pm 1} \notin N_{k-1}$, so we cannot have $\epsilon_k \neq \epsilon'_k$. Hence, $\epsilon_k = \epsilon'_k$, and by induction the other $\epsilon_u = \epsilon'_u$. Furthermore, each of these products are words of length $\le 3k+1$; hence, the growth function $\mathfrak{g}_S(3k+1) \ge 2^k$, contrary to G having subexponential growth. Thus, for some k, $N = N_k$, which is f.g. □

LEMMA 17.64. *If G is a f.g. group of subexponential growth, then G' is f.g.*

Proof. Since G/G' is f.g. Abelian, there is a chain of normal subgroups $G = G_0 \supset G_1 \supset \cdots \supset G_m = G'$, with each G_i/G_{i+1} cyclic. By induction applied to Lemma 17.63, each G_i is f.g., so we conclude that G' is f.g. □

A solvable group G is called **polycyclic** if there is a descending subnormal chain of subgroups

$$G = G_0 \supset G_1 \supset \cdots \supset G_t, \tag{17.14}$$

with G_i/G_{i+1} cyclic, for each i.

Any f.g. Abelian group is polycyclic, by Theorem 2.77 of Volume 1.

PROPOSITION 17.65. *If a f.g. group G is solvable of subexponential growth, then G is polycyclic.*

Proof. By induction on the degree of solvability. By Lemma 17.63, G' is f.g., clearly of subexponential growth, and so by induction is polycyclic, and G/G' is f.g. Abelian and thus polycyclic. Hence G is polycyclic. □

Proof of Theorem 17.61. By definition, G has a solvable subgroup H of finite index, which is f.g. by Remark 00.1, and by Lemma 17.62 has the same growth rate as G. Thus, we may assume that G is solvable. In view of Proposition 17.65, G is polycyclic. Taking the subnormal chain (17.14), we induct on t; if $t = 1$, then G is f.g. Abelian, and thus nilpotent.

In general, we see by induction that G_1 is virtually nilpotent. Thus, G_1 has a nilpotent subgroup N of finite index, and in view of Proposition 00.3, we may assume that $N \lhd G$. If $[G : G_1] < \infty$, then G is virtually

nilpotent, so we may assume that $[G:G_1]$ is infinite. Take $a \in G$ such that $G_1 a$ generates G/G_1. Then $N\langle a\rangle$ has finite index in G, so we may assume that $G = N\langle a\rangle$ and $G_1 = N$.

Each characteristic subgroup of N is normal in G. Thus, we can refine the lower central series of N to a series of normal subgroups of G of the form

$$N = N_0 \supset N_1 \supset \cdots \supset N_v = \{1\},$$

having the property that if L is a normal subgroup of G between N_{i+1} and N_i, then $[N_i : L] < \infty$. Conjugation by a yields an automorphism σ_i of N_i/N_{i+1}.

We claim that each σ_i has some finite order m_i; then, taking $m = m_1 \cdots m_{v-1}$, one would have the commutator set $(a^m, N_i) \subseteq N_{i+1}$ for each i, implying that $N\langle a^m\rangle$ is a nilpotent subgroup of G of finite index $\leq m$, concluding the proof.

So it remains to prove the claim. Write $\sigma = \sigma_i$. Let $A = N_i/N_{i+1}$, an Abelian group of subexponential growth, and form $A_{\mathbb{Q}}$ as in Definition 00.4. By Remark 00.5, $A_{\mathbb{Q}}$ is a vector space over $\mathbb{Q}$, invariant under σ; and by the hypothesis of the last paragraph, $A_{\mathbb{Q}}$ has no σ-invariant $\mathbb{Q}$-subspaces (since this would yield a normal subgroup between L_i and L_{i+1}).

Consider the set $\mathcal{D}$ of linear transformations $\varphi\colon A \to A$ preserving σ, in the sense that $\varphi(\sigma(c)) = \sigma(\varphi(c))$, $\forall c \in A$. As in Proposition 13.39, any such φ is an isomorphism, implying that $\mathcal{D}$ is a division ring that clearly contains σ itself. Let K be the subfield of $\mathcal{D}$ generated by $\mathbb{Q}$ and σ; then A can be viewed as a module over the subring $\mathbb{Z}[\sigma]$ of K, where addition is the group operation of A. Since $\sigma A = A$, we see by Theorem 5.21 of Volume 1 that σ is integral over $\mathbb{Z}$. The claim clearly holds unless σ has infinite order; Proposition 00.7 then implies that there is some embedding of K into $\mathbb{C}$ such that the image μ of σ has absolute value > 1; replacing a by a suitable power, we may assume that $|\mu| > 2$.

Now pick $c \neq 0$ in A. Since $\sum \mathbb{Q}\sigma^k(c)$ is a σ-invariant $\mathbb{Q}$-subspace of $A_{\mathbb{Q}}$, we get $\sum \mathbb{Q}\sigma^k(c) = A_{\mathbb{Q}}$. Put $b_k = \sigma^k(c) = a^k c a^{-k} \in N_i$. Arguing as in the proof of Lemma 17.63, we consider words of the form $ac^{\epsilon_1}ac^{\epsilon_2}\cdots ac^{\epsilon_m}$ for $\varepsilon_u = \pm 1$, $1 \leq u \leq m$. If these were all distinct, then, with respect to a generating set of G containing a and ac, we would have $\mathfrak{g}_S(m) \geq 2^m$, contrary to G having subexponential growth. Hence, there are $m \leq n$ such that

$$ac^{\epsilon_1}ac^{\epsilon_2}\cdots ac^{\epsilon_m} = ac^{\epsilon'_1}ac^{\epsilon'_2}\cdots ac^{\epsilon'_n} \quad \big((\epsilon_1,\dots,\epsilon_m) \neq (\epsilon'_1,\dots,\epsilon'_n)\big),$$

which can be rewritten as

$$b_1^{\epsilon_1}b_2^{\epsilon_2}\cdots b_m^{\epsilon_m}a^m = b_1^{\epsilon'_1}b_2^{\epsilon'_2}\cdots b_n^{\epsilon'_n}a^n.$$

It follows that $a^{m-n} \in N_i$, implying $m = n$, and since A is Abelian, we have

$$b_1^{\epsilon'_1-\epsilon_1}b_2^{\epsilon'_2-\epsilon_2}\cdots b_n^{\epsilon'_m-\epsilon_m} \in N_{i+1}.$$

Letting $j_k = (\epsilon'_k - \epsilon_k)$ for $1 \leq u \leq m$, we may assume that $j_m > 0$, and get $\sum_{k=1}^m j_k\sigma^k(c) = 0$ on $\sum \mathbb{Q}\sigma^k(c) = A_{\mathbb{Q}}$. But then $j_m\mu^m = -\sum_{k=1}^{m-1} j_k\mu^k$, implying the following inequality (since $|\mu| - 1 > 1$):

$$j_m|\mu|^m \leq \sum_{k=1}^{m-1}|\mu|^k < \frac{|\mu|^m - 1}{|\mu| - 1} < |\mu|^m - 1,$$

a contradiction. The claim is proved. □

(Although each step is elementary, this proof foreshadows the use of a module over a group, namely the group $\langle\sigma\rangle$; this notion plays a very important role in Chapter 19.) Putting things together so far already yields an interesting result.

COROLLARY 17.65′. *Any f.g. virtually solvable group G either has exponential growth or is virtually nilpotent; the latter is true iff G has polynomial growth.*

Curiously enough, a classical theorem of Mal′cev says that any virtually polycyclic group has a subgroup H of finite index for which H' is nilpotent. A nice proof is given in Passman [**Pas**] (using the structure of group algebras).

One important class of groups, to be discussed in Appendix 19A, is the class of **linear** groups, i.e., those groups embeddible into $\mathrm{GL}(n, F)$. Tits proved the **Tits alternative**, given below as Theorem 19B.21, that any f.g. linear group either is virtually solvable or contains a free group; we thereby obtain the same dichotomy of growth for all f.g. linear groups:

COROLLARY 17.66. *Every f.g. linear group of subexponential growth is of polynomial growth.*

Grigorchuk [**Gri2**] has constructed abstract groups of intermediate growth. On the other hand, Gromov [**Grom1**] proved that every group of polynomial growth is virtually nilpotent. Gromov's proof relies heavily on the geometry of the Cayley graph, and has led researchers to search for more elementary proofs. An alternative proof using nonstandard analysis was found by van den Dries-Wilkie [**vdDW**], but still is quite difficult.

Another important measure of growth in a finite group is the number of subgroups of order n, as $n \mapsto \infty$; cf. Lubotzky–Segal [**LuS**].

Appendix 17A. Presentations of groups

We have already defined a presentation $\phi\colon \mathcal{F} \to A$ of any algebraic structure $\mathcal{A}$ in terms of the corresponding free structure $\mathcal{F}$. In this appendix, we study such presentations, focusing on groups and using topological methods.

Finitely presented structures.

Our main concern is $\mathcal{N} = \ker \phi$. For example, if A is finitely generated, with $\mathcal{F}$ also taken to be finitely generated, when is $\mathcal{N}$ finitely generated in the appropriate sense? If so, we say that A is **finitely presented**. From this perspective, finitely presented structures are the ones that should be most amenable to investigation. In the commutative theory, finite presentation comes automatically.

Example 17A.1. (i) Any commutative affine algebra R can be written in the form $F[\lambda_1, \dots, \lambda_n]/I$; by Theorem 7.17 of Volume 1, any ideal I of the polynomial algebra $F[\lambda_1, \dots, \lambda_n]$ is f.g., and thus R is finitely presented as an algebra.

(ii) Any f.g. R-module M has the form $R^{(n)}/K$ where the $K \leq R^{(n)}$. If the ring R is Noetherian, then K is f.g., so M is finitely presented as a module. Thus, for $R = \mathbb{Z}$, any f.g. Abelian group is finitely presented.

(iii) Any finite group of order n is finitely presented, since all the relations obviously are generated by the relations described in its multiplication table of n^2 relations.

Of course, in describing a finitely presented group, we would want to take some "canonical" presentation. There is considerable ambiguity, as seen via a few examples.

Example 17A.2. (i) The (finite) dihedral group

$$D_n = \langle a, b : a^n = 1 = b^2,\ bab^{-1} = a^{n-1} \rangle$$

is also generated by b and $c = ab$; but $c^2 = 1$ and $a = cb$, so the presentation becomes

$$\langle b, c : b^2 = c^2 = 1, (cb)^n = 1 \rangle.$$

This presentation generalizes naturally to infinite groups, and we define the **infinite dihedral group**

$$D_\infty = \langle b, c : b^2 = c^2 = 1 \rangle.$$

D_∞ is solvable, since $\langle cb \rangle$ is a normal subgroup and $D_\infty/\langle cb \rangle$ is cyclic.

From this presentation, we see that any group G generated by two elements b and c of order 2 must be dihedral; indeed, we can use the relations $b^2 = 1$ and $c^2 = 1$ to reduce any word to an alternating product of b and c. Let $w = 1$ be an extra relation with w written in shortest form. It suffices to show that w must be of the form $(bc)^n$ or $(cb)^n$. If, say, $(bc)^n b = 1$ were the relation in shortest form, then also $1 = b(bc)^n = b^2c(bc)^{n-1} = c(bc)^{n-1}$, yielding a shorter relation, contrary to assumption on w.

(ii) The symmetric group S_n is generated by the $n-1$ transpositions $\{\sigma_i = (i\ i{+}1) : 1 \leq i < n\}$. These satisfy the relations $\sigma_i^2 = 1$, $(\sigma_i\sigma_{i+1})^3 = 1$, and $(\sigma_i\sigma_j)^2 = 1$ for $|j - i| > 1$. On the other hand, these relations imply all the relations of the symmetric group; cf. Exercise A1. This approach leads us to the braid group of Exercise 19A.24ff. as well as the Coxeter groups of Chapter 22.

Groups as fundamental groups

Various breakthroughs in the combinatoric theory of groups have been attained by viewing abstract groups as fundamental groups of topological structures; one of several applications is an insightful proof of the Nielsen-Schreier theorem, which states that every subgroup of a free group is free. To see how this proof works, we review some simplicial topology. We outline the theory, relying on some basics of topology with which the reader may already be familiar. The assertions are not particularly difficult, and their verifications are left for Exercises A2–A9. (See Rotman [**Rot1**] for more details.)

Definition 17A.3. We work with a set V of **vertices**. An **n-simplex** is the power set $\mathcal{S} = \mathcal{P}(S)$, where $S \subset V$ is a subset of $n+1$ vertices; a **(simplicial) complex of dimension** n is a union of m-simplices for various m, the largest of which is n.

We have a category whose objects are complexes and whose morphisms $f\colon \mathcal{K}_1 \to \mathcal{K}_2$ are maps such that if S is a simplex of $\mathcal{K}_1$, then $f(S)$ is a simplex of $\mathcal{K}_2$; we say that $f(\mathcal{K}_1)$ is a **subcomplex** of the complex $\mathcal{K}_2$.

For example, any point $\{v\}$ is a 0-simplex, whereas the pair $\{v_0, v_1\}$ is a 1-simplex, often identified with a line. From this point of view, an undirected graph is merely a complex of dimension 1. In general, we identify an n-simplex with the convex hull of its vertices in n-space, viewed thereby as a topological space. The following definition is intuitively clear.

Definition 17A.4. Two simplices are **connected** iff they have a common vertex. By transitivity, we define the **connected component** of a complex; when this is the whole complex, we say that the complex is **connected**.

From now on, we assume that our complex $\mathcal{K}$ is connected. We are ready for the main application to groups. Fixing one vertex v of $\mathcal{K}$, we define **homotopic equivalence** on v-circuits by saying that circuits $p_1 \sim p_2$ if $p_1\bar{p}_2$ is homotopic (read, "can be shrunk") to a point. (We do not give the formal definition of "homotopic," a well-known concept from algebraic topology, but we interpret it in the next two examples.) The equivalence classes are called **homotopy classes**. The **fundamental group** $\pi_1(\mathcal{K}, v)$ is defined to be the set of homotopy classes of circuits, with the group operation coming from composition of paths.

Example 17A.5. In the case when $\dim \mathcal{K} = 1$, i.e., $\mathcal{K}$ is an (undirected) graph, homotopic equivalence is the same as the equivalence of Definition 17.24, defined via tacking on or deleting trivial paths. For example, the fundamental group of a tree is trivial, whereas a nontrivial circuit is *not* homotopic to a point.

The fundamental group of any finite graph $\mathcal{K}$ can be seen easily to be a free group, generated by the "minimal" nontrivial circuits. Explicitly, if T is a maximal tree in $\mathcal{K}$, then adjoining any edge in $\mathcal{K} \setminus T$ produces a new circuit, and thus the rank of the free group $\pi_1(\mathcal{K}, v)$ is precisely the number of edges in $\mathcal{K} \setminus T$.

Conversely, we define the n-**bouquet** to be the union of n v-circuits that are disjoint except at v. The fundamental group of the n-bouquet is the free group on n generators. Thus we have classified f.g. free groups as fundamental groups of finite graphs.

Example 17A.6. More generally, in any complex $\mathcal{K}$, homotopic equivalence still has a clear-cut description, although more complicated than the definition of equivalence in Definition 17.24. Paths $p = p(a, b)$ and $q = q(a, b)$ are homotopically equivalent if $p\bar{q}$ lies in a simplex of $\mathcal{K}$; if we replace a sub-path of a circuit c by an equivalent path, then the new circuit is also homotopically equivalent to c.

Along the same lines as in Example 17A.5, but with more effort, one sees in Exercise A9 that any finitely presented group is the fundamental group of a complex of dimension 2. So the study of arbitrary finitely presented groups is thereby reduced to the combinatorics of two-dimensional complexes. These ideas gain depth with the introduction of another notion.

Definition 17A.7. A **covering complex** of a complex $\mathcal{K}$ is a (connected) complex $\tilde{\mathcal{K}}$ together with an onto simplicial map $\phi\colon \tilde{\mathcal{K}} \to \mathcal{K}$ such that the inverse image $\phi^{-1}(\mathcal{S})$ of any simplex $\mathcal{S}$ of $\mathcal{K}$ is a finite disjoint union of simplices isomorphic to $\mathcal{S}$, called the **sheets** of $\mathcal{S}$.

In particular, $\tilde{\mathcal{K}}$ has the same dimension as $\mathcal{K}$. The map $\phi\colon \tilde{\mathcal{K}} \to \mathcal{K}$ induces a group homomorphism

$$\phi_{\#}\colon \pi_1(\tilde{\mathcal{K}}, v) \to \pi_1(\mathcal{K}, v).$$

Since each sheet in $\tilde{\mathcal{K}}$ is locally isomorphic to its image in $\mathcal{K}$, we cannot lose any circuits, so we see at once that $\phi_{\#}$ is 1:1. (On the other hand, ϕ^{-1} can "pull apart" circuits to paths connecting different points in the preimage of v, so we would not expect $\phi_{\#}$ to be onto.) This reasoning yields

Remark 17A.8. $[\pi_1(\mathcal{K}, v) : \phi_{\#}\pi_1(\tilde{\mathcal{K}}, v)] = |\phi^{-1}(v)|$.

Given a subgroup G of $\pi_1(\mathcal{K}, v)$, one can construct a covering complex $\phi\colon \tilde{\mathcal{K}} \to \mathcal{K}$ such that $\phi_{\#}\pi_1(\tilde{\mathcal{K}}, v) = G$; cf. Exercise A9. This paves the way for a celebrated theorem.

THEOREM 17A.9 (NIELSEN-SCHREIER). *Every subgroup G of a free group $\mathcal{G}$ is free.*

Proof. By Example 17A.5, there is a complex $\mathcal{K}$ such that $\mathcal{G} = \pi(\mathcal{K}, v)$. Taking a covering complex $\phi\colon \tilde{\mathcal{K}} \to \mathcal{K}$ with $\phi_{\#}\pi(\tilde{\mathcal{K}}, v) = G$, one notes that $\dim \tilde{\mathcal{K}} = 1$, so G is free, again by Example 17A.5. □

If the subgroup G has finite index, then G is f.g. by Remark 00.1; its precise rank is given in Exercise A7. Another powerful method for studying groups (especially finitely presented groups) via graph theory, discovered by Stallings, is described briefly in Exercises A10–A14.

Appendix 17B. Decision problems and reduction procedures

In Example 13.49 we represented algebras explicitly as subrings of endomorphism rings, and used this description in the subsequent chapters as a springboard for Jacobson's structure theory. In this appendix, we take a more intrinsic approach to presentations of algebraic structures, leading directly to basic questions concerning the effectiveness of a presentation. Just as in the commutative case (Appendix 7A of Volume 1), we should like to obtain an explicit, computable description of an arbitrary algebra by a careful study of its generators and relations. Although the process formalized by Newman [**Ne**] and refined by Bergman [**Berg**] is rather complicated at every stage and is unmanageable at times, it often is surprisingly effective.

Definition 17B.1. For any partially ordered set $(A, <)$ satisfying the minimum condition, a **reduction procedure** $\mathcal{R}$ on A is a set of **reduction functions** $\rho\colon A \to A$ satisfying $\rho(r) \leq r$ for each $r \in A$.

A **chain of reductions** of an element $r \in A$ is a sequence

$$r, \quad \rho_1(r), \quad \rho_2\rho_1(r), \quad \dots$$

for various $\rho_i \in \mathcal{R}$; r is **irreducible** if $\rho(r) = r$ for all $\rho \in \mathcal{R}$.

For any reduction functions $\rho_1, \rho_2, \dots$ and any $r \in A$, the chain of reductions $r, \rho_1(r), \rho_2\rho_1(r), \dots$ decreases in A and thus must stabilize. Consequently, any element $r \in A$ can be reduced to an irreducible element r' after a finite number of reductions; r' is called **reduction-final** for r.

Unfortunately, reduction procedures are fraught with ambiguity, since we might be able to apply different reductions to r, and a priori, we do not know which initial reduction is "best." In other words, although we arrive at some reduction-final element, we might have been able to get something even lower by changing the order of the reductions and perhaps applying different reductions. This raises the fundamental question:

Reduction Question 17B.2. Do all chains of reductions on r eventually arrive at the same reduction-final element?

In that case, we call r **reduction-unique**. A **reduction-unique** procedure is a reduction procedure with respect to which each element of A is reduction-unique.

One major obstacle to understanding reduction procedures is that although the reduction relation is antisymmetric, we want to view it somehow as an equivalence, as illustrated in the next example.

Example 17B.3. Let $A = \{a_1, a_2, \dots\}$ be a countable set. We consider reduction functions of the form $\rho_{i,j}$, which means we replace a_i by a_j and leave all other elements fixed; our reduction procedure $\mathcal{R}$ is a collection of certain $\rho_{i,j}$ for $i \geq j$. This procedure need not be reduction-unique; for example, $\mathcal{R}$ could contain both $\rho_{5,3}$ and $\rho_{5,2}$ but not $\rho_{3,2}$, leading to ambiguity in reducing a_5.

The obvious remedy is to extend ρ first to a symmetric relation and then by transitivity to an equivalence relation; then we define the extended reduction procedure $\hat{\mathcal{R}}$ by admitting $\rho_{i,j}$ whenever $a_i \sim a_j$ for $i \geq j$. $\hat{\mathcal{R}}$ is reduction-unique, since the reduction-final elements are merely the smallest representatives of their equivalence classes. In our example, we would have a_3 equivalent to a_2 via

$$a_3 \sim a_5 \sim a_2,$$

and would introduce the new reduction $\rho_{3,2}$.

Unfortunately, a serious difficulty lies under the surface, now well-known from cryptography — in practice, the reverse of the reduction procedure $\mathcal{R}$ might be difficult to ascertain. In order to reduce a_3 we had to cope with an ambiguity on a_5. In this way, we see that in order to reduce a_i we might need to cope with ambiguous reductions on a_j for $j > i$, perhaps arbitrarily large. Although there are only countably many reductions, we might have to search arbitrarily many of them before being able to define the new reduction that we need; in practice, after any given number of steps, we might never be sure whether some given a_i is irreducible.

At least when A is a finite set, we have only finitely many possible reductions, so we can always refine a reduction procedure to a reduction-final procedure (although the implementation could be unwieldy). But for A infinite, we might have no effective way of determining all new reductions.

Reduction procedures on natural numbers.

The easiest nontrivial example is $\mathbb{N}$; a reduction procedure would be a non-increasing function $\rho\colon \mathbb{N} \to \mathbb{N}$. The first difficulty is to provide a suitable definition of "computable function" from $\mathbb{N}$ to $\mathbb{N}$. This was done formally by Kleene, by means of **recursive functions**, which we describe only briefly, since the subject pertains more to set theory. Intuitively, one would want to define a recursive function f on n (using the operations of arithmetic) in terms of $f(n-k), \dots, f(n-1)$ for suitable k. For example, the famous Fibonacci function f is defined recursively as $f(0) = 0$, $f(1) = 1$, and $f(n) = f(n-1) + f(n-2)$ for all $n \geq 2$. Such functions are called **primitive recursive functions**. Clearly the set of primitive recursive functions is enumerable. This class does not contain all the functions we want, and recursive functions are defined more generally in terms of existential elementary formulas, in a language that includes the successor function $n \mapsto n+1$. One can also show that the set of recursive functions is enumerable. An alternative, more explicit way of defining recursive functions is through Turing machines, which provide a way of "computing" $f(n)$ in finite time, from $f(n-k), \dots, f(n-1)$.

Definition 17B.4. A set $S \subseteq \mathbb{N}$ is **recursively enumerable** (r.e.) if it is the range of a suitable recursive function. A set $S \subseteq \mathbb{N}$ is **recursive** if both S and $\mathbb{N} \setminus S$ are r.e. For example, the set of even natural numbers is recursive.

The motivation for the definition is as follows: Given a recursive reduction procedure ρ, let $S = \{\rho(n) : n \in \mathbb{N}\}$. We have the problem of computing whether some given $n \in S$. We could try computing $\rho(1), \rho(2), \dots$, until n showed up as $\rho(k)$ for some k. Although this should happen eventually if

indeed $n \in S$, if after arbitrarily many evaluations we do not come up with k such that $n = \rho(k)$, we still do not know whether $n \notin S$. But if S is recursive, then $\mathbb{N} \setminus S$ is the range of some recursive function ψ, so by alternating evaluations for ρ and ψ and performing enough evaluations, we find out for sure whether $n \in S$. Thus, decision problems can only be solved effectively for reduction procedures whose images are recursive sets.

There exists an r.e. set S that is not recursive, and consequently the word problem has a negative answer for sets. This theorem, proved independently by Post, Markov, and Turing, is very well-known because it is one of the pillars of computer science. A thorough treatment can be found in Bridges [**Bridg**]. (See Mal′cev [**Mal, Theorem 6.3**] for a mathematically rigorous treatment; the argument is a modification of the kind of the self-referential argument introduced by G. Kantor. See Hofstadter [**Hof**] for a very readable account.)

Reduction procedures on monoids.

We turn to richer structures, hoping that the extra algebraic structure might give us a better hold on the reduction procedure. We start with f.g. monoids, since the free f.g. monoid $\mathcal{M}$ is so easy to describe (cf. Definition 17.2), and the crossword dictionary order satisfies the minimum condition by Lemma 17.10.

Suppose for example that we have the reduction $x_2^2 \mapsto x_1x_2$ in $\mathcal{M}$. How do we apply this to x_2^3? (Read as $x_2^2x_2$, it reduces to $(x_1x_2)x_2$; read as $x_2x_2^2$, it reduces to $x_2(x_1x_2)$.) In order for our reduction system to be well-defined, we treat $x_2^2 \mapsto x_1x_2$, $x_2^3 \mapsto x_1x_2^2$, and $x_2^3 \mapsto x_2x_1x_2$ as separate reductions. In other words, we define **basic reductions** that fix all but one word of $\mathcal{M}$; moreover, we stipulate that if we have a basic reduction $w \mapsto v$, then we also have the basic reductions $awb \mapsto avb$ for all words a, b.

With this proviso in mind, we now can apply the same considerations from recursion theory. Any reduction procedure ρ on $\mathbb{N}$ can be used to formulate a reduction procedure on the free monoid, and vice versa. But the reduction procedure is effective only when the range of ρ is recursive. Thus, any nonrecursive r.e. set yields a noneffective reduction procedure on the free monoid.

Famous decision problems.

So far, our discussion has been at a rather abstract level. In order to proceed, we pose some specific problems arising from these considerations:

The Word Problem. Is there an "effective" procedure for determining whether two elements of a given algebraic structure $\mathcal{F}/\mathcal{I}$ are equal?

The Isomorphism Problem. Is there an "effective" procedure for determining whether two given algebraic structures (defined in terms of generators and relations) are isomorphic?

Back in 1912, Dehn [**Deh**] discussed presentations of groups and posed these two fundamental decision questions, as well as a third:

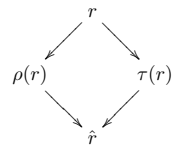
The Conjugacy Problem. Are two given elements of a group conjugate?

Of course the word problem is trivial for the free monoid, since by definition, every two distinct words are different. But the situation changes drastically when we introduce relations, which can be viewed as identifications of words. As explained above (proved independently by Markov and Post), the word problem is unsolvable for monoids in general, even when the relation congruence is finitely generated.

The Diamond Lemma.

To have hope in handling decision problems, we need to formulate a method of arriving at a reduction-unique procedure. We start with a necessary and sufficient condition, due to Newman, for a reduction procedure to be reduction-unique.

Proposition 17B.5 (The Diamond Lemma). *A reduction procedure $\mathcal{R}$ is reduction-unique on A iff for each $r \in A$ and any reductions $\rho, \tau \in \mathcal{R}$, the elements $\rho(r)$ and $\tau(r)$ have chains of reductions arriving at the same element $\hat{r}$. In other words, we have the "diamond" condition*



Proof. Clearly, a reduction-unique procedure satisfies the diamond condition. We prove the other direction by induction on the minimal length m of a chain of reductions

$$r, \quad \rho_1(r), \quad \rho_2\rho_1(r), \quad \ldots, \quad r' = \rho_m \cdots \rho_2\rho_1(r),$$

from r to a reduction-final element r'. The assertion is obvious for $m = 0$, since then r itself is reduction-final. In general, we apply the diamond diagram to $\rho_1(r)$ and an arbitrary reduction $\rho(r)$ to get a common reduction $\hat{r}$.

But $\rho_1(r)$ has a chain of reductions of length $m-1$ to r', so by induction $\rho_1(r)$ is reduction-unique, and thus $\hat{r}$ reduces to r'. □

Let us rephrase the condition of the Diamond Lemma. The reduction procedure is called **ambiguous** on r if there are reductions ρ_1, ρ_2 such that $\rho_1(r) \neq \rho_2(r)$. The diamond condition states that even when the reduction procedure is ambiguous on r, there is a common reduction, which thus **resolves** the ambiguity. Let us gain intuition by seeing how this idea applies to specific algebraic theories.

Reduction procedures on the free group.

For our first application, we apply the Diamond Lemma to describe the free group itself (as an image of the free monoid).

Example 17B.6 (The free group revisited). We fix a generating set $\{x_i : i \in I\}$. Most (but not all) of the time, we want to construct the free group $\mathcal{G}$ on n generators, in which case I is a finite set $\{x_1, \dots, x_n\}$.

We start with the free monoid $\mathcal{M}$ in an alphabet $\{x_i, y_i : i \in I\}$. We want to make Example 17.15 more rigorous, introducing relations such that $y_i = x_i^{-1}$ for each i. The obvious basic reductions are $ax_iy_ib \mapsto ab$ and $ay_ix_ib \mapsto ab$ for all $a, b \in \mathcal{M}$. We claim that the ensuing reduction procedure is reduction-unique, according to Proposition 17B.5. The only possible ambiguities arise for words having one of the following forms:

(1) $w_1x_iy_ix_iw_2$;
(2) $w_1y_ix_iy_iw_2$;
(3) $w_1x_iy_iw_2x_jy_jw_3$;
(4) $w_1x_iy_iw_2y_jx_jw_3$;
(5) $w_1y_ix_iw_2x_jy_jw_3$;
(6) $w_1y_ix_iw_2y_jx_jw_3$.

In (1) we could reduce x_iy_i to 1 or y_ix_i to 1, but we get the same result $w_1x_iw_2$, thereby resolving any ambiguity; likewise with (2). (3) has a genuine ambiguity, since reducing x_iy_i to 1 yields $w_1w_2x_jy_jw_3$, whereas reducing x_jy_j to 1 yields $w_1x_iy_iw_2w_3$. However, both have the common reduction to $w_1w_2w_3$, as desired. The same argument holds for (4), (5), and (6). Thus the diamond condition is verified, and we have a reduction-unique procedure. Its effect is to create inverses for the generators in the most general possible way, thereby providing us with the free group $\mathcal{G}$; namely, any group generated by $|I|$ generators can be written naturally as a homomorphic image of $\mathcal{G}$.

The free group $\mathcal{G}$ is harder to work with than the free monoid $\mathcal{M}$, precisely because of the extra relations $x_iy_i = 1 = y_ix_i$ that we have imposed. For example, in the language of monoids we have tacitly assumed that the

reduction $w \mapsto \rho(w)$ also entails the reductions $awb \mapsto a\rho(w)b$ for all a, b. But since we are working in the free group, taking $a = w^{-1}$ would force us to accept the reduction $w^{-1}wb \mapsto w^{-1}\rho(w)b$. Since $b = w^{-1}wb$ in $\mathcal{G}$, we would always have a reduction $b \mapsto w^{-1}\rho(w)b$, leading to the paradox that no element of the free group is irreducible. Thus, it is preferable to backtrack and do our reduction procedure only in the free monoid $\mathcal{M}$, permitting the reduction $w'wb \mapsto b$ in $\mathcal{M}$ whenever $w'w$ reduces to 1 in $\mathcal{G}$. (This leaves $w'wb$ ambiguous, since it reduces both to b and $w'\rho(w)b$, but we can resolve this ambiguity by introducing the reductions corresponding to $w'\rho(w) \mapsto 1$ for each word w.)

From this point of view, we can take any subset $\mathcal{R}$ of the free group $\mathcal{G}$ and use it to define reductions $asb \mapsto ab$ for all $a, b \in \mathcal{G}$, all $s \in \mathcal{R}$. We let G denote the resulting set of irreducible elements, made into a group by taking the product (in $\mathcal{G}$) and then reducing. Let $\mathcal{N}$ denote the normal subgroup of $\mathcal{G}$ generated by $\mathcal{R}$. By Noether's isomorphism theorem, the natural homomorphism $\mathcal{G} \to G$ induces an onto homomorphism $\mathcal{G}/\mathcal{N} \to G$. But each element of $\mathcal{G}/\mathcal{N}$ is represented by an element of G, proving that $G \cong \mathcal{G}/\mathcal{N}$.

The word problem for groups.

The word problem for groups is equivalent to determining when a word w is trivial, since $a = b$ iff $ab^{-1} = 1$. In his paper, Dehn found a class of groups in which the word problem is solvable.

Write $|w|$ for the length of a word w; for example, $|x_1x_2x_1^{-1}x_2| = 4$. A group $G = \mathcal{G}/\mathcal{N}$ is said to satisfy **Dehn's algorithm** if there are words $u_1, v_1, \dots, u_n, v_n$ with $|u_i| < |v_i|$ such that $\{u_1^{-1}v_1, \dots, u_n^{-1}v_n\}$ generate $\mathcal{N}$, and for each w in $\mathcal{N}$, either w or w^{-1} contains some v_i as a subword.

PROPOSITION 17B.7. *The word problem is solvable in any group G satisfying Dehn's algorithm.*

Proof. Write $G = \mathcal{G}/\mathcal{N}$ as above. Suppose $a \in G$ is the image of a word w. We proceed by induction on $|w|$; if $|w| = 0$, then $w = 1$ and there is nothing to prove, so assume that $|w| > 0$. If none of the v_i is a subword of w or w^{-1}, then $a \neq 1$; thus, replacing w by w^{-1} if necessary, we may assume that some v_i is a subword, and we write $w = w'v_iw''$. But then we reduce w to $w'u_iw''$, thereby decreasing the length, and conclude by induction. □

On the face of it, Dehn's algorithm seems rather naive. Nevertheless, it has turned out to be surprisingly versatile.

Digression 17B.8: Hyperbolic groups. Rips and Gromov [**Grom2**] introduced an important class of groups which, remarkably, satisfy Dehn's

algorithm (and thus have a positive answer to the word problem). The definition arises from consideration of the Cayley graph of a group (Example 17.29).

A shortest path joining vertices a and b is called a **geodesic**. This might not be unique; for example, for G Abelian, if $i \neq j$, there are at least two geodesics between 1 and g_ig_j, obtained by passing either through g_i or through g_j respectively. Any graph has a metric associated with it, where the **distance** $d(a, b)$ is the length of a geodesic connecting a and b. For example, by our convention of Remark 17.29′, $d(g_i, g_j) > 1$ for $i \neq j$; in fact, $d(g_i, g_j) = 2$ since $g_j = g_i g_i^{-1} g_j$. Obviously,

$$d(a, b) + d(b, c) \geq d(a, c)$$

for all a, b, c in G.

If S is a set of vertices and v is any vertex, define

$$d(v, S) = \min\{d(v, s) : s \in S\}.$$

Thus, $d(v, S) = 0$ iff $v \in S$.

Given three vertices a, b, and c, one defines the **triangle** formed by three geodesics: from a to b, from b to c, and from c to a. Define the **thickness** of the triangle to be the maximal distance from a vertex on one geodesic to the closest vertex on the union of the other two geodesics. For example, a tree has thickness 0. Intuitively, the greater the curvature of space, the thicker the triangles, thereby leading us to the following important definition:

A group G is **hyperbolic** if there is some number δ such that every triangle of its Cayley graph has thickness $< \delta$.

Obviously, any finite group is hyperbolic, since its graph only has finitely many triangles. Free groups are hyperbolic, since their graphs are trees. But hyperbolic groups encompass much more. Bridson-Haefliger [**BridsH**] is a good general reference that describes geometric and topological methods used in investigating hyperbolic groups.

To produce a finitely presented group in which the word problem is unsolvable, one would like to define a finite set of relations on a free group and show that the previous r.e. versus recursive argument would yield a suitable example. Unfortunately, the verifications become very difficult. The first examples of finitely presented groups having a negative solution to the word problem were discovered independently by Novikov (1955) and Boone (1954–57); cf. Boone [**Boo**]. Boone showed that an example of Post can be embedded into a finitely presented group for which the word problem is unsolvable, via a well-known construction in group theory called the HNN

construction; details can be found in Rotman [**Rot1**], the standard introductory reference for decision problems on groups. Other examples can be culled from the papers in the collection Boone-Cannonito-Lyndon [**BooCL**].

Reduction procedures on algebras: Bergman's Method.

Having utilized Newman's Diamond Lemma to analyze monoids (and also having noted its limitations), let us apply these ideas to associative algebras. We work in the free associative algebra $F\{X\}$ over a field, thereby enabling us to normalize polynomials so that the leading coefficient is 1; the same general setup could be made for algebras over an arbitrary commutative ring, but the reduction procedure is messier.

As in Appendix 7A of Volume 1, for any $N \lhd F\{X\}$, we write $\widehat{N}$ for the (graded) ideal of $F\{X\}$ generated by the set of leading monomials $\{\widehat{f} : f \in N\}$. Although the situation is now vastly more complicated than in Example 7A.2, Proposition 7A.3 of Volume 1 generalizes (with the same proof) to the following result:

PROPOSITION 17B.9. *Suppose $K \leq N$ are ideals of $F\{X\}$. Then $K = N$ iff $\widehat{K} = \widehat{N}$.*

Suppose we are trying to describe an algebra $R = F\{X\}/N$, where $N \lhd F\{X\}$. Given a set of generators $\{f_i : i \in I\}$ of N, we write $f_i = \widehat{f}_i + \widetilde{f}_i$, where $\widehat{f}_i$ is the leading monomial of f_i, and define the reduction

$$\rho_i \colon \widehat{f}_i \mapsto -\widetilde{f}_i.$$

Note that ρ_i replaces the highest monomial of f_i by a linear combination of smaller monomials.

In order to be precise, we define ρ_i as the F-linear map of $F\{X\}$ sending $\widehat{f}_i \mapsto -\widetilde{f}_i$ and leaving all other monomials fixed. Thus for any words u, v, we also define the reduction $u\rho_i v$ corresponding to uf_iv, given by the F-linear map sending $u\widehat{f}_i v \mapsto -u\widetilde{f}_i v$ and leaving all other monomials fixed.

Remark 17B.10. To apply the Diamond Lemma, one need only check the diamond condition on the leading monomials, since the reductions are linear maps.

Definition 17B.11. An **inclusion ambiguity** consists of relations f_i, f_j, where $\widehat{f}_i = u\widehat{f}_j v$. An **overlap ambiguity** consists of relations f_i, f_j where $\widehat{f}_i = uv$ and $\widehat{f}_j = vw$; we denote the ambiguity as uvw.

According to the Diamond Lemma and Remark 17B.10, we need only be concerned when a given monomial h can be reduced in two different ways. This could happen in the following situations:

1. $h = g_1 \widehat{f}_i g_2 \widehat{f}_j g_3$;
2. $h = g_1 \widehat{f}_i g_2$ where there is an inclusion ambiguity $\widehat{f}_i = u \widehat{f}_j v$;
3. $h = g_1 uvw g_2$ where f_i and f_j have the overlap ambiguity uvw.

Remark 17B.12. In the list above, (1) is not ambiguous, since

$$\rho_j \rho_i(h) = -g_1 \widetilde{f}_i g_2 \widehat{f}_j g_3 = g_1 \widetilde{f}_i g_2 \widetilde{f}_j g_3;$$
$$\rho_i \rho_j(h) = -g_1 \widehat{f}_i g_2 \widetilde{f}_j g_3 = g_1 \widetilde{f}_i g_2 \widetilde{f}_j g_3,$$

yielding the desired diamond.

As explained in Remark 7A.10 of Volume 1, (2) can be avoided by first reducing the relation f_i by f_j and then taking its reduction. (We might have to do this several times.)

This leaves (3). Should we be able to resolve all overlap ambiguities, the diamond condition would be satisfied.

Proposition 17B.13 (Bergman). *Any set of relations can be expanded to a set of relations for which any given word h becomes reduction-unique.*

Proof. Otherwise, by Remark 17B.12, we may assume that h contains an overlap ambiguity uvw, i.e., $h = g_1 uvw g_2$. Suppose $\widehat{f}_i = uv$ and $\widehat{f}_j = vw$. Then we also have relations $f_i w$ and $u f_j$; $(\widehat{f_i w - u f_j}) < uvw$ since the leading terms cancel. We adjoin the relation $g_1 \widetilde{f}_i w g_2 - g_1 u \widetilde{f}_j g_2$ to our set of relations. But $\rho_i(h) = -g_1 \widetilde{f}_i w g_2$ and $\rho_j(h) = -g_1 u \widetilde{f}_j g_2$ now have a common reduction, so our overlap ambiguity uvw is resolved. Doing this for all overlap ambiguities for h concludes the proof. □

In fact, there are only finitely many possible overlap ambiguities for h, so any word can be made reduction-unique by adding a finite number of reductions. Unfortunately this does not mean that there is an effective reduction procedure, since the resolution of ambiguities could lead to new ambiguities that are more and more complicated to analyze. More formally, it could be impossible to make the reduction system recursive, as we discussed above. Nevertheless, in practice, Bergman's reduction procedure often is effective. Furthermore, it is a direct generalization of Remark 7A.10 of Volume 1. We could reformulate this discussion in terms of $F\{X\}$ being an algebra filtered over $\mathcal{M}$, and we would get the noncommutative analog of a Gröbner basis; cf. Exercises B1–B3.

Appendix 17C: An introduction to the Burnside Problem

Even when the word problem fails to have a solution, perhaps other problems can be decided. One very basic question, asked over 100 years ago by Burnside [**Bur1**], is

The Burnside Problem (BP). If a f.g. group G has exponent n, then is G finite?

A group is called **periodic** if every element has finite order. Burnside [**Bur2**] asked more generally,

Generalized BP. Is every periodic f.g. group G necessarily finite?

The sister problems for algebras:

The Kurosh Problem. If R is an affine algebra and every element of R is algebraic over F, then is R finite-dimensional?

The Levitzki Problem. If R is an affine algebra without 1 and every element of R is nilpotent, then is R nilpotent?

Although both of these problems have negative answers as originally formulated, each has led to much beautiful mathematics.

An equivalent formulation of Burnside's Problem: Define $B(m, n)$ to be the group of exponent n, freely generated by m elements. (In other words, G satisfies the relations $w^n = 1$ for all words w in the generators.) Then is $B(m, n)$ finite?

When $n = 2$ it is an undergraduate exercise to see that G is Abelian and thus finite. The case $n = 3$ also is easy, cf. Exercise C1, and its solution was known to Burnside. The case $n = 4$ was verified in 1940 by Sanov, who used a clever elementary argument given in Exercise C2. M. Hall proved the case $n = 6$. Full details are given in Hall [**Halm, Ch. 18**]. Incidentally, the case $n = 5$ remains open. Positive information also can be had for the generalized BP.

Remark 17C.1. The generalized BP has an affirmative answer for Abelian groups, by the fundamental theorem of f.g. Abelian groups (Theorem 2.77 of Volume 1).

Proposition 17C.2. *The generalized BP has an affirmative answer for solvable groups.*

Proof. Induction on the solvability degree t of G. If $t = 1$, then G is Abelian.

In general, G/G' is f.g. Abelian and thus finite. Remark 00.1 then implies G' is f.g., of solvability degree $t-1$, and thus is finite, by induction. □

Conversely, if G is finite of prime exponent p, then obviously G is a p-group, which is nilpotent and hence solvable.

Another classical result is Schur's positive solution for linear groups of characteristic 0; cf. Exercise 19A.6. (In fact, Schur proved the stronger result that if G is a periodic subgroup of $\mathrm{GL}(n,\mathbb{C})$, then G has an Abelian subgroup of finite index bounded by a function on n.)

On the negative side, in a remarkably concise paper, Golod [**Go**] found an infinite group generated by two elements, in which the order of every element is some power of p (but not bounded). This effectively disposed of the generalized BP (as well as the Kurosh Problem, although it remains open when R has no zero-divisors). An amazingly simple counterexample, discovered by Grigorchuk [**Gri1**], is given in Exercises C5–C7. BP itself was settled negatively in the 1960's, when Novikov-Adjan and Britton independently produced very difficult counterexamples for large p.

In the 1930s and 1940s, interest increased in the following more restricted version of Burnside's Problem, which was posed formally by Magnus in 1950:

Restricted Burnside Problem: RBP(m,n). Are there only finitely many groups (up to isomorphism) of exponent n generated by m elements?

Remark 17C.3. In the language of $B(m,n)$, the Restricted Burnside Problem asks, "Does $B(m,n)$ have only finitely many finite homomorphic images?" Or, equivalently, "Does $B(m,n)$ have a finite homomorphic image $\bar{B}(m,n)$ such that every finite group of exponent n generated by m elements is a homomorphic image of $\bar{B}(m,n)$?" (Indeed, if $G_1,\dots,G_k$ are the finite homomorphic images of $B(m,n)$, write $G_j \cong B(m,n)/N_j$ for suitable subgroups $N_j \lhd B(m,n)$ of finite index, and take $N = \bigcap_{j=1}^{k} N_j$. Then N has finite index, and $B(m,n)/N$ is the desired finite group $\bar{B}(m,n)$.)

Hall-Higman [**HalpH**] proved the following major reduction.

RBP(m,n) *holds if the following conditions are satisfied, where* $n = p_1^{k_1}\cdots p_\ell^{k_\ell}$:

(1) RBP$(m,p_i^{k_i})$ *holds for each* $1 \le i \le \ell$.
(2) *There are finitely many simple groups* G *of exponent* n.
(3) $\mathrm{Aut}(G)/\mathrm{InnAut}(G)$ *is solvable for all such simple* G.

In other words, given conditions (2) and (3), RBP(m,n) is reduced to the case for n a prime power. This theorem was remarkably prescient in view of the immense strides in the study of simple groups in the subsequent thirty years. In particular, the Feit-Thompson theorem shows that there are no simple groups of odd order, so conditions (2) and (3) are vacuous for n odd. Furthermore, the proposed classification of finite simple groups, which claims that there are only finitely many sporadic simple groups, would yield conditions (2) and (3), thereby reducing the Restricted Burnside Problem to the case where n is a prime power, and thus for G nilpotent. This leads us to the following problem:

Problem 17C.4. For $n = p^k$, is there an upper bound $t(m,n)$ for the nilpotence class of all groups of exponent n generated by m elements?

LEMMA 17C.5. *An affirmative answer to Problem 17C.4 yields a positive solution to the RBP.*

Proof. Let $\gamma_j(G)$ denote the j-subgroup in the lower central series of G, and $\bar{G} = B(m,n)/\gamma_{t(m,n)}(B(m,n))$. Problem 17C.4 implies that every group of exponent n generated by m elements is a homomorphic image of $\bar{G}$. But we claim that $\gamma_j(\bar{G})$ is f.g. for each j. Indeed, this is clear for $j = 0$, so we proceed by induction. $\gamma_{j-1}(\bar{G})/\gamma_j(\bar{G})$ is f.g. Abelian, and thus finite, by the fundamental theorem of Abelian groups (Theorem 2.77 of Volume 1). Hence, $\gamma_j(\bar{G})$ is f.g. by Remark 00.1.

We conclude that $|\bar{G}| = \prod_{j=1}^{t(m,n)} |\bar{G}_{j-1}/\bar{G}_j| < \infty$. □

In 1959, Kostrikin produced a proof of RBP(m,p) for p prime (using Exercise C4 as a springboard), but the proof contained gaps which he later filled in collaboration with Zelmanov. Finally, in 1989, Zelmanov solved Problem 17C.4 affirmatively and thus RBP(m,p^k) for any prime power, thereby concluding the solution of the Restricted Burnside Problem; this major achievement earned him the Fields Medal in 1992. The method of proof of both results was to pass to identities of Lie algebras, so we postpone the remainder of this discussion until Appendix 23B.

One might hope that all groups of prime exponent are solvable, but this is also false, by an example of Bachmuth, Mochizuki, and Walkup. (Their group has exponent 5, but all of its f.g. subgroups are solvable and thus finite, so it is not a counterexample to the Burnside Problem.) This result and other seminal papers (such as Britton's counterexample) appear in the collection Boone-Cannonito-London [**BooCL**].

Chapter 18

Tensor Products

The direct sum of two vector spaces A and B over a field F has the property that the dimension $\dim_F(A\oplus B) = \dim_F A + \dim_F B$, since the union of bases of $A\times\{0\}$ and $\{0\}\times B$ is a base of $A\oplus B$. Analogously, we already have noted in Definition 0.11 of Volume 1 that the vector space $A\otimes B$, called the **tensor product,** has dimension $\dim_F A \dim_F B$.

In this chapter we deal with various aspects of the tensor product $A\otimes B$. When A and B are algebras, just as $A\times B$ turns out to be an algebra, also $A\otimes B$ has a natural structure as an algebra. The tensor product has a huge range of applications, too numerous to describe in this overview.

As with most of linear algebra, one has the usual choice between a concrete, base-oriented definition, which lends itself to immediate computation, or a more abstract, base-free definition. The more abstract definition has clear advantages, since we can use it for important, more general applications. Nevertheless, to build intuition, our point of departure is the concrete definition of the tensor product of vector spaces over a field.

Definition 18.1. Suppose A and B are vector spaces over a field F, having respective bases $\{a_1,\dots,a_m\}$ and $\{b_1,\dots,b_n\}$. Recall from Definition 0.11 of Volume 1 that $A\otimes B$ is defined as the vector space of dimension mn over F with base labelled

$$\{a_i\otimes b_j : 1\le i\le m,\ 1\le j\le n\}. \tag{18.1}$$

For any $a=\sum\alpha_i a_i\in A$ and $b=\sum\beta_j b_j\in B$ with α_i,β_j in F, we define the **simple tensor** $a\otimes b$ to be $\sum_{i,j}\alpha_i\beta_j\, a_i\otimes b_j$. Thus, $\alpha(a\otimes b)=\alpha a\otimes b = a\otimes\alpha b$, $\forall\alpha\in F$.

When A and B are algebras, we introduce a multiplication on $A\otimes B$, starting with the simple tensors:

$$(a\otimes b)(a'\otimes b') = aa'\otimes bb'.$$

By Exercise 13.1, this extends to a product on all of $A\otimes B$. One could check that this multiplication is well-defined and satisfies associativity and distributivity over addition, and that change of base of A and B yields an isomorphic construction of the tensor product. However, this is tedious at best, and it all will become obvious when we obtain a base-free construction.

The role of simple tensors is rather mysterious. Although the simple tensors of $A\otimes B$ clearly include the given base (18.1), as well as many other tensors when $\dim_F A, B > 1$, it is rather easy to concoct tensors that are not simple tensors.

Example 18.2. Suppose $A=B=\mathbb{Q}[\sqrt{2}]$. Then $A\otimes_{\mathbb{Q}} B$ is a commutative $\mathbb{Q}$-algebra, and it is easy to see that no simple tensor is a zero divisor. Nevertheless, $A\otimes B$ is not an integral domain, since if we take

$$r=\sqrt{2}\otimes 1+1\otimes\sqrt{2}\quad\text{and}\quad s=\sqrt{2}\otimes 1-1\otimes\sqrt{2},$$

we have

$$\begin{aligned} rs &= (\sqrt{2}\otimes 1+1\otimes\sqrt{2})(\sqrt{2}\otimes 1-1\otimes\sqrt{2})\\ &= (\sqrt{2}\otimes 1)^2-(1\otimes\sqrt{2})^2\\ &= 2\otimes 1-1\otimes 2 = 2(1\otimes 1)-2(1\otimes 1)=0. \end{aligned}$$

In particular, r and s are not simple tensors.

The basic construction

Despite its intuitive immediacy, Definition 18.1 hampers development of the theory because of its reliance on the choice of base. Fundamental assertions can require intricate calculations, and the specter of well-definedness haunts every proof. Accordingly, we turn to an abstract approach, which provides direct, computation-free proofs of all the important properties. An extra benefit is the ability to work with modules and algebras over arbitrary commutative rings, since the existence of a base no longer is required to define the tensor product. After obtaining many important general properties, we recover Definition 18.1 as a special case, in Theorem 18.21. Before proceeding, let us record some basic properties of tensor products:

$$\begin{aligned} (a_1+a_2)\otimes b &= a_1\otimes b\ +a_2\otimes b;\\ a\otimes(b_1+b_2) &= a\otimes b_1\ +a\otimes b_2;\\ \alpha a\otimes b &= a\otimes\alpha b \end{aligned}$$

for all $a \in A,\ b \in B,\ \alpha \in F$. Thus, the map $(a,b) \mapsto a \otimes b$ resembles a bilinear form, but with values in $A \otimes B$. We use this as our starting point.

Suppose R is any ring, not necessarily commutative. Recall that "module" means "left module." Given a right R-module M and an R-module N, we want to produce an Abelian group $M \otimes_R N$ spanned by "simple tensors" $a \otimes b$ for $a \in M, b \in N$, satisfying the following properties:

$$(a_1 + a_2) \otimes b = a_1 \otimes b \ + a_2 \otimes b; \tag{T1}$$

$$a \otimes (b_1 + b_2) = a \otimes b_1 \ + a \otimes b_2; \tag{T2}$$

$$ar \otimes b = a \otimes rb \tag{T3}$$

for all $a_i \in M,\ b_i \in N,\ r \in R$.

Properties (T1)–(T3) form the foundation of our construction, which is to be "universal" in the following sense:

A **balanced map** is a function $\psi\colon M \times N \to G$ (where $G = (G,+)$ is an Abelian group), satisfying the following properties:

$$\psi(a_1 + a_2, b) = \psi(a_1, b) + \psi(a_2, b); \tag{B1}$$

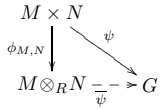
$$\psi(a, b_1 + b_2) = \psi(a, b_1) + \psi(a, b_2); \tag{B2}$$

$$\psi(ar, b) = \psi(a, rb) \tag{B3}$$

for all $a_i \in M,\ b_i \in N,\ r \in R$. The tensor product is to be a group denoted $M \otimes_R N$, endowed with a balanced map $\phi_{M,N}\colon M \times N \to M \otimes_R N$, such that for any balanced map $\psi\colon M \times N \to G$ there is a unique group homomorphism $\overline{\psi}\colon M \otimes_R N \to G$ such that

$$\overline{\psi}\phi_{M,N} = \psi, \tag{18.2}$$

i.e., the diagram



is commutative. Our universal construction is achieved by taking the free Abelian group, i.e., the free $\mathbb{Z}$-module, generated by what are to become the simple tensors, and factoring out all the desired relations.

Definition 18.3. Given a right R-module M and an R-module N over an arbitrary ring R, define $M \otimes_R N$ to be $\mathcal{F}/\mathcal{K}$, where $\mathcal{F}$ is the free $\mathbb{Z}$-module with base $M \times N$, written as the set of ordered pairs $\{(a,b) : a \in M,\ b \in N\}$,

and where $\mathcal{K}$ is the submodule generated by the elements

$$\begin{gathered}(a_1 + a_2, b) - (a_1, b) - (a_2, b);\\ (a, b_1 + b_2) - (a, b_1) - (a, b_2);\\ (ar, b) - (a, rb)\end{gathered}$$

for all $a_i \in M,\ b_i \in N,\ r \in R$. When R is understood, we write $M \otimes N$ instead of $M \otimes_R N$.

Our definition is tailored to satisfy (T1), (T2), and (T3) in the Abelian group $M \otimes N$, where we write $a \otimes b$ for the coset $(a,b) + \mathcal{K}$. Let us verify the desired universal property.

PROPOSITION 18.4. *Given a right R-module M and an R-module N, an Abelian group G, and a balanced map $\psi\colon M \times N \to G$, we get a unique group homomorphism $\overline{\psi}\colon M \otimes N \to G$ satisfying (18.2) given by $\overline{\psi}(a \otimes b) = \psi(a,b)$.*

Proof. We define a map $\hat{\psi}\colon \mathcal{F} \to G$ by its action on the base of $\mathcal{F}$, namely $\hat{\psi}(a,b) = \psi(a,b)$. By definition $\hat{\psi}(\mathcal{K}) = 0$, so Noether's first isomorphism theorem gives us the unique desired map $\overline{\psi} : \mathcal{F}/\mathcal{K} \to G$. □

However abstract the construction of $M \otimes_R N$, it must be "correct," being unique up to isomorphism, by "abstract nonsense" (cf. Proposition 8.6 of Volume 1).

We have an immediate dividend: Whereas it often is difficult to check computationally that a given map defined on $M \otimes N$ is well-defined, balanced maps from $M \times N$ are automatically well-defined since they are defined set-theoretically. Thus, our strategy is to intuit some desired map $M \otimes N \to G$; then, working backwards, we define the corresponding map $M \times N \to G$, check that it is balanced, and apply Proposition 18.4 to produce our desired map. One can expedite this procedure by means of the following key observation.

PROPOSITION 18.5. *Suppose $f\colon M \to M'$ is a map of right R-modules and $g\colon N \to N'$ is a map of R-modules. Then there is a group homomorphism denoted*

$$f \otimes g\colon M \otimes N \to M' \otimes N'$$

given by $(f \otimes g)(a \otimes b) = f(a) \otimes g(b)$.

Proof. Define $\psi\colon M\times N \to M'\otimes N'$ by $\psi(a,b) = f(a)\otimes g(b)$. ψ is balanced since

$$\begin{aligned}\psi(a_1+a_2,b) &= f(a_1+a_2)\otimes g(b)\\ &= f(a_1)\otimes g(b) + f(a_2)\otimes g(b) = \psi(a_1,b)+\psi(a_2,b);\\ \psi(a,b_1+b_2) &= \psi(a,b_1)+\psi(a,b_2) \quad \text{(analogously);}\\ \psi(ar,b) &= f(ar)\otimes g(b) = f(a)r\otimes g(b) = f(a)\otimes rg(b) = \psi(a,rb).\end{aligned}$$

Thus, $\overline{\psi}$ is the desired homomorphism $f\otimes g$. □

Since the simple tensors span $M\otimes N$, we can verify properties of $f\otimes g$ by checking how they act on the simple tensors, as illustrated in the next observation.

Remark 18.6. If, in Proposition 18.5, f and g are isomorphisms, then $f\otimes g$ is also an isomorphism, whose inverse is given by $f^{-1}\otimes g^{-1}$. (Indeed,

$$\begin{aligned}(f^{-1}\otimes g^{-1})(f\otimes g)(a\otimes b) &= (f^{-1}\otimes g^{-1})(f(a)\otimes g(b))\\ &= f^{-1}f(a)\otimes g^{-1}g(b) = a\otimes b\end{aligned}$$

for $a\in M$ and $b\in N$, and likewise

$$(f\otimes g)(f^{-1}\otimes g^{-1})(a'\otimes b') = ff^{-1}(a')\otimes gg^{-1}(b') = a'\otimes b'$$

for $a'\in M'$ and $b'\in N'$. Thus, $(f^{-1}\otimes g^{-1})(f\otimes g)$ and $(f\otimes g)(f^{-1}\otimes g^{-1})$ fix all simple tensors, and so are the identity maps.)

In other words, if $M\cong M'$ and $N\cong N'$, then $M\otimes_R N\cong M'\otimes_R N'$.

Tensor product of bimodules.

So far we have defined only an additive group structure for $M\otimes N$. We can obtain more structure when we start with extra structure on M and N.

PROPOSITION 18.7. *Suppose R and T are arbitrary rings, M is a T,R-bimodule, and N is an R-module. Then $M\otimes_R N$ is a T-module under the multiplication*

$$t(a\otimes b) = ta\otimes b \tag{18.3}$$

for t in T, a in M, and b in N.

Proof. Given t in T, we have the left multiplication map $\ell_t\colon M\to M$ given by $\ell_t(a) = ta$, clearly a right R-module map. Then

$$\ell_t\otimes 1_N\colon M\otimes_R N\to M\otimes_R N$$

satisfies (18.3). □

Remark 18.8. When T is a ring containing R and $M=T$, the T-module $T\otimes_R N$ is called the T-module **extended from** N. This situation arises more generally whenever there is a ring homomorphism $\psi\colon R\to T$; we define T as a T,R-bimodule via Example 13.35″.

Since any right R-module is a $\mathbb{Z},R$-bimodule, this remark generalizes Proposition 18.5. On the other hand, the most common application is when $R=T$ is a commutative ring, since then any right R-module is an R,R-bimodule in the obvious way ($r_1ar_2 = ar_1r_2$ for $r_i\in R,\ a\in M$).

Example 18.9. In the notation of Proposition 18.7, any R-module map $f\colon N\to P$ yields a T-module map $1_M\otimes f\colon M\otimes_R N\to M\otimes_R P$. From the categorical point of view, $M\otimes_R \_$ yields a functor from R-**Mod** to T-**Mod**.

Remark 18.9′. Proposition 18.7 also has a right-handed version: If N is an R,T-bimodule, then $M\otimes_R N$ is a right T-module via $(a\otimes b)t = a\otimes bt$. Then the right-handed version of Example 18.9 is available.

Remark 18.9″. If M is a right R-module and N is an R-module, we can also view $M\otimes_R N$ as an R-module via the **trivial** action for $a\in M$, $b\in N$, and $r\in R$:

$$r(a\otimes b) = a\otimes rb = (1_M\otimes \ell_r)(a\otimes b),$$

where ℓ_r denotes left multiplication by r.

Isomorphisms of tensor products.

Our next objective is to compare various constructions involving tensor products. We start with a fundamental observation in algebras, rings, and modules: Multiplication is a balanced map (because of distributivity and associativity).

PROPOSITION 18.10. *$R\otimes_R N\cong N$ for any R-module N.*

Proof. The balanced map $\psi\colon R\times N\to N$ given by $\psi(r,a) = ra$ yields a map $\overline{\psi}\colon R\otimes_R N\to N$. This has an inverse $\phi\colon N\to R\otimes_R N$ given by $a\mapsto 1\otimes a$. Indeed, for $r\in R$ and $a\in N$,

$$\overline{\psi}(\phi(a)) = \overline{\psi}(1\otimes a) = 1a = a;$$

$$\phi(\overline{\psi}(r\otimes a)) = \phi(ra) = 1\otimes ra = r\otimes a.$$ □

PROPOSITION 18.11. *For any right R-modules M_i and R-module N, there is an isomorphism*

$$\phi\colon (M_1\oplus\ldots\oplus M_t)\otimes N\cong (M_1\otimes N)\oplus\cdots\oplus(M_t\otimes N)$$

given by

$$(a_1, \dots, a_t) \otimes b \mapsto (a_1 \otimes b,\ \dots,\ a_t \otimes b).$$

Proof. Let π_i denote the natural projection $M_1 \oplus \cdots \oplus M_t \to M_i$. Then, by Proposition 18.5, we have projections

$$\pi_i \otimes 1_N \colon (M_1 \oplus \cdots \oplus M_t) \otimes N \to M_i \otimes N,$$

so our desired map ϕ is $(\pi_1 \otimes 1_N, \dots, \pi_t \otimes 1_N)$. The inverse map

$$(M_1 \otimes N) \oplus \cdots \oplus (M_t \otimes N) \to (M_1 \oplus \cdots \oplus M_t) \otimes N$$

is obtained by taking the canonical injections $\nu_j \colon M_j \to M_1 \oplus \cdots \oplus M_t$; then we form $\nu_j \otimes 1_N \colon M_j \otimes N \to (M_1 \oplus \cdots \oplus M_t) \otimes N$, $1 \le i \le t$, and in view of Remark 2.14 of Volume 1, $\bigoplus_{i=1}^{t} (\nu_i \otimes 1_N)$ is ϕ^{-1}, as desired. □

Symmetrically, we have

PROPOSITION 18.11′. $M \otimes (N_1 \oplus \cdots \oplus N_t) \cong (M \otimes N_1) \oplus \cdots \oplus (M \otimes N_t)$.

The same argument yields distributivity of tensor product over arbitrary direct sums; cf. Exercise 2. We can make Proposition 18.11 more explicit in the case of free modules.

PROPOSITION 18.12. *Suppose M is a free right R-module with base $B = \{b_i : i \in I\}$, and N is an R-module. Then every element of $M \otimes N$ can be written uniquely in the form $\sum_{i \in I} b_i \otimes v_i$ for v_i in N. In particular, if $\sum_{i \in I} b_i \otimes v_i = 0$, then each $v_i = 0$.*

Proof. Any element of $M \otimes N$ can be written as

$$\sum_k \sum_i b_i r_{ik} \otimes v_k = \sum_i b_i \otimes \left(\sum_k r_{ik} v_k \right)$$

for $v_k \in N$. To prove uniqueness, it suffices to verify the last assertion. First note that the projection $\pi_j \colon M \to R$ onto the j-th coordinate given by $\sum_i b_i r_i \mapsto r_j$ yields a projection

$$\pi_j \otimes 1_N \colon M \otimes N \to R \otimes N \cong N$$

given by $\sum_k (\sum_i b_i r_{ik}) \otimes v_k \mapsto \sum_k r_{jk} v_k$ (where each $v_k \in N$).

If $\sum_i b_i \otimes v_i = 0$, then, for each j,

$$0 = (\pi_j \otimes 1_N) \left(\sum_i b_i \otimes v_i \right) = v_j. \qquad \square$$

COROLLARY 18.13. *If C is commutative and M and N are free C-modules with respective bases $\{b_i : i \in I\}$ and $\{b'_j : j \in J\}$, then $M \otimes_C N$ is also a free C-module, with base $\{b_i \otimes b'_j : (i, j) \in I \times J\}$. In particular,*

$$C^{(m)} \otimes_C C^{(n)} \cong C^{(mn)}.$$

Thus, when C is a field, $\dim(M \otimes N) = \dim M \dim N$.

Proof. We need only prove the first assertion. Since $\{b_i \otimes b'_j : i \in I,\ j \in J\}$ clearly span $A \otimes B$, it remains to prove that they are independent. If $0 = \sum_{i,j} c_{ij} b_i \otimes b'_j = \sum_i b_i \otimes (\sum_j c_{ij} b'_j)$, then $\sum_j c_{ij} b'_j = 0$ for each i, by Proposition 18.12, implying that each $c_{ij} = 0$. □

Let us sneak in a quick example.

Example 18.14. If V is a vector space over a field F and V^* is the dual space, then

$$V \otimes_F V^* \cong \operatorname{End}_F V \cong M_n(F).$$

(Indeed, given $w \in V$ and $f \in V^*$ we define $\psi_{w,f} \in \operatorname{End}_F V$ by

$$\psi_{w,f}(v) = f(v) w.$$

Then $(w, f) \mapsto \psi_{w,f}$ defines a balanced map $\psi \colon V \times V^* \to \operatorname{End}_F V$, which induces a map $\overline{\psi} : V \otimes V^* \to \operatorname{End}_F V$. Taking a base $e_1, \dots, e_n$ of V and dual base $e_1^*, \dots, e_n^*$ of V^*, we see that the base of $V \otimes V^*$ is comprised of the $e_i \otimes e_j^*$, which $\overline{\psi}$ sends to the map

$$\sum_{k=1}^{n} \alpha_k e_k \mapsto \alpha_j e_i,$$

which in turn is identified with left multiplication by the matrix unit e_{ij}. Since the e_{ij} are a base of $M_n(F)$, $\overline{\psi}$ sends base to base and thus is an isomorphism.)

This gives us more insight on simple tensors, for any simple tensor must have rank 1 as a matrix. Since the matrices of rank $\ne 1$ are dense in the Zariski topology of $M_n(F)$, "most" elements of $V \otimes_F V^*$ are not simple tensors. See Exercise 5 to describe the trace in terms of this correspondence.

PROPOSITION 18.15. *Suppose R_1, R_2, R_3, and R_4 are rings. For any R_i, R_{i+1}-bimodules M_i, $1 \le i \le 3$, we have*

$$M_1 \otimes_{R_2} (M_2 \otimes_{R_3} M_3) \cong (M_1 \otimes_{R_2} M_2) \otimes_{R_3} M_3$$

as R_1, R_4-bimodules, via the correspondence

$$a \otimes (b \otimes c) \mapsto (a \otimes b) \otimes c.$$

Proof. Given any a in M_1, one can define a right R_3-module map

$$f_a\colon M_2 \to M_1 \otimes M_2$$

by $f_a(b) = a \otimes b$. By Proposition 18.7 and Remark 18.9′, $f_a \otimes 1_{M_3}$ defines a right R_4-module map $M_2 \otimes M_3 \to (M_1 \otimes M_2) \otimes M_3$, and we have a balanced map

$$M_1 \times (M_2 \otimes M_3) \to (M_1 \otimes M_2) \otimes M_3$$

given by

$$(a,\ b \otimes c) \mapsto f_a(b) \otimes c = (a \otimes b) \otimes c.$$

Now Proposition 18.4 provides a map

$$M_1 \otimes (M_2 \otimes M_3) \to (M_1 \otimes M_2) \otimes M_3$$

given by

$$a \otimes (b \otimes c) \mapsto (a \otimes b) \otimes c,$$

clearly an R_1, R_4-bimodule map whose inverse map is defined symmetrically. □

Because of this natural isomorphism, we write $M_1 \otimes M_2 \otimes M_3$ without parentheses.

Proposition 18.16. *For any modules A, B over a commutative ring C, there is a* **twist isomorphism** $\tau\colon A \otimes_C B \to B \otimes_C A$ *given by* $\tau(a \otimes b) = b \otimes a$.

Proof. Via the balanced map given by $(a, b) \mapsto b \otimes a$. □

Corollary 18.17. $(A \otimes B) \otimes (A' \otimes B') \cong (A \otimes A') \otimes (B \otimes B')$ *for any C-modules A, A', B, B', via*

$$(a \otimes b) \otimes (a' \otimes b') \mapsto (a \otimes a') \otimes (b \otimes b').$$

Proof. Follow the isomorphisms of Propositions 18.15 and 18.16:

$$\begin{aligned}(A \otimes B) \otimes (A' \otimes B') &\cong A \otimes (B \otimes (A' \otimes B')) \cong A \otimes ((B \otimes A') \otimes B') \\ &\cong A \otimes (A' \otimes B) \otimes B' \\ &\cong ((A \otimes A') \otimes B) \otimes B' \\ &\cong (A \otimes A') \otimes (B \otimes B').\end{aligned}$$ □

Here is an important special case.

Corollary 18.17′. $(A \otimes B) \otimes (A \otimes B) \cong (A \otimes A) \otimes (B \otimes B)$, *via*

$$(a_1 \otimes b_1) \otimes (a_2 \otimes b_2) \mapsto (a_1 \otimes a_2) \otimes (b_1 \otimes b_2).$$

Suppose K is a right R-submodule of M, and N is an R-module. It is customary by abuse of notation to write $K \otimes N$ for the submodule of $M \otimes N$ spanned by simple tensors of the form $a \otimes b$, where $a \in K$ and $b \in N$. However, this is *not* the same as the construction $K \otimes N$, as we see in the next example. (Nevertheless, we shall abuse the notation and explain the context if there is ambiguity.)

Example 18.18. Let $N = \mathbb{Z}/2$. We have the isomorphism $\mathbb{Z} \otimes_{\mathbb{Z}} N \cong N$ as $\mathbb{Z}$-modules and thus as Abelian groups. Since $\mathbb{Z} \cong 2\mathbb{Z}$, we have $2\mathbb{Z} \otimes_{\mathbb{Z}} N \cong N$, by Remark 18.6. However, viewing $2\mathbb{Z} \otimes_{\mathbb{Z}} N$ as a submodule of $\mathbb{Z} \otimes_{\mathbb{Z}} N \cong N$, we see that

$$2m \otimes [n] = m \otimes [2n] = m \otimes [0] = 0;$$

i.e., $2\mathbb{Z} \otimes N = 0$.

Put more formally, the fact that $f\colon K \to M$ is monic does not necessarily imply that the map $f \otimes 1\colon K \otimes N \to M \otimes N$ is monic. Compare this with Exercise 3.

Algebras as tensor products.

Tensor products of algebras lead to the very essence of multiplication in algebra.

Remark 18.19. Suppose R is a C-algebra, and let $\mu\colon R \times R \to R$ be the multiplication map

$$\mu(a, b) = ab.$$

Clearly μ is a balanced map, and thus, we have a C-module map

$$\bar{\mu}\colon R \otimes_C R \to R$$

given by $a \otimes b \mapsto ab$. For later use, note that associativity of multiplication over R was not needed to obtain $\bar{\mu}$.

Conversely, given a C-module R and a map $\bar{\mu}\colon R \otimes_C R \to R$, we can define multiplication on R via $r_1 r_2 = \bar{\mu}(r_1 \otimes r_2)$. It is easy to verify that $(R, +, \bar{\mu})$ satisfies distributivity of multiplication over addition, as well as all of the other algebra axioms *except* possibly the following two axioms:

(A1) associativity of multiplication:

$$\bar{\mu}(\bar{\mu}(r_1 \otimes r_2) \otimes r_3) = \bar{\mu}(r_1 \otimes \bar{\mu}(r_2 \otimes r_3)).$$

(A2) Existence of the unit element 1 of R, satisfying

$$\bar{\mu}(1\otimes r) = r = \bar{\mu}(r\otimes 1),\ \forall r \in R.$$

These properties can be written respectively in terms of commutative diagrams. Associativity is given by commutativity of the diagram:

$$\begin{array}{ccc} R\otimes R\otimes R & \xrightarrow{1\otimes\bar{\mu}} & R\otimes R \\ {\scriptstyle \bar{\mu}\otimes 1}\downarrow & & \downarrow{\scriptstyle \bar{\mu}} \\ R\otimes R & \xrightarrow{\bar{\mu}} & R \end{array} \tag{18.4}$$

The unit element 1 is described in terms of the map $\iota\colon C \to R$ given by $c \mapsto c\cdot 1$ and commutativity of the diagram

$$\begin{array}{ccccc} C\otimes_C R & \xrightarrow{\iota\otimes 1_R} & R\otimes_C R & \xleftarrow{1_R\otimes\iota} & R\otimes_C C \\ & {\scriptstyle\cong}\searrow & \downarrow{\scriptstyle\bar{\mu}} & \swarrow{\scriptstyle\cong} & \\ & & R & & \end{array} \tag{18.5}$$

This approach is dualized in Appendix 19B (algebraic groups) and more generally in Chapter 26 (Hopf Algebras).

Tensor products of algebras

Our next goal is to see how the algebra structure passes to tensor products. One could do this by a clever manipulation of balanced maps, but here is a more conceptual method.

Remark 18.20. For any C-algebra R with $\bar{\mu}\colon R\otimes_C R \to R$ as in Remark 18.19, together with C-module maps $f\colon A \to R$ and $g\colon B \to R$, the composite

$$A\otimes_C B \xrightarrow{f\otimes g} R\otimes_C R \xrightarrow{\bar{\mu}} R$$

defines a map $A\otimes_C B \to R$ given by $a\otimes b \mapsto f(a)g(b)$ for $a \in A$, $b \in B$.

THEOREM 18.21. *If A and B are algebras over a commutative ring C, then $A\otimes_C B$ is also a C-algebra whose multiplication is given by*

$$\begin{aligned}(a\otimes b)(a'\otimes b') &= aa'\otimes bb';\\ c(a\otimes b) &= ca\otimes b.\end{aligned}$$

Proof. Let $\bar{\mu}_A\colon A\otimes_C A \to A$ and $\bar{\mu}_B\colon B\otimes_C B \to B$ denote the customary multiplications in A and B respectively. Now consider the following composite (using Corollary 18.17′):

$$(A\otimes B)\otimes(A\otimes B) \cong (A\otimes A)\otimes(B\otimes B) \xrightarrow{\bar{\mu}_A\otimes\bar{\mu}_B} A\otimes B.$$

This sends $(a_1\otimes b_1)\otimes(a_2\otimes b_2) \mapsto a_1a_2\otimes b_1b_2$ and thus defines multiplication in $A\otimes B$.

Associativity in $A\otimes B$ follows from associativity in A and B:

$$\begin{aligned}((a_1\otimes b_1)(a_2\otimes b_2))(a_3\otimes b_3) &= (a_1a_2)a_3\otimes(b_1b_2)b_3\\ &= a_1(a_2a_3)\otimes b_1(b_2b_3)\\ &= (a_1\otimes b_1)((a_2\otimes b_2)(a_3\otimes b_3)).\end{aligned}$$

Likewise, if 1_A and 1_B are the respective unit elements of A and B, then $1_A\otimes 1_B$ is clearly the unit element of $A\otimes B$. □

Remark 18.22. When studying nonassociative algebras in Chapter 21, we need the observation that associativity of A and B were needed only in the verification that $A\otimes B$ is associative. Hence the tensor product of nonassociative algebras is a nonassociative algebra.

At this stage, we return to check that the earlier tensor isomorphisms are actually isomorphisms of algebras. The following observation helps us verify that maps from tensor products are algebra homomorphisms:

Remark 18.23. Suppose algebra homomorphisms $f\colon A \to R$ and $g\colon B \to R$ satisfy $f(a)g(b) = g(b)f(a)$ for all $a \in A, b \in B$. Then the map

$$\varphi\colon A\otimes_C B \to R$$

given by $a\otimes b \mapsto f(a)g(b)$ (cf. Remark 18.20) is an algebra homomorphism. (Indeed,

$$\begin{aligned}\varphi((a_1\otimes b_1)(a_2\otimes b_2)) = \varphi(a_1a_2\otimes b_1b_2) &= f(a_1)f(a_2)g(b_1)g(b_2)\\ &= f(a_1)g(b_1)f(a_2)g(b_2)\\ &= \varphi(a_1\otimes b_1)\varphi(a_2\otimes b_2).)\end{aligned}$$

Example 18.24. $R[\lambda] \cong R \otimes_C C[\lambda]$ for any C-algebra R. (Indeed, by Remark 18.23, the natural injections $R \to R[\lambda]$ and $C[\lambda] \to R[\lambda]$ give a homomorphism $\varphi\colon R \otimes_C C[\lambda] \to R[\lambda]$, which, viewing both sides as (free) R-modules, sends the base $\{1 \otimes \lambda^i : i \in \mathbb{N}\}$ to the base $\{\lambda^i : i \in \mathbb{N}\}$; hence φ is an isomorphism.)

THEOREM 18.25. *The following algebra isomorphisms hold for any algebras A_1, A_2, and A_3 over a commutative ring C:*

$$A_1 \otimes_C C \cong C \otimes_C A_1 \cong A_1;$$
$$A_1 \otimes A_2 \cong A_2 \otimes A_1;$$
$$A_1 \otimes (A_2 \otimes A_3) \cong (A_1 \otimes A_2) \otimes A_3.$$

Proof. The maps in Propositions 18.10, 18.16, and 18.15 preserve multiplication. □

Example 18.26 (Central localization as a tensor product). Recall the definition of $S^{-1}R$ from Definition 8.1 of Volume 1, where S is a submonoid of $C \subseteq \mathrm{Cent}(R)$. By the universal property of $S^{-1}C$, stated in Proposition 8.4 of Volume 1, the natural map $C \to S^{-1}R$ given by $c \mapsto \frac{c}{1}$ induces a homomorphism $S^{-1}C \to S^{-1}R$ given by $\frac{c}{s} \mapsto \frac{c}{s}$, and thus (via the balanced map argument), a homomorphism $\varphi\colon R \otimes_C S^{-1}C \to S^{-1}R$ given by $r \otimes \frac{c}{s} \mapsto \frac{rc}{s}$. In the opposite direction, applying the universal property of localization to the natural homomorphism $R \to R \otimes_C S^{-1}C$ yields a map $S^{-1}R \to R \otimes_C S^{-1}C$ given by $\frac{r}{s} \mapsto r \otimes \frac{1}{s}$, inverse to φ. Thus, φ is an isomorphism. This identification is useful since it enables us to transfer general properties of tensor products to central localization.

Tensor products are also used for changing the base ring of an algebra.

Remark 18.27 (Change of scalars). Suppose R is an algebra over a commutative ring C, and H is a commutative C-algebra. Then $H \otimes_C R$ is an H-algebra, under the operation

$$h(h_1 \otimes r) = (hh_1) \otimes r;$$

cf. Equation (18.3). Furthermore, if R is free over C, then $H \otimes_C R$ is free over H, of the same rank, by Proposition 18.12.

Often we write the tensor product from the other side: $R \otimes_C H \cong H \otimes_C R$, by Theorem 18.25.

One can also change scalars in the other direction by means of structure constants; cf. Exercise 8. We also want to study the structure of the tensor product with respect to its components.

Remark 18.28. If $A_1 \lhd R_1$, then $A_1 \otimes R_2$ is an ideal of $R_1 \otimes R_2$, and there is an isomorphism $\psi\colon (R_1 \otimes R_2)\Big/(A_1 \otimes R_2) \xrightarrow{\sim} (R_1/A_1) \otimes R_2$. (Indeed, the natural map $R_1 \otimes R_2 \to (R_1/A) \otimes R_2$ has kernel containing $A_1 \otimes R_2$, yielding the homomorphism ψ; the inverse map is induced from the balanced map $R_1/A_1 \times R_2 \to (R_1 \otimes R_2)\Big/(A_1 \otimes R_2)$ given by

$$(r_1 + A_1,\, r_2) \mapsto r_1 \otimes r_2 + A_1 \otimes R_2.)$$

Likewise for the second component. See Exercise 7 for more precise information.

Applications of tensor products

The rest of this chapter describes different ways in which the tensor product is used.

Tensor products over a field.

Let us consider the tensor product $R_1 \otimes_F R_2$ of algebras over a field F. In this case, the canonical maps $R_1 \to R_1 \otimes R_2$ and $R_2 \to R_1 \otimes R_2$ (given respectively by $a \mapsto a \otimes 1$ and $b \mapsto 1 \otimes b$) are monic, in view of Corollary 18.13 (since we may choose respective bases of R_1 and R_2 containing 1).

We saw in Example 18.2 that the tensor product of fields could have 0-divisors, and now we describe the structure explicitly.

Example 18.29. (i) Suppose $K = F[a]$ where $a \notin F$ is separable over F, and L/F is any field extension. We factor the minimal polynomial of a over F as $f = g_1 \cdots g_m$, a product of irreducibles in $L[\lambda]$. Since a is separable, the g_i are distinct and thus pairwise relatively prime. But $K \cong F[\lambda]/F[\lambda]f$, implying, by Remark 18.28 and the Chinese Remainder Theorem:

$$L \otimes K \cong L \otimes (F[\lambda]/F[\lambda]f) \cong L[\lambda]/L[\lambda]f \cong \prod_{i=1}^{m} L[\lambda]/L[\lambda]f_i,$$

a direct product of fields. When $L \supseteq K$, we note that $m > 1$ since $\lambda - a$ divides f in $K[\lambda]$, and the corresponding direct sum component is $L[\lambda]/L[\lambda](\lambda - a) \cong L$.

(ii) If, in (i), f splits into n linear factors over L, then $L[\lambda]/L[\lambda]f_i \cong L$, for each $1 \le i \le n$, implying that $L \otimes_F K \approx L^{(n)}$. In particular, if $[K\colon F] = 2$ with $\mathrm{char}(F) \neq 2$, then $K \otimes_F K \cong K \times K$.

(iii) If fields $K_1, K_2 \supset F$ have a common subfield K algebraic over F, then $K_1 \otimes_F K_2$ contains $K \otimes_F K$ which, by (i), has zero divisors.

(iv) On the other hand, notation as in (iii), for any maximal ideal P of $K_1 \otimes_F K_2$, the algebra $(K_1 \otimes K_2)/P$ is a field in which we can embed both K_1 (via the map $a \mapsto (a \otimes 1) + P$) and K_2 (via $a \mapsto (1 \otimes a) + P$).

We can elaborate this argument to get a criterion for separability.

PROPOSITION 18.29′. *A finite field extension $K \supseteq F$ is separable iff the ring $K \otimes_F K$ is semisimple.*

Proof. ($\Rightarrow$) Using Corollary 4.80 of Volume 1, we may write $K = F[a]$, implying by Example 18.29(i) that $K \otimes_F K$ is a direct product of fields.

($\Leftarrow$) First assume that K is a purely inseparable extension of F, and take $a \in K \setminus F$ with $a^p \in F$, where $p = \mathrm{char}(F)$. Then

$$(a \otimes 1 - 1 \otimes a)^p = a^p \otimes 1 - 1 \otimes a^p = a^p(1 \otimes 1 - 1 \otimes 1) = 0,$$

proving that$K \otimes_F K$ has the nilpotent ideal generated by $(a \otimes 1 - 1 \otimes a)$ and so cannot be semiprime.

In general, K is a purely inseparable extension of the separable closure L of F in K, and the non-semisimple ring $K \otimes_L K$ is a homomorphic image of $K \otimes_F K$, proving that $K \otimes_F K$ is not semisimple. □

Thus, tensor products could be used to make certain aspects of field theory more precise; for example, the compositum of fields is constructed in Exercise 13, using tensor products. One basic application of Remark 18.27 is to "improve" the structure of an algebra.

Definition 18.30. Suppose R is a f.d. F-algebra. A field extension K/F is a **splitting field** of R if $R \otimes_F K$ is split semisimple (as a K-algebra); cf. Definition 15.24.

Warning. Let $J = \mathrm{Jac}(R)$, which is nilpotent since R is a f.d. algebra. If $R \otimes_F K$ is split semisimple, then $J \otimes_F K$ is a nilpotent ideal of $R \otimes_F K$, and thus 0, implying that R is semisimple. On the other hand, a f.d. semisimple algebra R need not have a splitting field. For example, if R is a purely inseparable field extension of F, then $R \otimes_F K$ cannot be split semisimple, in view of Exercise 15. This is a delicate issue to be pursued further in Appendix 25B.

Sometimes algebraically closed fields are difficult to study arithmetically, so we might prefer to use smaller fields. The following proposition provides a way of cutting down splitting fields.

PROPOSITION 18.31. *If an F-algebra R has a splitting field K, then K contains some subfield K_0 f.g. over F such that K_0 also splits R. In particular, if K is algebraic over F, then $[K_0 : F] < \infty$.*

Proof. By considering structure constants. Writing $K \otimes R$ as $\prod_{k=1}^t M_{n_k}(K)$, we let $\{e_{ijk} : 1 \leq i, j \leq n_k\}$ denote the $n_k \times n_k$ matrix units of $M_{n_k}(K)$;

$$e_{ijk} = \sum_u a_{ijku} \otimes r_{ijku}$$

for suitable $a_{ijku} \in K$ and $r_{ijku} \in R$. Then, letting K_0 be the F-subfield of K generated by all the a_{ijku}, we see that each $e_{ijk} \in K_0 \otimes_F R$, implying that $K_0 \otimes_F R$ is split (since it contains the requisite sets of matrix units.)

□

We get more information about tensor products via a technical observation.

Remark 18.32. Suppose A, B are F-algebras over a field F, and write any element $y = \sum_{i=1}^t a_i \otimes b_i$ of $A \otimes B$ in such a way that t is minimal. Then $a_1, \ldots, a_t$ are linearly F-independent, as are $b_1, \ldots, b_t$. (Indeed if say $a_t = \sum_{i=1}^{t-1} \alpha_i a_i$, then we could replace $\sum_{i=1}^t a_i \otimes b_i$ by $\sum_{i=1}^{t-1} a_i \otimes (b_i + \alpha_i b_t)$, thereby lowering t, contrary to our choice of t.)

PROPOSITION 18.33. *Suppose R is simple with center a field F, and W is an F-algebra. Any nonzero ideal I of the tensor product $R \otimes_F W$ contains $\{1 \otimes w : w \in V\}$, for some nonzero ideal V of W.*

Proof. Take $0 \neq y = \sum_{i=1}^t r_i \otimes w_i \in I$ with t minimal.

We first show that $t = 1$. Indeed, assume on the contrary that $t \geq 2$. We claim that there are $a_i, b_i \in R$ such that $\sum a_j r_1 b_j = 0$ but $\sum a_j r_2 b_j \neq 0$. If not, then $\sum a_j r_1 b_j = 0$ always implies $\sum a_j r_2 b_j = 0$, so we have a well-defined map $f\colon R \to R$ given by $\sum a_j r_1 b_j \mapsto \sum a_j r_2 b_j$ (where the $a_j, b_j \in R$ are arbitrary) which is defined on all of R since $Rr_1R = R$. Clearly f is an R, R-bimodule map, so writing $z = f(1)$, we see

$$zr = f(1r) = f(r1) = rf(1) = rz$$

for all $r \in R$, implying that $z \in F$, and $r_2 = f(r_1) = f(1r_1) = f(1)r_1 = zr_1$, contrary to Remark 18.32, which says that the r_i are F-independent.

But now taking a_j, b_j as in the claim, we have

$$\sum_{i=1}^t \Big(\sum_j a_j r_i b_j \Big) \otimes w_i = \sum_i (a_j \otimes 1) y (b_j \otimes 1) \in I.$$

Letting $\tilde{r}_i = \sum_j a_j r_i b_j$, we have by hypothesis $\tilde{r}_1 = 0$ and $\tilde{r}_2 \neq 0$, so we have lowered t, contrary to hypothesis.

We have proved that $t = 1$; i.e., I contains a simple tensor $r \otimes w \neq 0$. But $RrR = R$ and

$$RrR \otimes w = (R \otimes 1)(r \otimes w)(R \otimes 1) \subseteq I,$$

implying that $1 \otimes w \in I$; hence, $1 \otimes RwR = (1 \otimes R)w(1 \otimes R) \subseteq I$. □

COROLLARY 18.34. *If, in Proposition 18.33, W is simple, then $R \otimes_F W$ is also simple; if W is semisimple, then $R \otimes_F W$ is semisimple.*

We have already seen in Example 18.29 that the conclusion may fail when $F \subset \mathrm{Cent}(R)$. These results are critical in the study of central simple algebras (Chapter 24). The idea of proof of Proposition 18.33 is quite useful, and already has been employed in Exercise 16.25. See Exercises 23 and 24 for structural results concerning tensor products and the Jacobson radical.

Tensor products of matrix algebras.

We turn to the tensor product of endomorphism rings.

PROPOSITION 18.35. *If M_i and N_i are modules over the algebra R_i, for $i = 1, 2$, then there is a canonical homomorphism*

$$\mathrm{Hom}_{R_1}(M_1, N_1) \otimes_C \mathrm{Hom}_{R_2}(M_2, N_2) \to \mathrm{Hom}_{R_1 \otimes_C R_2}(M_1 \otimes M_2,\ N_1 \otimes N_2).$$

Proof. We first define the balanced map

$$\Phi\colon \mathrm{Hom}_{R_1}(M_1, N_1) \times \mathrm{Hom}_{R_2}(M_2, N_2) \to \mathrm{Hom}_{R_1 \otimes_C R_2}(M_1 \otimes M_2, N_1 \otimes N_2)$$

by $(f, g) \mapsto f \otimes g$; thus Φ yields a C-module homomorphism

$$\bar{\Phi}\colon \mathrm{Hom}_{R_1}(M_1, N_1) \otimes_C \mathrm{Hom}_{R_2}(M_2, N_2) \to \mathrm{Hom}_{R_1 \otimes_C R_2}(M_1 \otimes M_2, N_1 \otimes N_2),$$

given by $f \otimes g \mapsto f \otimes g$. □

Unfortunately, the notation of the proof is confusing; the simple tensor $f \otimes g$ was sent to the tensor product of maps, again denoted $f \otimes g$.

Remark 18.35′. As a special case of Proposition 18.35, if M and N are modules over C, then there is a canonical homomorphism

$$\mathrm{End}_C M \otimes \mathrm{End}_C N \to \mathrm{End}_C(M \otimes N), \tag{18.6}$$

which is seen to be a ring homomorphism.

COROLLARY 18.36. *$M_m(C) \otimes M_n(C) \cong M_{mn}(C)$, for any commutative ring C.*

Proof. (18.6) yields a ring homomorphism

$$\bar{\Phi}\colon \mathrm{End}_C C^{(m)} \otimes \mathrm{End}_C C^{(n)} \to \mathrm{End}_C C^{(mn)}. \tag{18.7}$$

Let us compare bases of both sides, taking a standard base $\{e_1, \dots, e_m\}$ of $C^{(m)}$. Using Proposition 13.43(i), we define $f_{ij} \in \mathrm{End}_C C^{(m)}$ sending $e_i \mapsto e_j$, yielding the base $\{f_{ij} \in \mathrm{End}_C C^{(m)}\colon 1 \leq i, j \leq m\}$. Analogously, taking a standard base $\{e'_1, \dots, e'_n\}$ of $C^{(n)}$, we get the base $\{f'_{k\ell} \in \mathrm{End}_C C^{(n)} : 1 \leq k, \ell \leq n\}$ of $\mathrm{End}_C C^{(m)}$ where $f'_{k\ell}$ sends $e'_k \mapsto e'_\ell$. Thus, $f_{ij} \otimes f'_{k\ell}$ sends $e_i \otimes e'_k$ to $e_j \otimes e'_\ell$. But by Corollary 18.13, the $e_i \otimes e'_k$ (resp. $e_j \otimes e'_\ell$) constitute a base of $C^{(m)} \otimes_C C^{(n)}$, which is isomorphic to $C^{(mn)}$, so $\bar{\Phi}$ sends a base of $\mathrm{End}\, C^{(m)} \otimes \mathrm{End}\, C^{(n)}$ to a base of $\mathrm{End}\, C^{(mn)}$ and thus is an isomorphism. We are done in view of the identification $M_m(C) \cong \mathrm{End}_C C^{(m)}$. □

Remark 18.37. Let us describe the isomorphism $\bar{\Phi}$ of (18.7) in matrix notation. Our base of $C^{(m)} \otimes C^{(n)} \cong C^{(mn)}$ is

$$e_1 \otimes e'_1, \dots, e_1 \otimes e'_n,\ e_2 \otimes e'_1, \dots, e_2 \otimes e'_n,\ \dots\dots,\ e_m \otimes e'_1, \dots, e_m \otimes e'_n.$$

Matrices $A = (a_{ik}) \in M_m(C)$ and $B = (b_{j\ell}) \in M_n(C)$ can be viewed, respectively, as the maps $e_i \mapsto \sum_k a_{ik} e_k$ and $e'_j \mapsto \sum_\ell b_{j\ell} e'_\ell$. Thus $\Phi(A \otimes B)$ sends

$$e_i \otimes e'_j \mapsto \sum_{k,\ell} a_{ik} b_{j\ell}\, e_k \otimes e'_\ell,$$

and the corresponding matrix, usually denoted $A \otimes B$, is

$$\begin{pmatrix}
a_{11}b_{11} & \dots & a_{11}b_{1n} & \dots & \dots & a_{1m}b_{11} & \dots & a_{1m}b_{1n} \\
a_{11}b_{21} & \dots & a_{11}b_{2n} & \dots & \dots & a_{1m}b_{21} & \dots & a_{1m}b_{2n} \\
\vdots & \ddots & \vdots & \dots & \dots & \vdots & \ddots & \vdots \\
a_{11}b_{n1} & \dots & a_{11}b_{nn} & \dots & \dots & a_{1m}b_{n1} & \dots & a_{1m}b_{nn} \\
a_{21}b_{11} & \dots & a_{21}b_{1n} & \dots & \dots & a_{2m}b_{1n} & \dots & a_{2m}b_{1n} \\
a_{21}b_{21} & \dots & a_{21}b_{2n} & \dots & \dots & a_{2m}b_{2n} & \dots & a_{2m}b_{2n} \\
\vdots & \ddots & \vdots & \dots & \dots & \dots & \ddots & \vdots \\
a_{21}b_{n1} & \dots & a_{21}b_{nn} & \dots & \dots & a_{2m}b_{nn} & \dots & a_{2m}b_{nn} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{m1}b_{11} & \dots & a_{m1}b_{1n} & \dots & \dots & a_{mm}b_{1n} & \dots & a_{mm}b_{1n} \\
a_{m1}b_{21} & \dots & a_{m2}b_{2n} & \dots & \dots & a_{mm}b_{2n} & \dots & a_{mm}b_{2n} \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
a_{m1}b_{n1} & \dots & a_{mm}b_{nn} & \dots & \dots & a_{mm}b_{n1} & \dots & a_{mm}b_{nn}
\end{pmatrix},$$

which can be rewritten in block form as

$$\begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mm}B \end{pmatrix}.$$

Tensor algebras.

Example 18.38. Given a commutative ring C and an R-module M, we write $M^{\otimes 0}$ for C, $M^{\otimes 1}$ for M, and $M^{\otimes i}$ for $M \otimes_C \cdots \otimes_C M$, with i tensor factors, $i \geq 2$. The natural identification

$$M^{\otimes i} \otimes M^{\otimes j} \cong M^{\otimes (i+j)}$$

yields a multiplication

$$M^{\otimes i} \times M^{\otimes j} \to M^{\otimes (i+j)}$$

given by $(a, b) \mapsto a \otimes b$. We define the **tensor algebra**

$$T(M) = \bigoplus_{i \geq 0} M^{\otimes i},$$

endowed with this natural multiplication. $T(M)$ is an associative algebra, in view of Proposition 18.15, and is $\mathbb{N}$-graded, where the i component is $M^{\otimes i}$.

PROPOSITION 18.39. *Let $F\colon C$-**Alg**$\to C$-**Mod** be the forgetful functor, i.e., forgetting the algebra multiplication. Then the tensor algebra $T(M)$ is the universal from M to F; cf. Definition 8.5 of Volume 1.*

Proof. We have to show that for any C-algebra R, any C-module map $f\colon M \to R$ extends to a unique algebra homomorphism $\widehat{f}\colon T(M) \to R$. But clearly uniqueness forces

$$\widehat{f}(a_1 \otimes \cdots \otimes a_i) = \widehat{f}(a_1) \cdots \widehat{f}(a_i) = f(a_1) \cdots f(a_i),$$

and this can be used as the formula to define $\widehat{f}$; one checks easily that this is an algebra homomorphism. □

This can also be viewed in terms of adjoint pairs; cf. Exercise 25.37. There are many cases in which we want to embed a module into an algebra, so this construction will prove very useful. Here are some famous illustrations.

Example 18.40. (i) (cf. Definition 17.5.) The free algebra $C\{X\}$ on a set of noncommuting indeterminates $X = \{x_i : i \in I\}$ is $T(M)$, where M is the free C-module with base X.

(ii) The **symmetric algebra** $S(M)$ of a module is defined as $T(M)/\mathcal{I}$, where $\mathcal{I}$ is the ideal generated by $\{v \otimes w - w \otimes v : v, w \in M\}$. This is easily seen to be the universal from M to the forgetful functor (from the category of commutative C-algebras to the category of C-modules). Indeed, for any commutative C-algebra R, by Proposition 18.39, any module map $M \to R$ extends to a unique homomorphism $T(M) \to R$ whose kernel contains $\mathcal{I}$, so we get an induced homomorphism $T(M)/\mathcal{I} \to R$.

When M is the free module $C^{(n)}$, the symmetric algebra $S(M)$ is naturally isomorphic to the commutative polynomial algebra $C[\lambda_1, \dots, \lambda_n]$.

(iii) The **Clifford algebra** $C(V, Q)$ of a quadratic space (V, Q) is the algebra $T(V)/\mathcal{I}$, where $\mathcal{I}$ is the ideal generated by $\{v \otimes v - Q(v) : v \in V\}$. It can be viewed as a universal; cf. Exercise 9.

(iv) The **Grassmann algebra** or **exterior algebra**, $E(V)$, is the special case of (iii) where $Q = 0$. (This is easily generalized to any module V over a commutative ring C.) In other words, $v^2 = 0$ in $E(V)$ for each $v \in V$. But then $0 = (v_1 + v_2)^2 = v_1^2 + v_2^2 + v_1 v_2 + v_2 v_1$, implying that $v_1 v_2 + v_2 v_1 = 0$, and thus

$$v_1 v_2 = -v_2 v_1, \qquad \forall v_i \in V. \tag{18.8}$$

Now $E(V)$ can be described explicitly. Taking a base $\{e_i : i \in I\}$ of V, we well-order I and note that $E(V)$ is spanned by the elements 1 and

$$\{e_{i_1} \cdots e_{i_m} : i_1 < i_2 < \cdots < i_m\}.$$

In fact, this comprises a base; cf. Exercise 11. For any two words w, w' of respective lengths m, n, we have

$$ww' = (-1)^{mn} w'w, \tag{18.9}$$

seen by induction on length; in particular, the even words are in the center.

The exterior algebra plays a very important role in differential geometry (where the product in $E(V)$ often is called the **wedge product**, denoted as $\wedge$), as well as in the theory of graded algebras and in polynomial identities (Chapter 23).

An application to algebraic geometry.

Tensor products play a fundamental role in the study of coordinate algebras, which we record here for further use in Appendix 19B. First, an innocuous fact.

PROPOSITION 18.41. *The tensor product $A \otimes_F B$ of two integral domains A and B over an algebraically closed field F is an integral domain.*

Proof. If not, suppose $rs = 0$, where $r = \sum a_i \otimes b_i$ and $s = \sum c_j \otimes d_j$. By Remark 18.32, one may assume that the a_i are linearly independent over F, and likewise for the c_j. Replacing A by the subalgebra generated by the a_i and c_j, and replacing B by the subalgebra generated by the b_i and d_j, we may assume that A and B are affine. Any homomorphism $\varphi\colon B \to F$ yields

$$0 = (1_A \otimes \varphi)(rs) = \sum a_i c_j \varphi(b_i d_j) = \sum a_i \varphi(b_i) \sum c_j \varphi(d_j),$$

implying that $\sum a_i \varphi(b_i) = 0$ or $\sum c_j \varphi(d_j) = 0$. Hence, either $\varphi(b_i) = 0$ for each i or $\varphi(d_j) = 0$ for each j. Note, by Proposition 6.14 of Volume 1, that every maximal ideal of B is the kernel of some φ and so contains the b_i or the d_j. Let $\mathcal{P}_1 = \{\text{maximal ideals of } B \text{ containing all } b_i\}$, $\mathcal{P}_2 = \{\text{maximal ideals of } B \text{ containing all } d_j\}$, and $N_u = \bigcap\{P \in \mathcal{P}_u\}$ for $u = 1, 2$. Then $N_1 N_2 \subseteq N_1 \cap N_2 = 0$ by Proposition 6.37 of Volume 1, implying that $N_1 = 0$ or $N_2 = 0$. Hence all $b_i = 0$ or all $d_j = 0$, implying that $r = 0$ or $s = 0$. □

PROPOSITION 18.42. *If X and Y are affine varieties over an algebraically closed field F, then $X \times Y$ is an affine variety, with $F[X] \otimes F[Y] \cong F[X \times Y]$. Under this correspondence, the following facts are true:*

(i) *If $f \in F[X]$ and $g \in F[Y]$, then $(f \otimes g)(x, y) = f(x)g(y)$.*

(ii) *If $X = Y$ and $\bar{\mu}\colon F[X] \otimes F[X] \to F[X]$ is the multiplication map, then for any $h \in F[X] \otimes F[X]$,*

$$\bar{\mu} h(x) = h(x, x).$$

Proof. We display $F[X]$ and $F[Y]$ using different indeterminates λ_i, i.e., $F[X] = F[\lambda_1, \ldots, \lambda_m]/\mathcal{I}(X)$ and $F[Y] = F[\lambda_{m+1}, \ldots, \lambda_{m+n}]/\mathcal{I}(Y)$. Now identifying $F[\lambda_1, \ldots, \lambda_{m+n}]$ with $F[\lambda_1, \ldots, \lambda_m] \otimes F[\lambda_{m+1}, \ldots, \lambda_{m+n}]$ via Example 18.24 (iterated), we obtain a natural onto homomorphism $\varphi\colon F[X] \otimes F[Y] \to F[X \times Y]$, and both are integral domains corresponding to the algebraic set $X \times Y$; hence $\ker \varphi = 0$, and φ is an isomorphism.

(i) follows since $(f, g)(x, y) = f(x)g(y)$.

(ii) can be verified by assuming $h = f \otimes g$, in which case

$$\bar{\mu} h(x) = f(x)g(x) = (f \otimes g)(x, x) = h(x, x),$$

by (i). □

See Exercises 19–21 for variations on this theme.

The adjoint isomorphism.

There is a sublime connection between tensor products and the functor Hom, which, although based on a general categorical principle to be given in Definition 25.6, can be stated rather quickly.

Remark 18.43. If R, S, and T are rings, M is an R, S-bimodule, and N is an R, T-bimodule, then $\mathrm{Hom}_R(M, N)$ is an S, T-bimodule, where for $f\colon M \to N$, $s \in S$, and $t \in T$, we define $sf, ft\colon M \to N$ by

$$(sf)(a) = f(as); \qquad (ft)(a) = f(a)t.$$

(Indeed, there are three straightforward verifications:

$$((s_1 s_2)f)(a) = f(as_1 s_2) = (s_2 f)(as_1) = (s_1(s_2 f))(a);$$

$$((sf)t)(a) = (sf(a))t = (f(as))t = (ft)(as) = (s(ft))(a);$$

$$(f(t_1 t_2))(a) = f(a)t_1 t_2 = (ft_1)(a)t_2 = ((ft_1)t_2)(a).)$$

PROPOSITION 18.44. *Given rings R, S, T, and W, suppose A is an R, S-bimodule, B is an S, T-bimodule, and C is an R, W-bimodule. Given $b \in B$ and $f \in \mathrm{Hom}_R(A \otimes_S B, C)$, define $f_b\colon A \to C$ by $f_b(a) = f(a \otimes b)$. Then there is an isomorphism of T, W-bimodules*

$$\Phi : \mathrm{Hom}_R(A \otimes_S B, C) \to \mathrm{Hom}_S(B, \mathrm{Hom}_R(A, C))$$

given by $\Phi(f) : b \mapsto f_b$.

Proof. First, using Remark 18.43, we note that $\mathrm{Hom}_R(A, C)$ is an S, W-bimodule, implying that $\mathrm{Hom}_S(B, \mathrm{Hom}_R(A, C))$ is a T, W-bimodule. Likewise, by Proposition 18.7 and Remark 18.9′, $A \otimes_S B$ is an R, T-bimodule, so $\mathrm{Hom}_R(A \otimes_S B, C)$ is also a T, W-bimodule, and checking the definitions one sees easily that Φ is a T, W-bimodule map. It remains to construct Φ^{-1}. Toward this end, given $g\colon B \to \mathrm{Hom}_R(A, C)$, define the balanced map $A \times B \to C$ by $(a, b) \mapsto g(b)(a)$. This gives a map $A \otimes B \to C$, which is $\Phi^{-1}(g)$. □

The notation of Proposition 18.44 is a little bothersome. If the ring R is commutative, then every R-module is a bimodule, and Proposition 18.44 becomes much easier to state. In general, given a C-algebra R, here is a sublime way to pass back and forth from modules to bimodules. We write R^e for $R \otimes_C R^{\mathrm{op}}$.

Remark 18.45. Any R, R-bimodule M is an R^e-module, under the multiplication

$$(r_1 \otimes r_2)a = r_1 a r_2.$$

Conversely, any R^e-module N is an R, R-bimodule under the multiplications

$$r_1 a = (r_1 \otimes 1)a; \qquad a r_2 = (1 \otimes r_2)a.$$

In particular R, being an R, R-bimodule, is an R^e-module. Furthermore, R^e is a module over itself and thus an R, R-bimodule, with multiplication $r(r_1 \otimes r_2) = r r_1 \otimes r_2$ and $(r_1 \otimes r_2)r = r_1 \otimes r_2 r$ for $r \in R$ and $r_1 \otimes r_2 \in R^e$.

Remark 18.46. The ideals of R are precisely the R, R sub-bimodules, and thus the R^e-submodules, of R. This enables us to apply module theory to the structure of ideals, sometimes to significant effect. We pursue this further in Appendix 25B.

Exercises – Part IV

Chapter 13

1. Suppose R is any ring and W is a free R-module with base $B = \{b_i : i \in I\}$. Given a binary operation $\mu: B \times B \to W$, extend μ to multiplication on W by the rule
$$\sum_i r_i b_i \sum_j s_j b_j = \sum_{i,j} (r_i s_j) b_i b_j$$
for $r_i, s_j \in R$. If $(b_1 b_2) b_3 = b_1 (b_2 b_3)$, $\forall b_i \in B$, this multiplication is associative and W thereby becomes a ring. In this way, conclude that condition (MU1) of Definition 13.3 "determines" multiplication in $M_n(R)$.
2. For $W = M_n(R)$, show that $We_{11} = We_{21}$, and $W(e_{11} + e_{22}) = We_{11} + We_{22}$, but $W(e_{11} + e_{12}) \neq We_{11} + We_{12}$.
3. If ab is invertible in R, show that a is *right* invertible but not necessarily left invertible. In fact, if V is an infinite-dimensional vector space over a field F, then $\mathrm{End}_F V$ has an element a such that $ab = 1$ but $ba \neq 1$. This can be seen in a colorful setting known as **Hilbert's Hotel**. Suppose an (infinite) hotel is fully booked, and suddenly a VIP arrives. How can the manager take care of the VIP without putting out any of the guests? The manager tells the occupants of room 1 that they must move on to room 2, and similarly bumps the occupants of room 2 to room 3, etc.

Idempotents

4. In any domain, show that $ab = 1$ implies $ba = 1$. Give an example in a ring, where ab is idempotent but ba is not idempotent. (But $(ba)^2$ must be idempotent.)

5. If $a+b = 1$ and $ab = 0$, show that a and b are idempotents. Generalize to $\sum_{i=1}^n a_i = 1$.
6. For the next two exercises, say that a central idempotent e of a ring R is **simple** if Re is simple (as a ring with unit element e). Show for any simple central idempotent e and any other nonzero central idempotent f of R, that either e and f are orthogonal or $ef = e$.
7. Suppose $R = R_1 \times \cdots \times R_n$, a direct product of simple rings. Show that the decomposition of R into a direct product of simple rings is unique in the sense that R has a unique 1-sum set of orthogonal simple central idempotents. In fact, if $S = \{e_1, \ldots, e_n\}$ is a 1-sum set of orthogonal simple central idempotents, then *every* simple central idempotent of R belongs to S. (Hint: It suffices to prove the last assertion, in view of Exercise 6.)

Peirce decomposition

8. For any simple ring R, show that the ring eRe is simple for any idempotent e. (Hint: If $A \triangleleft eRe$, then A plus the other Peirce components is an ideal of R.)
9. ("Peirce decomposition for n orthogonal idempotents") For any 1-sum set of orthogonal idempotents $e_1, \ldots, e_n$, define the **Peirce decomposition**
$$R = \bigoplus_{i,j=1}^n e_i R e_j,$$
generalizing Remark 13.6. Verify that the Peirce decomposition of $R = M_n(T)$ with respect to $\{e_{11}, \ldots, e_{nn}\}$ is just $R = \bigoplus_{i,j=1}^n T e_{ij}$.
10. If e and f are idempotents of R with $fe = 0$, show that e and $(1-e)f$ are orthogonal idempotents, and moreover $Rf = R(1-e)f$. (Hint: $f = f(1-e)f$.)

Idempotents and Hom

11. Show that any map $f: Re \to M$ of R-modules is given by right multiplication by the element $f(e) = ef(e) \in eM$. In particular, $f(Re) \subseteq ReM$.
12. If e, e' are idempotents of a ring R, then $\mathrm{Hom}(Re, Re') \cong eRe'$ as Abelian groups. (Hint: Define the map $\Phi: \mathrm{Hom}(Re, Re') \to eRe'$ via $\Phi(f) = f(e) = ef(e) \in eRe'$.)

Minimal left ideals (also cf. Exercises 15.19ff.)

13. In any domain R that is not a division ring, show that $La < L$ for any left ideal L and any noninvertible element a of L. (Hint: $a \notin La$.) Thus, R has no minimal left ideals.
14. Show that $M_n(F)a$ is a minimal left ideal of the matrix algebra $M_n(F)$ over a field F, iff the matrix a has rank 1.

The opposite ring

15. Show that $M_n(C) \cong M_n(C)^{\mathrm{op}}$ for any commutative ring C.
16. Let $R = \begin{pmatrix} \mathbb{Q} & \mathbb{R} \\ 0 & \mathbb{R} \end{pmatrix}$. Prove that R has infinite ascending and descending chains of left ideals, since $\begin{pmatrix} 0 & V \\ 0 & 0 \end{pmatrix}$ is a left ideal of R, for every $\mathbb{Q}$-subspace V of $\mathbb{R}$. (On the other hand, as a right R-module, R has composition length 3. Find a composition series.) In particular, R is right Artinian and right Noetherian, but neither left Artinian nor left Noetherian. Conclude that R is not isomorphic to R^{op}.

The regular representation

17. Suppose R has a subring W such that R is a free right W-module of rank n. Then there is an injection of R into $M_n(W)$.
18. Notation as in Exercise 17, suppose $b_1, \dots, b_n$ is a base of R over W. Then $\{\sum_{i=1}^n b_i w_{ij} : 1 \leq i, j \leq n\}$ span R iff the matrix (w_{ij}) is right invertible.
19. For any f.g. right W-module $M = \sum_{i=1}^n a_i W$, prove that the ring $\operatorname{End} M_W$ is a homomorphic image of a subring R of $M_n(W)$. (Hint: A modification of Proposition 13.42. R is the set of matrices that acts on the set $a_1, \dots, a_n$ like suitable endomorphisms of M; the homomorphism $R \to \operatorname{End} M_W$ sends the matrix to its corresponding endomorphism.)
20. Suppose R has a subring W such that R is generated by n elements as a right W-module. Prove that R is isomorphic to a homomorphic image of some subring of $M_n(W)$. (Hint: Apply the regular representation to Exercise 19.)

PLID's (following results from Volume 1)

21. Show that every submodule of a free module over a PLID is free. (Hint: As in Theorem 2.40 of Volume 1.)
22. Generalize Corollary 2.41 of Volume 1 to f.g. modules over a PLID.
23. Show that every f.g. module over a PLID is isomorphic to a direct sum of cyclic modules, given as in Theorem 2.54 of Volume 1.

Power series rings in one indeterminate

24. Show that the power series ring $R[[\lambda]]$ is a domain when R is a domain; also show that $R[[\lambda]]$ is Noetherian when R is Noetherian. (Hint: Use lowest order monomials instead of highest order monomials.)
25. Show that any power series in $R[[\lambda]]$ whose constant term is invertible, is invertible in $R[[\lambda]]$. Conclude that if D is a division ring, then $D[[\lambda]]$ has a unique maximal left ideal I which also satisfies $I \lhd D[[\lambda]]$ and $D[[\lambda]]/I \cong D$.

Appendix 13A

1. Show that $\operatorname{Cent}(\mathcal{A}_1(F)) = F[\lambda^p, \mu^p]$ for any field F of characteristic p; conclude that $\mathcal{A}_1(F)$ is f.g. over its center, spanned by $\{\lambda^i \mu^j : 0 \leq i, j < p\}$.
2. Show that Dixmier's conjecture D_n implies the Jacobian conjecture for $F[\lambda_1, \dots, \lambda_n]$. (Hint: Define the homomorphism given by $\lambda_i \mapsto f_i$, $\mu_i \mapsto \sum_{j=1}^n M_{ij}\mu_j$, where M_{ij} is the i, j minor of the Jacobian matrix $J(f_1, \dots, f_n)$.)

Ore extensions

3. Given an injection $\sigma\colon R \to R$, define a **σ-derivation** to be a C-module map $\delta : R \to R$ satisfying

$$\delta(ab) = \sigma(a)\delta(b) + \delta(a)b.$$

(Compare with Definition 6B.1 of Volume 1.) Given a σ-derivation δ, define the **Ore extension** $R[\lambda; \sigma, \delta]$ to have the R-module structure of $R[\lambda]$ but multiplication given by

$$\lambda r = \sigma(r)\lambda + \delta(r).$$

Prove that $R[\lambda; \sigma, \delta]$ is a ring, by means of the regular representation, and show that the natural degree function with respect to λ satisfies

$$\deg fg = \deg f + \deg g. \tag{13E.1}$$

Ore extensions include skew polynomial rings (taking $\delta = 0$). On the other hand, when $\sigma = 1$, the Ore extension is called a **differential polynomial ring**.

4. Conversely to Exercise A3, show that when the R-module $T = R[\lambda]$ has a multiplication making it into a ring having a degree function satisfying (13E.1), then T is an Ore extension of the ring R. (Hint: Writing $\lambda r = r'\lambda + r''$, define $\sigma(r) = r'$ and $\delta(r) = r''$.)
5. Let δ be the usual derivative with respect to μ, taken on the polynomial algebra $R = F[\mu]$. (Thus, $\delta(\mu) = 1$.) Show that $\mathcal{A}_1(F)$ is isomorphic to the differential polynomial ring $R[\lambda; \delta]$.
6. In the Ore extension $R[\lambda; \sigma, \delta]$, show that the injection σ is an automorphism, iff every element can also be written in the form $\sum \lambda_i r_i$, for $r_i \in R$.
7. Prove that any Ore extension of a left Noetherian ring is left Noetherian. More generally, if a ring W is generated by a left Noetherian subring R and an element a such that $W = R + aR = R + Ra$, then W also is left Noetherian. (Hint: To show that any left ideal L of

W is f.g., define $L_u = \{r_u \in R : \sum_{i=0}^{u} a^i r_i \in L$, for suitable $r_i \in R\}$. L_u is a left ideal, and one can now mimic the proof of Theorem 7.18 of Volume 1.)

8. Show that Ore extension satisfies the following version of Euclid's algorithm: if $f, g \in R[\lambda, \sigma, \delta]$ and the leading coefficient of g is invertible, then one can find q, r in $R[\lambda, \sigma, \delta]$ with $r = 0$ or $\deg r < \deg g$ such that $f = qg + r$. (Hint: By induction on degree.) Conclude that any Ore extension $D[\lambda; \sigma, \delta]$ of a division ring D is a PLID.
9. Let $F = \text{Cent}(D)$. Show that the skew polynomial ring $\text{Cent}(D[\lambda; \sigma])$ has center F unless some power σ^n of σ is inner, in which case the center is $F[a\lambda^n]$ for suitable $a \in D$. In both cases each ideal I of $D[\lambda; \sigma]$ has the form $D[\lambda; \sigma] f \lambda^m$ for suitable $f \in \text{Cent}(D[\lambda; \sigma])$ and $m \geq 0$. (Hint: As in Example 13A.3. For the last assertion, write $I = D[\lambda; \sigma]f$ where f has minimal degree and minimal number of nonzero terms; one may assume that f is monic. Compute the ring commutators $[\lambda, f]$ and $[d, f]$, $\forall d \in D$, which are in I and thus 0.)
10. Suppose K is a field with automorphism σ, and F is the fixed subfield. If f is an irreducible polynomial of $F[\lambda]$, show that $K[\lambda, \sigma]/\langle f(\lambda) \rangle$ is a simple ring.
11. If the derivation δ is not inner and $\text{char}(D) = 0$, show that the differential polynomial ring $D[\lambda, \delta]$ is a simple ring.

Chapter 14

1. Suppose $n = p_1^{i_1} \cdots p_t^{i_t}$ is the prime factorization of n. Then show that $\text{soc}(\mathbb{Z}/n\mathbb{Z}) = k\mathbb{Z}/n\mathbb{Z}$, where $k = \frac{n}{p_1 \cdots p_t}$.
2. Show that $\text{soc}(R) = \cap\{\text{nonzero ideals of } R\}$ when R is prime with $\text{soc}(R) \neq 0$.
3. Prove that $\text{soc}(R) \lhd R$, for any ring R. (Hint: One needs to check for each minimal left ideal L that Lr is a minimal left ideal or 0.)
4. Prove that a module M is semisimple iff M is the direct sum of simple submodules. (Hint: Define a relation saying that $N_1 \prec N_2$ if $N_1 < N_2$ and $N_2 = N_1 \oplus S$ where S is a direct sum of simple submodules. Apply Zorn's lemma and the reasoning in the proof of Theorem 14.16.)
5. Reprove Exercise 4 by observing that independence of submodules defines an abstract dependence relation on the set of simple submodules of a given module M; $\text{soc}(M)$ has a base of simple submodules.
6. Fill in the details for the following short argument for (3) $\Rightarrow$ (1) of the Wedderburn-Artin Theorem: Suppose R is a simple ring having a minimal left ideal L. Write $1 = \sum_{i=1}^{n} a_i r_i$ for $a_i \in L_i$, with n minimal. The map $L^{(n)} \to R$ defined by $(b_1, \ldots, b_n) \mapsto \sum b_i r_i$ is onto, and is 1:1 since L is a simple module. Hence, $R \cong \text{End}_R L^{(n)} \cong M_n(\text{End}_R L)$.

Large submodules

7. Show that the intersection of all large $\mathbb{Z}$-submodules of $\mathbb{Z}$ is 0.
8. For any module M, prove that $\text{soc}(M) = \bigcap\{\text{Large submodules of } M\}$. (Hint: $\bigcap\{\text{Large submodules of } M\}$ is complemented, since any submodule has an essential complement in M.)

Quaternion algebras

9. (The generalized quaternion algebra) Generalizing Example 14.29, define the algebra of **generalized quaternions** $(\alpha, \beta)_2$ over an arbitrary field F as $F + Fx + Fy + Fz$ where $xy = -yx = z$, $x^2 = \alpha$, and $y^2 = \beta$. (Then $z^2 = -\alpha\beta$). This algebra plays an important role in Chapter 24.
10. If σ is the automorphism of $\mathbb{H}$ given by conjugation by i, what is the structure of $\text{Cent}(\mathbb{H}[\lambda; \sigma])$?

Rings with involution

11. Suppose V is a finite-dimensional vector space with a bilinear form $\langle\ ,\ \rangle$ over the field F, and let $R = \text{End}_F V$. Show that the map $T \mapsto T^*$ sending a transformation to its adjoint (cf. Remark 0.20 of Volume 1) defines an involution of R. If $\langle\ ,\ \rangle$ is symmetric (resp. Hermitian) and V has an orthonormal base $\{e_1, \ldots, e_n\}$, this gives us Example 14.33(ii) (resp. (iv)). (Hint: e_{ij} sends e_i to e_j.)
12. Apply Exercise 11 to alternate bilinear forms to obtain the canonical symplectic involution (Example 14.33(iii)).
13. As a special case of Exercise 12, show that the canonical symplectic involution on $M_2(F)$ is given by $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^* = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$; the only matrices symmetric under this involution are scalar, and thus r^*r must be scalar for each $r \in M_2(F)$. In fact $r^*r = \det r \cdot I$; hence, $r^{-1} = \frac{1}{\det r} r^*$ for any nonsingular matrix r.
14. If $\frac{1}{2} \in R$, show that $R = S \oplus K$. (Hint: $r = \frac{r + r^*}{2} + \frac{r - r^*}{2}$.)
15. In Example 14.33(ii), show that $\dim S = \frac{n(n+1)}{2}$ and $\dim K = \frac{n(n-1)}{2}$. For the canonical symplectic involution, show that $\dim S = \frac{n(n-1)}{2}$ and $\dim K = \frac{n(n+1)}{2}$.
16. Suppose $\frac{1}{2} \in R$. If S is the set of symmetric elements of R with respect to $(*)$, and $a \in S$, show that Sa is the set of symmetric elements with respect to the new involution of Example 14.33(iii). What happens if $a^* = -a$?

Structure theory for rings with involution

In Exercises 17–21, we use universal algebra to develop a structure theory for rings with involution $(R, *)$. A $(*)$**-homomorphism** $f\colon (R, *) \to (W, *)$ is a ring homomorphism $f\colon R \to W$ such that

$f(r^*) = f(r)^*$ for all $r \in R$. We write $A \lhd (R, *)$, called a $(*)$**-ideal**, if $A \lhd R$ with $A^* \subseteq A$.

17. Show that the $(*)$-ideals are just the kernels of the $(*)$-homomorphisms.
18. For any ring R, define $R \times R^{\rm op}$, together with the **exchange involution** $\circ$ given by $(a,b)^\circ = (b,a)$. Show that the $(\circ)$-ideals are precisely of the form $A \times A^{\rm op}$ for $A \lhd R$. This provides a faithful functor from the category of rings to the category of rings with involution.
19. Prove that $(R, *)$ is simple (i.e., R has no trivial $(*)$-ideals) iff R has a maximal ideal P such that $P \cap P^* = 0$. In this case, either R is simple or $(R, *) \cong (W \times W^{\rm op}, \circ)$ of Exercise 18, where W is a simple ring. (Hint: $(\Leftarrow)$ Use the CRT.)
20. For V and R as in Exercise 11, show that each isometry of V gives rise to an isomorphic ring with involution. In particular, if the bilinear form is alternate, then $R \cong (M_n(F), J)$ where J is the canonical symplectic involution. (Hint: Exercise 0.21 of Volume 1 and Exercise 12.)
21. Suppose R is simple and finite-dimensional over an algebraically closed field F, and R has an involution $(*)$. Show that $(R, *) \cong (M_n(F), J)$, where J is either the transpose or the canonical symplectic involution, and in the latter case n is even.

Chapter 15

Characterizations of primitive rings via the core

1. For any left ideal L of R, show that $\operatorname{Ann}_R R/L$ is the unique largest ideal of R contained in L. This is called the **core** of L.
2. Show that an ideal is primitive iff it is the core of a maximal left ideal. In particular, a ring R is primitive iff R has a maximal left ideal with core 0.
3. Say that a left ideal L of a ring R is **comaximal with prime ideals** if $L + P = R$ for every nonzero prime ideal P of R. Prove that R is primitive iff R has a left ideal L comaximal with prime ideals. (Hint: Any maximal left ideal containing L has core 0.)
4. Suppose that there is some non-nilpotent $r \in R$ such that every ideal of R contains a power of r. Show that either $1 - r$ is left invertible or the ring R is primitive. (Hint: Consider $R(1 - r)$.)
5. If K is a field with a homomorphism $\sigma\colon K \to K$ of infinite order, show that the skew polynomial ring $R = K[\lambda; \sigma]$ of Example 13A.3 is primitive. (Hint: $R(\lambda - 1)$ is comaximal with prime ideals.)
6. Prove that any prime ring having a faithful module of finite composition length is primitive.

Elements of finite rank

7. Suppose M is a module over a division ring D. Define the **rank** of an element $f \in W = \operatorname{End} M_D$ to be the dimension of $f(M)$ over D. Show that Wf is a minimal left ideal of W iff f has rank 1. Also, the set of elements of W having finite rank is an ideal of W, which is precisely $\operatorname{soc}(W)$.
8. In Exercise 7, if M is infinite-dimensional over D, show that the primitive ring $W = \operatorname{End} M_D$ is not simple.
9. Give an example of a primitive ring with socle 0. (Hint: Take any simple ring that is not Artinian. Alternatively, see Exercise 10.)
10. Define transformations x, y on a vector space V with infinite base $\{b_1, b_2, \dots\}$ by $xb_1 = 0$, $xb_{i+1} = b_i$, and $yb_i = b_{i^2+1}$. Show that x and y generate a primitive subalgebra of $\operatorname{End} V$ having socle 0.
11. Over a field F, define R to have the F-vector space structure of the commutative polynomial algebra $F[\lambda, \mu]$ but with multiplication satisfying $\mu\lambda - \lambda\mu = \lambda$. Display R as a skew polynomial ring, and show that it is primitive.
12. Embedding any algebra R into the primitive algebra $\operatorname{End}_F R$ via the left representation, show that $R + \operatorname{soc}(\operatorname{End}_F R)$ is a primitive algebra with nonzero socle. This observation provides many interesting examples of primitive rings.

Semiprime rings

13. Prove that the following conditions are equivalent for a ring R: (i) R is semiprime; (ii) $L^2 \neq 0$ for any nonzero left ideal L of R; (iii) $aRa \neq 0$ for any $0 \neq a \in R$. (Hint: Modify Remark 15.12.)
14. Suppose L is a left ideal of a ring R such that, for some $n \geq 2$, $a^n = 0$ for all $a \in L$. Show that $(Ra^{n-1}R)^2 = 0$, $\forall a \in L$. Conclude by induction that $L = 0$ if R is semiprime. (Hint: For any $r \in R$, let $r' = ra^{n-1}$. Then $r'a = 0$ and $(r')^n = 0$. But $0 = (r' + a)^n$; compute this to be $(1+b)a^{n-1}r'$, where $b \in Lr$ is nilpotent, implying that $1+b$ is invertible. Hence, $0 = a^{n-1}r' = a^{n-1}ra^{n-1}$.)

Semiprime rings with involution

15. Prove that $(R, *)$ has no nonzero nilpotent $(*)$-ideals iff R is a semiprime ring. (Hint: $(\Rightarrow)$ If $A^2 = 0$, then $(A + A^*)^3 = 0$.)
16. Suppose R is a ring with involution $(*)$, satisfying the condition that if $r^*r = 0$ then $r = 0$. If r^*r is nilpotent, show that $r = 0$.
17. Under the conditions of Exercise 16, conclude that R is semiprime. (Hint: If $A \lhd R$ with $A^2 = 0$, then for $a \in A$, $(a^*a)^*(a^*a) \in A^2 = 0$.)
18. Show that any $\mathbb{R}$-subalgebra of $M_n(\mathbb{C})$ that is invariant under the Hermitian transpose must be semisimple. (Hint: Apply Exercise 17.)

Minimal left and right ideals of semiprime rings

19. If Re is a minimal left ideal of a semiprime ring R, for e idempotent, show that eR is a minimal right ideal. (Hint: Digression 14.21(iv).)
20. Show that any minimal left ideal of a semiprime ring R has the form $L = Re$ for a suitable idempotent e. (Hint: $L = La$, for some $a \in L$, so $a = ea$ for some $e \in L$. Then $R(e^2-e)a = 0$, yielding $R(e^2-e) = 0$.)
21. For any semiprime ring, prove that $\operatorname{soc}(R)$ is also the sum of the minimal right ideals of R. (Hint: Combine Exercises 19 and 20.)

Elementary properties of left Artinian rings

22. In a left Artinian ring R, prove that if $\operatorname{Ann}_R r = 0$, then r is left invertible. (Hint: Consider $Rr \supseteq Rr^2 \supseteq \cdots$.)
23. Prove that $ab = 1$ implies $ba = 1$ in a left Artinian ring. (Hint: Use Exercise 22.)

The Jacobson radical

An element $a \in R$ is called **quasi-invertible** if $1 \in R(1-a)$. A set $A \subset R$ is called **quasi-invertible** if each element is quasi-invertible.

24. Prove that $\operatorname{Jac}(R)$ is a quasi-invertible ideal that contains every quasi-invertible left ideal of R, and thus is the unique largest quasi-invertible left ideal of R. (Hint: If $A < R$ is quasi-invertible, then any proper maximal left ideal L contains A, since otherwise $1 \in A+L$, implying that L contains some element $1 - a$.)
25. Show that $1 - a$ is invertible for every element a for every element A in a quasi-invertible left ideal of R. (Hint: If $r(1 - a) = 1$, then $r = 1 + ra$ also has a left inverse b, so $b = 1 - a$.)
26. Prove that $\operatorname{Jac}(R)$ is the intersection of all maximal right ideals of R. (Hint: Apply right-left symmetry to Exercise 24.)
27. Prove that $\operatorname{Jac}(R)$ contains every nil left ideal and every nil right ideal of R. (Hint: Every nilpotent element is quasi-invertible.)
28. Show that $\operatorname{Jac}(R)/A \subseteq \operatorname{Jac}(R/A)$, for any $A \lhd R$; equality holds if $A \subseteq \operatorname{Jac}(R)$. If $\operatorname{Jac}(R/A) = 0$, then $A \supseteq \operatorname{Jac}(R)$.
29. Show that $\operatorname{Jac}(R)$ contains no nontrivial idempotents. (Hint: $1-e$ is not invertible.)
30. Show that an element $a \in R$ is invertible iff $a + \operatorname{Jac}(R)$ is invertible in $R/\operatorname{Jac}(R)$.

Nil vs. nilpotent

31. (A commutative counterexample.) For each $n \in \mathbb{N}$, define $R_n = \mathbb{Z}/2^n$, let $I_n = 2R_n$, and let $R = \prod_{n\in\mathbb{N}} R_n$. Show that $I_n^n = 0$, and thus I_n is a nilpotent ideal of R (viewed in the proper component). Hence, $\sum I_n$ is a nil ideal of R that is not nilpotent, and $\prod I_n$ is not nil.

Local rings

A ring R is called **local** if $R/\operatorname{Jac}(R)$ is a division ring. The following exercises generalize Proposition 8.19 of Volume 1.

32. Show that a ring R is local iff it has a unique maximal left ideal L, which thus is $\operatorname{Jac}(R)$.
33. Prove that a ring R is local iff $a + b = 1$ implies a or b is invertible.
34. If every element of R is invertible or nilpotent, show that R is local. (Hint: Any proper left ideal is nil, and thus contained in $\operatorname{Jac}(R)$.)
35. Suppose $J = \operatorname{Jac}(R)$ is nil and R/J is simple Artinian. Prove that $R \cong M_n(T)$, where T is local. (Hint: Use Proposition 13.13.)

Appendix 15A

1. Suppose M is a faithful simple R-module, $D = \operatorname{End}_R M$, and M is infinite-dimensional over D. Show for every n that $M_n(D)$ is isomorphic to a homomorphic image of a subring W of R. (Hint: Take D-independent elements $b_1, \ldots, b_n$ of M, let $V = \sum_{i=1}^n b_i D$, and define $W = \{r \in R : rV \subseteq V\}$, a subring of R; apply density.)
2. Show that if R is primitive and every element of R is algebraic of degree $\leq n$ over a field, then R is simple Artinian. This exercise is a useful introduction to Kaplansky's Theorem in polynomial identities; cf. Chapter 23.

Normalizing extensions

Suppose $W \supset R$. A **normalizing element** of W is an element a such that $aR = Ra$. W is a **normalizing extension** of R if $W = \sum_{i\in I} Ra_i$ where each a_i is a normalizing element. W is a **finite normalizing extension** if the index set I is finite. For example, any skew polynomial ring $R[\lambda;\sigma]$ with σ onto is a normalizing extension but not a finite normalizing extension.

3. For any finite normalizing extension W of a ring R, show that any simple W-module is a finite direct sum of simple R-modules. (Hint: Taking $W = \sum_{i=1}^n Ra_i$ as above, write $M = Wb = \sum Ra_i b$, which by Zorn's Lemma has a maximal R-submodule N. Take $N_i = \{b' \in M : Ra_i b' \subseteq N\}$, a maximal R-submodule of M. Then $\bigcap N_i$ is also a W-submodule, so is 0.)

The Jacobson radical of overrings

4. Show that $\operatorname{Jac}(R) \subseteq \operatorname{Jac}(W)$ for any finite normalizing extension W of R. (Hint: One needs to show for each primitive ideal P of W that $P \cap R$ is an intersection of primitive ideals of R. But this follows from Exercise A3, since $P = \operatorname{Ann}_W M$, for M a simple W-module.)

5. Suppose R is a subring of W, and any element of R that is left invertible in W is already left invertible in R. Prove that $R \cap \mathrm{Jac}(W) \subseteq \mathrm{Jac}(R)$. (Hint: Use quasi-invertibility.)
6. Prove that $R \cap \mathrm{Jac}(W) \subseteq \mathrm{Jac}(R)$ whenever the ring R is a direct summand of W as an R-module. (Hint: Use Exercise A5.)

When is the Jacobson radical nil?

7. Suppose R is an Artinian subring of W, and every element of R that is left invertible in W is already left invertible in R. Show that $R \cap \mathrm{Jac}(W)$ is nilpotent. (Hint: By Exercise A5.)
8. For any algebra W over a field, prove that every element r of $\mathrm{Jac}(W)$ is either nilpotent or transcendental. (Hint: Apply Exercise A7, viewing $F[r] \subseteq W$.)

Jacobson radical of algebras over uncountable fields

9. (Amitsur) Prove that $\mathrm{Jac}(R)$ is nil whenever R is an algebra over an infinite field F satisfying the condition $\dim_F R < |F|$ (as infinite cardinals). (Hint: Take $r \in \mathrm{Jac}(R)$. Then $r - \alpha$ is invertible for each $0 \neq \alpha \in F$. It follows that r is algebraic, so apply Exercise A8.)
10. Prove that $\mathrm{Jac}(R)$ is nil whenever R is affine (cf. Definition 17.1) over an uncountable field.

Appendix 15B

Kolchin's Problem

Let D be a division ring. We view $M_n(D)$ as the endomorphism ring of the right D-module $V = D^{(n)}$. Suppose G is a subgroup of $\mathrm{GL}(n, D)$. Given a subspace S of V, define the **fixator** $\mathrm{Fix}(S) = \{g \in G : gv = v, \forall v \in S\}$.

1. If $(g-I)^k V = 0$ and $V_1 = (g-I)^{k-1}V \neq 0$, show that g fixes V_1. Thus, g unipotent implies $g \in \mathrm{Fix}(W)$ for some nonzero $V_1 \subseteq V$.
2. For $\mathrm{char}\, D = p$, show that an element $g \in G$ is unipotent iff $g^{p^t} = I$ whenever $p^t > n$. In particular, G is unipotent iff G is a p-group of bounded index. (Hint: $g^{p^t} - I = (g-I)^{p^t}$.)
3. Kolchin's Problem is trivial for $n = 1$. If one is to solve Kolchin's Problem by induction on n, show that it suffices to prove that V has a proper G-invariant subspace $W \neq 0$. In particular, one may assume that $GvD = V$ for every nonzero $v \in V$.
4. Show that a unipotent subgroup $G \subset \mathrm{GL}(n, D)$ can be put into simultaneous upper triangular form iff G is nilpotent as a group; in this case G has nilpotence index $\leq n$. (Hint ($\Rightarrow$) Take $0 \neq v \in V$ with $G \subseteq \mathrm{Fix}(v)$. Then G acts unipotently on V/vD, so apply induction to V_1 and V/V_1). ($\Leftarrow$) Apply Exercise B1 to a nontrivial element of $Z(G)$ and apply induction.)

5. In order to solve Kolchin's Problem for a monoid $G \subset \mathrm{GL}(n, D)$, one can replace D by any simple Artinian ring properly containing D. (Hint: Apply Exercise 4, matching coefficients in a base over D.) As a result, verify Kolchin's Problem whenever D is a finite-dimensional division algebra, by using the regular representation to embed D into a matrix algebra over a field.
6. For $H < G$ define $H_V = \{v \in V : hv = v,\ \forall h \in H\}$, a D-subspace of V. If $H = \mathrm{Fix}(S)$, then $S \subseteq H_V$ and $H = \mathrm{Fix}(H_V)$. Conclude that any chain of fixators has length $\leq n$, and in particular any fixator is contained in a maximal fixator.
7. Suppose $N \lhd G$ fixes the vector $v \in V$. Show that N fixes every vector in GvD. Furthermore, G/N acts naturally on vD, by $gN(vd) = gvd$.
8. Prove that Kolchin's Problem has an affirmative answer for solvable subgroups G of $\mathrm{GL}(n, D)$. (Hint: By induction on solvability degree. Assume by induction that G' is nilpotent and therefore fixes some $0 \neq v \in V$. But G/G' then acts on GvD and has some fixed vector v_1, so G also fixes v_1.)
9. Verify Kolchin's Problem for locally solvable groups and for locally metabelian groups.
10. Verify Kolchin's Problem for $n = 2$. (Hint: Otherwise G contains elements g_1 and g_2 with respective D-linearly independent eigenvectors v_1, v_2. Using v_1, v_2 as the base for $D^{(2)}$, one may assume that $g_1 = I + d_1 e_{12}$ and $g_2 = I + d_2 e_{21}$ with $d_1, d_2 \neq 0$; $g_1 g_2$ is not unipotent. This proof is misleadingly easy.)
11. (Derakhshan) For $\mathrm{char}\, D = p$, show that Kolchin's Problem has a positive solution for G iff, for any homomorphic image H of any subgroup of G, every two elements of H of order p generate a solvable subgroup. (Hint: Take two maximal fixators $\mathcal{F}_1 \neq \mathcal{F}_2$ chosen such that their intersection K has K_V minimal possible. Let $V_i = \mathcal{F}_{iV}$. Note that $K_V \supseteq V_1 + V_2$. By induction $\mathcal{F}_1$ and $\mathcal{F}_2$ are nilpotent groups, so the normalizer $\mathcal{N}$ of K contains elements $a_i \in \mathcal{F}_i \setminus K$ such that $a_i^p \in K$, for $i = 1, 2$. By hypothesis, $a_1 K$ and $a_2 K$ generate a solvable subgroup of $\mathcal{N}/K$, so K, a_1, a_2 generate a subgroup $\tilde{K}$ of G that is solvable and thus nilpotent by Exercises B8 and B4; hence it is contained in a maximal fixator $\mathcal{F}$. But a_1 does not fix V_2 since $a_1 \notin \mathcal{F}_2$; hence, $(\mathcal{F} \cap \mathcal{F}_1)_V \subseteq \tilde{K}_V \subset K_V$, a contradiction.)
12. (Derakhshan) Verify Kolchin's Problem in characteristic 2. (Hint: See Example 17A.2(i).)

Chapter 16

Prime ideals

1. Define the prime spectrum $\operatorname{Spec} R$ for any ring R (not necessarily commutative) and prove that Exercises 8.8–8.21 of Volume 1 carry over. $\operatorname{Spec} R$ is disconnected iff R has a nontrivial central idempotent.
2. The ring $\left(\begin{smallmatrix} \mathbb{Z} & m\mathbb{Z} \\ \mathbb{Z} & \mathbb{Z} \end{smallmatrix}\right)$ is prime Noetherian. What are its prime ideals?
3. An ideal $A \lhd R$ is called **semiprime** if R/A is a semiprime ring. Show for R satisfying ACC{semiprime ideals}, that every semiprime ideal A is a finite intersection of prime ideals. (Hint: Digression 9.2′ of Volume 1. If $P_1 \cdots P_t \subseteq A$ for prime ideals P_i, then $P_1 \cap \cdots \cap P_t \subseteq A$.)

The lower nilradical

4. (Levitzki) Prove that a ring R is semiprime iff $N(R) = 0$. (Hint: For any $r \neq 0$ in R, build a prime ideal P not containing r, as follows:

   Define $s_1 = r$ and, inductively, given s_i, take r_i in R such that $s_i r_i s_i \neq 0$, and define $s_{i+1} = s_i r_i s_i$. Take $P \lhd R$ maximal with respect to not containing any s_i.)
5. (The Baer radical, for those familiar with ordinals.) Define $N_0(R)$ to be the sum of all nilpotent ideals of R, and for each ordinal α define inductively:

   For $\alpha = \beta + 1$ (a successor ordinal), $N_\alpha(R)$ is that ideal I of R such that $I/N_\beta(R) = N_0(R/N_\beta(R))$;

   For α a limit ordinal, $N_\alpha(R) = \bigcup_{\beta<\alpha} N_\beta(R)$.

   At some stage $N_\alpha(R) = N_\gamma(R)$ for all $\gamma > \alpha$; define the **Baer radical** to be this $N_\alpha(R)$. Prove that the Baer radical equals $N(R)$. (Hint: Pass to $R/N_\alpha(R)$ and apply transfinite induction to Exercise 4.)

The upper nilradical

6. Say that a prime ring is **prime**$^+$ if it contains no nonzero nil ideals. A prime ideal of a ring R is **prime**$^+$ if the ring R/P is prime$^+$. The intersection of the prime$^+$ ideals is called the **upper nilradical** $\operatorname{Nil}(R)$. Show that $\operatorname{Nil}(R)$ is a nil ideal that contains all the nil ideals of R. Thus, $\operatorname{Nil}(R)$ is the sum of the nil ideals of R. (Hint: As in Proposition 16.8: Show for any non-nilpotent element r in R that there is a prime$^+$ ideal that does not contain r.)
7. Let us call a ring R **weakly primitive** if R is prime, with $\operatorname{Nil}(R) = 0$, and there exists non-nilpotent $r \in R$ such that every nonzero ideal of R intersects a power of r. Strengthening Exercise 6, show that if $\operatorname{Nil}(R) = 0$, then R is a subdirect product of weakly primitive rings. In other words, the weakly primitive ideals are dense in the Spec topology.

Algebras over uncountable fields, revisited

8. Suppose R is an algebra over an infinite field F, with $\dim_F R < |F|$ (as infinite cardinals). Prove that if R is weakly primitive (Exercise 7), then R is primitive. (Hint: One can replace r by αr for any $\alpha \neq 0$ in F, so one is done by Exercise 15.5 unless $1 - \alpha r$ is invertible for every α in F; use Proposition 15A.7.)

   Exercises 7 and 8 comprise an improved version of Exercise 15A.10.

Irreducible rings

9. A ring R is called **irreducible** if $A \cap B \neq 0$ for any nonzero ideals A, B of R. Show that any Noetherian ring is a finite subdirect product of irreducible rings. (Hint: The usual Noetherian induction argument: Assume that R is a counterexample to the assertion, but R/I is not a counterexample, for all $0 \neq I \lhd R$.)
10. If R is irreducible and satisfies ACC$\{\ell(r) : r \in R\}$, show that any central element r of R is either nilpotent or regular. (Hint: Apply Lemma 16.27 to the powers of r.)

Rings of fractions

11. When the Ore condition holds, show that multiplication in the ring of left fractions is given via $(s_1^{-1}r_1)(s_2^{-1}r_1) = (ss_2)^{-1}rr_1$, where $r_1 s_2^{-1}$ has been rewritten as $s^{-1}r$.

    In the following exercises, rings of fractions are built by means of two elementary axioms. An arbitrary submonoid S of R is called a **(left) denominator set** if it satisfies the following two conditions:

    (i) (Compare with the Ore condition.) For any $r \in R$ and $s \in S$, there are $r' \in R$ and $s' \in S$ such that $s'r = r's$;

    (ii) If $rs = 0$ for r in R and s in S, then $S \cap \ell(r) \neq \emptyset$.

    Replacing S by $S \cup \{1\}$, we assume that $1 \in S$; note that (ii) is vacuous if S is regular.
12. Suppose S is a denominator set for R. Show that there is an equivalence relation on $R \times S$, where we say that $(r_1, s_1) \sim (r_2, s_2)$ iff there exist $a, a' \in R$ such that

$$ar_1 = a'r_2 \quad \text{and} \quad as_1 = a's_2 \in S.$$

    Prove that the set $S^{-1}R$ of equivalence classes of $R \times S$ is endowed with a natural ring structure, and there is a natural ring homomorphism $R \to S^{-1}R$ sending r to $[(r, 1)]$. (Hint: Verify the following steps:

STEP I. Reflexivity and symmetry of the relation $\sim$ are obvious. Transitivity requires the following technical lemma:

STEP II. If $(r_1, s_1) \sim (r_2, s_2)$ and $a, a' \in R$ with $as_1 = a's_2 \in S$, then we can find $b \in R$ with $bar_1 = ba'r_2$ and $bas_1 = ba's_2 \in S$.

STEP III. Define addition as in Remark 16.15 and multiplication as in Exercise 11.)

13. Conversely to Exercise 12, show that S is a denominator set if R has a ring of left fractions $S^{-1}R$.

Goldie rings

14. Suppose R is a semiprime ring satisfying ACC (Annihilator ideals). Show that the minimal prime ideals of R are precisely the maximal annihilator ideals, and there are only finitely many of these. (Hint: A finite set of maximal annihilator ideals has intersection 0.)
15. (Goldie's Second Theorem.) Write ACC(Ann) for ACC on left annihilators. Write $ACC(\oplus)$ for ACC on direct sums of left ideals; i.e., there is no infinite direct sum

$$L_1 \subset L_1 \oplus L_2 \subset L_1 \oplus L_2 \oplus L_3 \subset \cdots .$$

A ring satisfying ACC(Ann) and ACC($\oplus$) is called a **left Goldie ring**. Prove that R has a semisimple left ring of fractions iff R is a semiprime left Goldie ring. (Hint: (G1) is as in Theorem 16.23. To see (G2), given $L <_e R$, pick $r_1 \in L$ with $\ell(r_1)$ maximal possible. Then $Rr_1 \cap \ell(r_1) = 0$, and pick $r_2 \in Rr_1$ with $\ell(r_2)$ maximal possible. This process must terminate, and then $\ell(\sum r_i) = 0$.)

16. (Goldie's First Theorem derived from Goldie's Second Theorem.) Show that $Q(R) \cong Q(R/P_1) \times \cdots \times Q(R/P_m)$ where $P_1, \ldots, P_m$ are the minimal prime ideals of the semiprime left Goldie ring R. Conclude that the ring of fractions of any prime Goldie ring is simple Artinian.

Embeddings

17. Prove that $ab = 1$ implies $ba = 1$ in a left Noetherian ring. (Hint: Pass first to $R/\operatorname{Jac}(R)$ and then apply Goldie's Theorem.)
18. (Mal′cev) A domain R not embeddible in a division ring. First note the impossibility of a matrix equation

$$\begin{pmatrix} 0 & * \\ * & * \end{pmatrix} \begin{pmatrix} * & * \\ * & 0 \end{pmatrix} = \begin{pmatrix} 0 & d \\ 0 & 0 \end{pmatrix} \tag{16E.1}$$

in $M_2(D)$, for any $0 \neq d \in D$. Hence

$$\begin{pmatrix} * & * \\ y & * \end{pmatrix} \begin{pmatrix} * & * \\ z & * \end{pmatrix} = \begin{pmatrix} 0 & d \\ 0 & 0 \end{pmatrix} \tag{16E.2}$$

has no solution for $y, z, d \neq 0$, since one can reduce to (16E.1) by multiplying on the left and right by suitable upper triangular matrices. But Equation (16.2) has a "generic solution" in $M_2(R)$, for $R = \mathbb{Q}\{x_1, \ldots, x_8\}/A$, where A is the ideal generated by

$$x_1x_5 + x_2x_7, \quad x_3x_5 + x_4x_7, \quad x_3x_6 + x_4x_8,$$

seen by considering the product $\left(\begin{smallmatrix} \bar{x}_1 & \bar{x}_2 \\ \bar{x}_3 & \bar{x}_4 \end{smallmatrix}\right) \left(\begin{smallmatrix} \bar{x}_5 & \bar{x}_6 \\ \bar{x}_7 & \bar{x}_8 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & * \\ 0 & 0 \end{smallmatrix}\right)$. Show that R is a domain.

Von Neumann regular rings

19. A ring R is called **von Neumann regular** if for every $a \in R$ there is $b \in R$ such that $aba = a$. Prove that the following assertions are equivalent: (i) R is von Neumann regular; (ii) every principal left ideal of R is generated by an idempotent; (iii) every f.g. left ideal of R is generated by an idempotent. (Hint for (ii) $\Rightarrow$ (iii): It suffices to show that any left ideal $L = Re_1 + Re_2$ is principal for any idempotents e_1 and e_2. Write $R(e_1 - e_1e_2) = Re$ and, noting that $ee_2 = 0$, show that $L = R(e - e_2)$.)
20. Prove that any left Noetherian, von Neumann regular ring is semisimple.
21. Show that $\operatorname{End}_R M$ is von Neumann regular for any semisimple module M over an arbitrary ring R. But, for an infinite-dimensional vector space M over a field, the von Neumann regular ring $\operatorname{End}_F M$ lacks IBN. (Compare with Exercise 13.3.)

Rings of quotients

22. Define the **singular ideal** of R to be comprised of those elements of R whose left annihilator is a large left ideal of R. Prove that the singular ideal is an ideal. (Hint: Use Lemma 16.19(ii).) Also, any left Goldie ring has singular ideal 0.
23. (The maximal ring of quotients.) Given any ring R, carry out the construction of $Q(R)$ as in the proof of Theorem 16.23. Show that $Q(R)$ is von Neumann regular, and the natural ring homomorphism $\nu: R \to Q(R)$ has kernel equal to the singular ideal of R.
24. (Martindale's extended centroid of a prime ring.) Given a prime ring R, let

$$\mathcal{S} = \{(L, f) \mid L \lhd R, \ f: L \to R \quad \text{is a map of } R\text{-bimodules}\}.$$

Define an equivalence relation on $\mathcal{S}$ by putting $(L_1, f_1) \sim (L_2, f_2)$, iff f_1 and f_2 are equal on a nonzero ideal contained in $L_1 \cap L_2$. Define the **extended centroid** $\hat{C} = \hat{C}(R)$ to be the set of equivalence classes,

endowed with addition and multiplication as in the proof of Theorem 16.23. Show that $\hat{C}$ is a field, and there is a natural injection $\mathrm{Cent}(R) \to \hat{C}$, which is an isomorphism if R is simple.

25. If R is a prime ring and $a, b \in R$ with $arb = bra$ for all $r \in R$, show that $a = cb$ for some c in the extended centroid. (Hint: The bimodule map $f\colon RaR \to R$ given by $f(a) = b$ is well-defined; this foreshadows the proof of Proposition 18.33.)

LE-modules and the Krull-Schmidt Theorem

26. For maps $f\colon M \to N$ and $g\colon N \to M$, show that the map $1_N + fg$ is an isomorphism of N iff $1_M + gf$ is an isomorphism of M. (Hint: $(\Rightarrow)$ $(1_M + gf)^{-1} = 1_M - g(1_N + fg)^{-1}f$.)
27. Suppose M is an LE-module. If $f\colon M \to N$ and $g\colon N \to M$ satisfy $fg \notin \mathrm{Jac}(\mathrm{End}_R N)$, show that f is split monic and g is split epic. (Hint: Take $h\colon N \to N$ such that $1_N - hfg$ is not an isomorphism. Then $1_M - ghf$ is not an isomorphism; hence, ghf is an isomorphism.)
28. (The exchange property.) Suppose $M = N \oplus N' = \bigoplus M_i$, where N is indecomposable and each M_i is an LE-module. Prove that, for some j, $N \cong M_j$ and

$$M = \left(\bigoplus_{j \neq i} M_j\right) \oplus N = M_j \oplus N'.$$

(Hint: Let $\pi\colon M \to N$ be the projection, and $\pi_i = \pi|_{M_i}$. Likewise, let $e_i\colon N \to M_i$ be the restriction of the projection $M \to M_i$, and let $f = \sum \pi_{i_u} e_{i_u}$. Then $(1 - f)(N) = 0$, so some $\pi_{i_u} e_{i_u} \notin \mathrm{Jac}(\mathrm{End}\, M_{i_u})$. Thus, π_{i_u} and e_{i_u} are isomorphisms, and apply Exercise 27.)

29. (Wedderburn-Krull-Schmidt-Azumaya-Beck) If M is a finite direct sum of LE-modules, prove that every other decomposition of M as a direct sum of indecomposables is the same, up to isomorphism and permutation of summands. In particular, this is true when M has finite composition length. (Hint: Induction on the number of summands, using the exchange property.)
30. Suppose the ring R is written in two ways as a direct sum of indecomposable left ideals; to wit,

$$R = Re_1 \oplus \cdots \oplus Re_t = Re'_1 \oplus \cdots \oplus Re'_{t'}.$$

Prove that $t' = t$, and there is some invertible element $u \in R$ and permutation $\pi \in S_t$ such that $e'_{\pi(i)} = ue_iu^{-1}$ for each $1 \leq i \leq t$. (Hint: Identify e_i with the projection $\hat{e}_i \in \mathrm{End}_R R$ sending $r \mapsto re_i$. By Exercise 29, there is $\varphi \in \mathrm{End}_R R$ such that $\varphi(e_i) = e'_i$ for $1 \leq i \leq t$. But then $\varphi^{-1}\hat{e}'_i\varphi = \hat{e}_i$ and $\mathrm{End}_R R \cong R^{\mathrm{op}}$.)

Graded rings and modules (cf. Definition 7.24 of Volume 1)

A left ideal L of a graded ring R is called a gr-**left ideal** if every element of L is the sum of homogeneous elements of L, i.e., L is a graded submodule of R. We say that a graded ring R is a gr-**domain** if the product of nonzero homogeneous elements is always nonzero. R is called a gr-**division ring** if every homogeneous element is invertible; i.e., if R has no nontrivial gr-left ideals. A graded module M is called gr-**simple** if M has no graded submodules other than 0 and M itself. Define gr-soc(M) to be the sum of all gr-simple submodules of M. M is called gr-**semisimple** if $M = \mathrm{gr\,soc}(M)$. We write gr-**left Artinian** for DCC(gr-left ideals), and gr-**left Noetherian** for ACC(gr-left ideals).

31. If R is a graded ring and M is a graded R-module, show that $\mathrm{Ann}_R M$ is a gr-left ideal of R. (Hint: Suppose $r = \sum r_i \in \mathrm{Ann}_R M$. For any homogeneous $a \in M$, note that each $r_i a = 0$.)
32. A homomorphism $f\colon M \to N$ of graded modules is said to be **graded of degree** t if $f(M_i) \in M_{i+t}$ for every i. Define $\mathrm{END}_R(M)$ as the ring of graded homomorphisms from M to itself; show that this is a graded ring. If M is gr-simple, then $\mathrm{END}_R(M)$ is a gr-division ring.
33. Call a graded submodule of a graded module M **gr-large** if it intersects every nonzero graded submodule of M nontrivially. Show that a graded module M is gr-semisimple iff every graded submodule has a graded complement, iff M has no proper gr-large graded submodules.
34. (Graded Wedderburn-Artin.) Show that any gr-left Artinian, gr-simple ring has the form $\mathrm{END}(M)_D$, where M is f.g. over a gr-division ring D.
35. If R satisfies ACC$(\ell(r) : r \in R$ homogeneous$)$, show that any nonzero nil gr-right ideal contains a homogeneous element $r \neq 0$ with $rRr = 0$. Thus, any nonzero nil right gr-ideal contains a nonzero nilpotent gr-ideal. (Hint: As in Lemma 16.25, noting that R is spanned by its homogeneous elements.)
36. (Graded First Goldie Theorem – Goodearl and Stafford.) If R is graded by an Abelian group $\mathcal{G}$ and is gr-prime and left gr-Goldie, show that R has a gr-simple left gr-Artinian graded ring of (left) fractions. (Hint: As in the nongraded case, once one shows that any gr-large gr-left ideal L contains a homogeneous regular element. The difficulty is that in the proof and notation analogous to Theorem 16.29, even if all the a_i are homogeneous, they might come from different components, so their sum would not be homogeneous. The trick is to take homogeneous $r_i \in R$ such that $a_1^2 r_1 a_2^2 r_2 \cdots r_{k-1} a_k^2 r_k$

is not nilpotent. This is possible since R is assumed gr-prime. Let

$$b_i = a_i^2 r_i a_{i+1}^2 r_{i+1} \cdots r_{k-1} a_k^2 r_k a_1^2 r_1 a_2^2 r_2 \cdots r_{i-1}.$$

Then $\ell(b_i) = \ell(a_i)$ for each i, implying that $b_1 + \cdots + b_k$ is a homogeneous regular element in L.)

37. Show that any left gr-Goldie gr-domain has a graded ring of (left) fractions that is a gr-division ring.
38. (Counterexample to the graded version of Goldie's Second Theorem.) Take $R = F[\lambda_0] \times F[\lambda_1]$, a reduced commutative Noetherian ring that is $\mathbb{N}$-graded, where λ_i has grade i and F has degree 0. Show that the 0-part is $F[\lambda_0] \times F$. The only regular homogeneous elements of R are in R_0, and inverting them yields $F(\lambda_0) \times F[\lambda_1]$, which is not left gr-Artinian.
39. (Bergman) Suppose R is $\mathbb{Z}/n$-graded and $r \in \mathrm{Jac}(R)$. If $r = \sum_{i=0}^{n-1} r_i$ is the decomposition into homogeneous components, show that $nr_i \in \mathrm{Jac}(R)$ for each i. (Hint: Let $R' = R[\zeta]$, for ζ a primitive n-th root of 1; R' is $\mathbb{Z}/n$-graded by the powers of ζ. But $r \in \mathrm{Jac}(R')$ by Exercise 15A.4, implying that $\sum_{i=0}^{n-1} \zeta^{ij} r_i \in \mathrm{Jac}(R')$, for each j. Hence,

$$nr_u = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \zeta^{ij} r_i \right) \in R \cap \mathrm{Jac}(R') \subseteq \mathrm{Jac}(R);$$

cf. Exercise 15A.5.)
40. (Bergman) Show that $\mathrm{Jac}(R)$ is a graded ideal of any $\mathbb{Z}$-graded ring R. (Hint: Apply Exercise 39 for n and $n+1$, with n suitably big.)
41. (Bergman) Reprove Amitsur's Theorem 15A.5 using Exercise 40 and grading $R[\lambda]$ according to degree in λ. (Hint: Write $J = \mathrm{Jac}(R[\lambda])$ and $N = R \cap J$. If $r \in N$, then $1 - r\lambda$ is already invertible in $R[\lambda]$, but its unique inverse in the larger ring $R[[\lambda]]$ of formal power series is $1 + r\lambda + r^2\lambda^2 + \cdots$; hence the series truncates, so r is nilpotent. Thus, N is a nil ideal, and must be 0. But if $\sum r_i \lambda_i \in J$, then each $r_i \in N = 0$.)

Appendix 16A

1. Write the special case of Equation (16A.3) for $m = 1$.
2. Show that the **quantum affine space**, defined via the relations $\lambda_j \lambda_i = q \lambda_i \lambda_j$ for all $1 \leq i < j \leq n$, is a deformation of $F[\lambda_1, \ldots, \lambda_n]$. One also gets further deformations by defining more generally $\lambda_j \lambda_i = q_{ij} \lambda_i \lambda_j$, where $q_{ji} = q_{ij}^{-1}$.

3. Construct the **quantum torus** as the deformation of the Laurent polynomials $F[\lambda_1, \lambda_1^{-1}, \ldots, \lambda_n, \lambda_n^{-1}]$ via the relations of Exercise A2.
4. Show that 16A.3 and Exercises A2 and A3 all are Noetherian domains. (Hint: Write the quantum affine planes and quantum matrix algebras as skew polynomial algebras.)
5. Show that the Grassmann algebra in two indeterminates can be deformed to the **quantum exterior algebra** having generators x_1, x_2 satisfying $x_i^2 = 0$ and $x_2 x_1 = -q x_1 x_2$. How can this be generalized to more indeterminates? Use this to define a quantum determinant of arbitrary size.
6. Generalize Example 16A.2 to $\mathcal{O}(F^{(n)})$.

Chapter 17

Monoid algebras

1. Given any monoid M and commutative ring C, define the **monoid algebra** $C[M]$ (also denoted CM in the literature) to be the free C-module with base M, endowed with multiplication given by

$$\left(\sum_{w \in M} c_w w \right) \left(\sum_{v \in M} c'_v v \right) = \sum_{u \in M} c''_u u, \tag{17E.1}$$

where $c''_u = \sum_{wv=u} c_w c'_v$. Using Exercise 13.1, show that $C[M]$ is an associative algebra; alternatively, verify this by means of the regular representation.

Show that monoid algebras satisfy the following universal property: For any C-algebra R, any monoid homomorphism $M \to (R, \cdot)$ extends to a unique algebra homomorphism $C[M] \to R$.

2. Display the polynomial algebra as a monoid algebra, with respect to the monoid $M = \{\lambda^i : i \in \mathbb{N}\}$. Display the free (associative) algebra as a monoid algebra.
3. Display any monomial algebra as a monoid algebra.

Free algebras and free groups

4. Show that the free group $\mathcal{G}$ is the universal from the free monoid $\mathcal{M}$ to the forgetful functor from groups to monoids. In other words, any monoid homomorphism from $\mathcal{M}$ to a group G extends uniquely to a group homomorphism from $\mathcal{G}$ to G.
5. Show that the free algebra (or free monoid or free group) on a given set is unique up to isomorphism (if it exists). (Hint: As in Proposition 8.6 of Volume 1.)

6. Show for $|X|$ countable that $C\{X\}$ is isomorphic to a subalgebra of the free algebra $C\{x_1, x_2\}$. (Hint: There are no relations among the elements $\{x_1^i x_2 : i \in \mathbb{N}\}$.)
7. Show that the free algebra $F\{x_1, x_2\}$ on two generators over a field F is primitive, and is isomorphic to the algebra of Exercise 15.10.
8. Verify the injection from the free group on $X = \{x_1, x_2, \dots\}$ to the free group $\mathcal{G}$ on two letters x_1, x_2 given by $x_i \mapsto x_1^i x_2^i$. (Compare with Exercise 6.)
9. Let K denote the field of rational fractions $F(\lambda_1, \lambda_2, \lambda_3)$. Verify the group injection φ from the free group $\mathcal{G}$ on $\{x_1, x_2\}$ to $\mathrm{GL}(2, K)$ given by

$$x_1 \mapsto \begin{pmatrix} 1 & 0 \\ \lambda_1 & \lambda_3 \end{pmatrix}, \qquad x_2 \mapsto \begin{pmatrix} 1 & \lambda_2 \\ 0 & \lambda_3 \end{pmatrix}.$$

(Hint: φ is clearly a homomorphism, so it suffices to show that φ is 1:1; in fact, if $1 \neq g \in \mathcal{G}$, then the entry in the $1,2$ or $2,1$ position of $\varphi(g)$ is nonzero.)

Power series rings over ordered monoids

10. Given any monoid M and a ring R, define R^M to be the direct product of copies of R, indexed by the elements of M; this is identified with the set of functions from M to R. Given $f \in R^M$, define $\mathrm{supp}(f)$ to be $\{s \in M : f(s) \neq 0\}$. Show that $R[M]$ is the R-submodule of elements of R^M having finite support (but is not a subring!).
11. If M is an ordered monoid, define the **power series ring** $R[[M]] = \{f \in R^M : \mathrm{supp}(f) \text{ is well-ordered}\}$. Write any element $f = (r_s)$ of $R[[M]]$ formally as a sum $\sum_{s \in M} r_s s$. The **lowest order** term of f is defined as $r_s s$ for that minimal s in $\mathrm{supp}(f)$. Show that $R[[M]]$ is a ring under the multiplication of (17E.1), and $R[M]$ is a subring.
12. Show that an element $f \in R[[M]]$ is invertible iff, taking rs to be the lowest order term in f, one has both r invertible in R and s invertible in M. (Hint: Write $r^{-1}s^{-1}f = 1 - h$, where each element of $\mathrm{supp}(h)$ is greater than 1. It is enough to compute the inverse of $1 - h$, which formally is $1 + h + h^2 + \cdots$. This argument is easier to verify in the important case when M is "archimedean," i.e., for each $u \in M$ there is k such that $s^k > u$ for each $s > 1$ in M.)
13. If the ordered monoid M is a group and D is a division ring, show that $D[[M]]$ is a division ring. (Hint: Apply Exercise 12.)

Combinatorics of the free group

In these exercises, $\mathcal{G}$ denotes the free group on $x_1, \dots, x_n$, and its lower central series is denoted $\gamma_1, \gamma_2, \gamma_3, \dots$.

14. Let $\mathcal{I}$ be the ideal of the free ring $\mathbb{Z}\{X\}$ generated by $\{x_i^2 : i \in I\}$. Let $R = \mathbb{Z}\{X\}/\mathcal{I}$, and write $\bar{x}_i$ for the canonical image of x_i in R. Thus, $\bar{x}_i^2 = 0$ for each i. Show that R is spanned by all words in the $\bar{x}_i$ in which no $\bar{x}_i$ appears twice consecutively.
15. For any ring R, write $\mathrm{Unit}(R)$ for the multiplicative group of invertible elements of R. Taking R as in Exercise 14, define $\varphi\colon \mathcal{G} \to \mathrm{Unit}(R)$ via $x_i \mapsto \bar{x}_i + 1$ and $x_i^{-1} \mapsto \bar{x}_i - 1$. Show that the highest degree term of $\varphi(x_1^{m_1} \cdots x_k^{m_k})$ is $\pm m_1 \cdots m_k \bar{x}_1 \cdots \bar{x}_k + \dots$, where $m_i \in \mathbb{Z}$. Conclude that φ is a group injection.
16. Notation as above, if $g \in \gamma_t$, show that the component of $\varphi(g)$ in R_ℓ is 0, for each $1 \leq \ell \leq t$. Thus, $\varphi(g) = 1+$ terms of degree $> t$. (Hint: Induction on t.)
17. (Magnus' Theorem.) Prove that $\bigcap_{t \in \mathbb{N}} \gamma_t = (1)$. (Hint: Apply Exercise 16.)
18. For any t, show that any element in the free group can be written in the form $c_u^{\pm 1} \cdots c_1^{\pm 1} g$, where $c_1 \leq \cdots \leq c_u$ are basic commutators and $g \in \gamma_t$. (Hint: Use Equation (17.5) repeatedly to rearrange products of basic commutators, each time creating new basic commutators of higher weight.)
19. Given a basic commutator c in $\mathcal{G}$, define the **basic ring commutator** $\hat{c}$ in $\mathbb{Z}\{X\}$, where the group commutator (a, b) is replaced at each stage by the ring commutator $[a, b] = ab - ba$; define $\bar{c}$ to be the natural image of $\hat{c}$ in $R = \mathbb{Z}\{X\}/\mathcal{I}$, i.e., with x_i replaced by $\bar{x}_i$, notation as in Exercise 14. For c of weight t, show that the component of $\bar{c}$ in R_ℓ is 0 for each $1 \leq \ell < t$. (Hint: Same idea as Exercise 16.)
20. Combining Exercises 15 and 19, show for any basic group commutators $c_1 < \cdots < c_u$ and $k_i \in \mathbb{Z}$ that the lowest-order nonconstant part of $\varphi(c_1^{k_1} \cdots c_u^{k_u})$ is $\sum_{i=1}^u k_i \bar{c}_i$.
21. Prove that the basic ring commutators comprise a $\mathbb{Q}$-base of the free $\mathbb{Q}$-algebra. (Hint: Use the reduction procedure $\hat{c}_j \hat{c}_i \mapsto \hat{c}_i \hat{c}_j + [\hat{c}_i, \hat{c}_j]$ for $i < j$. Inductively, show that the number of reduced terms in $x_1, \dots, x_n$ of fixed degree t is the same as the dimension of the homogeneous subspace of degree t in $\mathbb{Q}\{x_1, \dots, x_n\}$.)
22. Prove that γ_t/γ_{t+1} is a free f.g. Abelian group, for every t, where the images of the basic group commutators are the generators. (Hint: Use Exercise 18 to define a reduction procedure on $\mathcal{G}$ which is reduction-unique, since otherwise one would get a dependence on the $\bar{c}_i$, contrary to Exercise 21.)
23. (Magnus-Witt Theorem.) Prove that the free group $\mathcal{G}$ is ordered. (Hint: Apply Exercises 17 and 22 to Exercise 0B.7 of Volume 1.)
24. For any field F, show that $F[[\mathcal{G}]]$ is a division ring containing the free algebra $F\{X\}$. (Hint: Apply Exercise 23 to Exercise 13.)

Free resolutions

25. Find a f.g. free resolution of finite length of the $\mathbb{Z}[\lambda]$-module $M = \mathbb{Z}/2$ (where $\lambda M = 0$).
26. Suppose $R = \mathbb{Z}[\lambda]/\langle\lambda^2 - 1\rangle$ and $M = \mathbb{Z}[\lambda]/\langle\lambda - 1\rangle$. Using a parity argument, show that any f.g. free resolution of M must have infinite length.

Graphs

27. Given a connected, undirected graph Γ and any vertex v, make Γ directed such that v is the unique initial vertex.
28. (König graph theorem for undirected graphs.) Suppose Γ is an infinite connected, undirected graph. Show that Γ has an infinite path (not equivalent to a finite path) starting at any given vertex.
29. Show that an undirected graph has a circuit iff the number of edges is at least the number of vertices.
30. Show that the Cayley graph of a monomial algebra (with respect to the natural presentation) is the same as the Cayley graph of the underlying monoid of Exercise 3.

Affine algebras

31. (Generalization of the Artin-Tate Lemma 5.15 of Volume 1.) Prove that if R is an affine algebra f.g. over a commutative (not necessarily central) subalgebra C, then C is affine. (Hint: As in the proof of Lemma 5.15, R is f.g. and thus Noetherian as a module over an affine subalgebra C_0 of C, so C is also f.g. over C_0, and thus is affine.)

Growth of algebras

32. For any affine algebra R that is f.g. over a commutative subalgebra C, show that R has a rational Hilbert series with respect to a suitable generating set. (Hint: C is affine, by Exercise 31.)
33. A filtered affine F-algebra $R = F\{a_1, \dots, a_\ell\}$ is called **almost commutative** if its associated graded algebra is commutative. (See Exercise 21A.5 for an interesting example.) Show that any almost commutative affine algebra has a rational Hilbert series.
34. Prove Example 17.47(ii) by showing that $\bar{d}_k(R[\lambda_1, \dots, \lambda_n]) = \binom{n+k}{k}$.
35. When R is f.g. as a module over a subalgebra W, show that the growth rate of R equals the growth rate of W. (Hint: Reduce to Remark 17.44, by means of Exercise 13.20.)
36. Reprove Lemma 17.62 using Exercise 35.
37. Prove that $\mathrm{GKdim}(R/I) \leq \mathrm{GKdim}(R) - 1$ for any $I \lhd R$ containing a regular element of R. In particular, this situation holds whenever I is a nonzero ideal of a prime Goldie ring R.

38. What is $\mathrm{GKdim}(R)$, where R is the monomial algebra $F\{x_1, x_2\}/\langle x_2 x_1^i x_2 x_1^j x_2,\ x_2 x_1^k x_2 : i, j, k \in \mathbb{N}\rangle$?
39. (Hyperword algebras.) Any word or hyperword h gives rise to a monomial algebra whose monomials are nonzero iff they are subwords of h. Given a growth rate $\geq p_2$, construct a hyperword algebra with that growth rate. (Hint: The same idea as in Exercise 38.)
40. (Smoktunowicz-Vishne) Define a **bidirectional** hyperword as extending indefinitely both to the left and right. Construct a bidirectional hyperword h in the free monoid on x and y by taking $h_1 = x$, $h_{n+1} = h_n y^{3^n} h_n$ for each n, and take the limit. Let R be the monomial algebra whose monomials are nonzero iff they are subwords of h. Show that the ring R is prime, but x generates a nonzero nil (in fact, locally nilpotent) ideal. In particular, the upper nilradical of R is nonzero, so R is not primitive. Also, R has quadratic growth.

Growth of arbitrary algebraic structures

41. We say that an algebraic structure A with a given signature is **generated** by a set S if it has no proper substructure (of the same signature) containing S. A is **finitely generated** if A is generated by some finite set. Show that this definition is consistent with Exercise 1A.28 of Volume 1 and generalizes "finitely generated" for groups and for modules.
42. Suppose A is a finitely generated algebraic structure whose signature contains a binary operation (called multiplication, for convenience). By considering the same signature excluding multiplication, show how to study the growth of A in terms of other structures containing its generating set.

Growth of groups

43. Suppose G is a nilpotent group of class t having center Z, and S is a finite set that generates G. Show that for any $z \in Z$, there is $\ell = \ell(z)$ such that, for all $k \in \mathbb{N}$, any power of z up to z^{k^t} can be written as a word of length at most ℓk in $S \cup S^{-1}$. (Hint: Induction on t.)
44. Prove under the hypotheses of Theorem 17.60 that N has polynomial growth of degree $\sum_j j d_j$. (Hint: Theorem 17.60 for the upper bound and Exercise 43 for the lower bound.)

Appendix 17A

1. Show that the symmetric group S_n has the **Coxeter presentation** $\sigma_i^2 = 1$, $(\sigma_i \sigma_{i+1})^3 = 1$, and $(\sigma_i \sigma_j)^2 = 1$ for $|j - i| > 1$. It is more efficient to use those relations with $j > i$. (Hint: Let $\widehat{S}_n$ denote the

group defined by the Coxeter presentation. It suffices to show that the homomorphism $\widehat{S}_n \to S_n$ given by $\sigma_i \mapsto (i\ i{+}1)$ has trivial kernel. The Coxeter relations permit us to replace $\sigma_{i+1}\sigma_i\sigma_{i+1}$ by $\sigma_i\sigma_{i+1}\sigma_i$, and to interchange σ_i and σ_j for $|j-i|>1$. Thus, one can always change the indices of a nontrivial element of $\widehat{S}_n$ so that the highest index occurs only once; it follows that the image of this element is nontrivial in S_n, as desired.)

Fundamental groups

2. Show that the smallest nontrivial circuit must have three vertices and three edges. Thus, the smallest bouquet of n circuits has $2n+1$ vertices and $3n$ edges. (Hint: One vertex is in common.)
3. Show that any graph has a covering complex that is a tree. (Hint: At each vertex, formally take the same number of edges leaving from that vertex.)
4. For any maximal tree T of a connected graph $\mathcal{K}$, show that all the vertices of $\mathcal{K}$ belong to T. (Hint: Otherwise enlarge T by adjoining a shortest path from an extra vertex of $\mathcal{K}$ to T.)
5. If $\mathcal{K}$ is a finite graph with n_1 edges and n_0 vertices, show that $\pi(\mathcal{K}, v)$ is free of rank $n_1 - n_0 + 1$. (Hint: Apply Exercise A4 to Example 17A.5.)
6. Suppose $G = \pi(\mathcal{K}, v)$ and $H < G$. Define two paths $p(v,b)$ and $q(v,b)$ to be H-**equivalent** if the homotopy class of the circuit $p\bar{q}$ belongs to H. Given a simplex $\mathcal{S}$ of $\mathcal{K}$, define $\mathcal{S}_H$ to be the set of H-equivalent homotopy classes obtained by taking a path that terminates in S followed by a path within S. Show that $H = \pi(\mathcal{K}_H, v)$, where $\mathcal{K}_H$ is defined as the union of the simplices of the form $\mathcal{S}_H$, for S a simplex of $\mathcal{K}$.

   Define $\phi_H\colon \mathcal{K}_H \to \mathcal{K}$ by sending the equivalence class of a path to its terminal vertex, thereby displaying $\mathcal{K}_H$ as a covering complex of $\mathcal{K}$ having fundamental group H. (This exercise requires many fairly straightforward verifications.)
7. If $\mathcal{G}$ is a free group of rank n and H is a subgroup of index m, prove that H is free of rank $mn - m + 1$. (Hint: By Exercise A2, $\mathcal{G}$ is the fundamental group $\pi(K, v)$ of a bouquet $\mathcal{K}$ having $2n+1$ vertices and $3n$ edges. By Exercise A6, H is the fundamental group of some complex $\mathcal{K}_H$, where $m = [\mathcal{G} : H]$ is the number of vertices lying over v. Using Exercise A5, show that H is free of rank $3mn - m(2n+1) + 1 = mn - m + 1$.)
8. Reprove Remark 00.1. (Hint: Write G as the image of a free group and apply Exercise A7.)

9. Prove that any group G is the fundamental group of a complex $\mathcal{K}$ of dimension 2. G is finitely presented iff $\mathcal{K}$ can be taken finite. (Hint: Suppose G is generated by $\{x_i : i \in I\}$ with relations $\mathcal{R}$. Let $\mathcal{K}$ be the bouquet having fundamental group $\mathcal{G}$ of rank $|I|$; its 2-simplices are $\{\{v, w_i, u_i\} : i \in I\}$. Taking relations ρ_j generating $\mathcal{R}$ as a normal subgroup, write the path corresponding to ρ_j as $e_{j_1} \cdots e_{j_\ell}$ for suitable edges e_{j_k} in $\mathcal{K}$. Then take some new letter z_j and adjoin all 2-simplices $\{z_j\} \cup e_{j_k}$ to obtain a complex $\mathcal{K}_1$. The inclusion map $\phi\colon \mathcal{K} \hookrightarrow \mathcal{K}_1$ induces a surjection $\phi_\#\colon \pi(K, v) \to \pi(K_1, v)$ whose kernel clearly contains $\mathcal{R}$. Show that $\ker \phi_\# = \mathcal{R}$. The second assertion follows from this proof.)

Stallings foldings, following Kapovich-Myasnikov [KapoM]

Notation as in Definition 17.23ff, consider graphs having the property that if $e \in E$ then $\bar{e} \in E$; cf. Remark 17.23′. A graph Γ is said to be **labelled** if there is a **labelling function** $\mathcal{L}$ from the set of edges to an alphabet $X \cup X^{-1}$; in other words, every edge e is labelled by x_i or x_i^{-1}. If e is labelled by x_i, then $\bar{e}$ is labelled by x_i^{-1} and vice versa, so by reversing directions we could use labels only from X. This yields a category that refines Remark 17.23′. Namely, our labelled graph is written $\Gamma = (V, E, \mathcal{L})$; a morphism $f\colon \Gamma_1 \to \Gamma_2$ now is required to satisfy the property that $\mathcal{L}(f(e)) = (e)$ for each $e \in E$.

10. Given a labelled graph $\Gamma = (V, E, \mathcal{L})$, define the **folding procedure** as the following equivalence relation:

    Two edges $e = (v, w)$ and $e' = (v, w')$ with the same initial vertex v are **equivalent** iff $\mathcal{L}(e) = \mathcal{L}(e')$, in which case we also say that $w \sim w'$. Define a new labelled graph, called the **initially folded graph** of Γ, consisting of vertices and equivalence classes of edges, with $\mathcal{L}([e])$ defined as $\mathcal{L}(e)$. In order to preserve symmetry, one does the same procedure with terminal vertices to obtain the **folded graph** $\hat{\Gamma}$. Show that $\hat{\Gamma}$ is a labelled graph, and that there is a natural morphism $\Gamma \to \hat{\Gamma}$.
11. For any folded graph, show that the labelling function $\mathcal{L}$ is locally 1:1 in the sense that two distinct edges with the same initial vertex (or the same terminal vertex) must be labelled differently.
12. Show that one can pass from a graph to its folded graph via a series of morphisms, each of which only identifies two vertices and two edges. This is called a **folding**.
13. Show that all paths in a folded graph are **reduced** in the sense that they do not contain an edge labelled by x_i followed by an edge labelled by x_i^{-1}, or vice versa.

14. Fix a distinguished vertex v_0 and define the **core graph** to be the union of the reduced paths from v_0 to v_0. Show that the set of reduced paths from v_0 to itself has a group structure, obtained by juxtaposing and then reducing. Any core graph gives rise to a subgroup of the free group. Conclude that there is a 1:1 correspondence between subgroups of the free groups and core graphs with a distinguished vertex, and the combinatorics of graph theory are available for studying presentations of groups as homomorphic images of the free group.

Appendix 17B

1. A **Gröbner-Shirshov(-Cohn-Bergman basis)** for an algebra $F\{X\}/N$ is a set $S \subseteq N$ that generates N as an ideal, such that $\widehat{S}$ spans $\widehat{N}$; i.e., the leading term of every element in N is spanned by the leading terms of elements of S. Using Bergman's theory, show that any set of relations can be expanded to a Gröbner-Shirshov basis (but the procedure is not necessarily recursive).
2. If $N \lhd F\{X\}$ has a set of relations yielding a reduction-unique procedure on $F\{X\}$, show that the algebra $R = F\{X\}/N$ is isomorphic as a vector space to the space V of irreducible elements of $F\{X\}$. In this case, V can be made into an algebra isomorphic to R by defining the product of a and b to be the reduction of ab in $F\{X\}$.
3. Carry over the theory of Gröbner bases of commutative algebras to almost commutative algebras (cf. Exercise 17.33).

Hyperbolic groups

4. Suppose that the hyperbolic group G is generated by the set $S = \{g_1, \dots, g_\ell\}$. Given a function $f\colon S \to \mathbb{R}^+$, show that the new distance function defined by $d(a, ag_i) = f(g_i)$ yields the same definition of "hyperbolic group."

Appendix 17C

1. Verify that $G = B(m, 3)$ is finite, for any number m of generators. (Hint: Induction on m. Suppose G is generated by $\{g_1, \dots, g_m\}$, and let H be the subgroup generated by $\{g_1, \dots, g_{m-1}\}$, which by hypothesis has some bounded order k. Let $g = g_m$. Then every element of G has the form

$$h_1 g h_2 g \cdots g h_j, \quad h_i \in H \cup gHg^{-1}. \tag{E17C.1}$$

But $(gh)^3 = 1$ implies $ghg = h^{-1}g^{-1}h^{-1}$, which can be applied to (E17C.1) to reduce j repeatedly; thereby, assume that $j \le 3$. Hence $|G| \le (2k)^3$; this bound can be reduced with a little care.)

2. (Sanov) If G is a group of exponent 4 generated by a finite subgroup H and an element g with $g^2 \in H$, show that G is finite. (Hint: Any element of G has the form

$$h_1 g h_2 g \cdots g h_j, \quad h_i \in H. \tag{E17C.2}$$

$(gh)^4 = 1$ implies $ghg = h^{-1}gh'gh^{-1}$, where $h' = g^2h^{-1}g^2 \in H$. Applying this to gh_ig in (E17C.2) yields a new string of the same length, but where h_{i-1} has been replaced by $h_{i-1}h_i^{-1}$. Looking at different positions enables one to run over all the values of H, if $j > |H| + 1$, and thus reduce some h_i to 1. Hence $|G| \le |H|^{|H|+1}$.)
3. Verify that $B(m, 4)$ is finite for any m by applying Exercise C2 twice.
4. For any $\bar{x}_1, \bar{x}_2$ in any group G of prime exponent p, show that the higher group commutator

$$\bar{w} = ((\dots((\bar{x}_2, \bar{x}_1), \bar{x}_1), \dots), \bar{x}_1),$$

where commutation by $\bar{x}_1$ is done $p-1$ times, lies in $\gamma_{p+1}(G)$. (Hint: Let $\gamma = \gamma_{p+1}(G)$. In the free group, one must check that w is a product of p-powers times an element of γ. In the Hall collection process applied to $(x_1x_2)^p$, any basic commutator of weight $< p$ occurs with power a multiple of p; Example 17C.4 is a special case. Ignoring p powers, write

$$w = v_1 \cdots v_t \pmod{\gamma}, \tag{E17C.3}$$

for other higher commutators v_i of weight p whose degrees in x_1 are between 1 and $p-2$. But the higher commutators of weight p commute modulo γ. Replacing x_1 by x_1^j for $p \nmid j$, one still gets w since $j^{p-1} \equiv 1 \pmod{p}$, but any $v = v_i$ is replaced by v^{t^d}, where d is the number of times x_1 appears in v. Taking each value $1 \le j \le p-1$ in turn and multiplying these $p-1$ versions of (E17C.3), one gets w^{p-1} on the left side, but on the right side, v appears to the power

$$1^j + 2^j + \cdots + (p-1)^j \equiv 1 + 2 + \cdots + (p-1) \equiv 0 \pmod{p},$$

as desired.)

Grigorchuk's Example

The following exercises provide Grigorchuk's example of an infinite, periodic, finitely generated group; the example is presented as a set of transformations of an infinite binary tree with a single initial vertex. (A directed tree is called **binary** if each vertex has two edges leading from it.) We consider the following binary tree Γ:

The vertices of Γ are labelled $\{v_i : i \in \mathbb{N}^+\}$ with v_1 the initial vertex. Any vertex v_i is the initial vertex of the two edges (v_i, v_{2i}) and (v_i, v_{2i+1}) of Γ. A connected sub-tree of Γ is called a **branch**. Clearly, any edge v_i is the initial vertex of a branch Γ_i isomorphic to Γ; note that $\Gamma = \Gamma_1$.

Let T_i denote the automorphism of Γ (of order 2) interchanging the subtrees Γ_{2i+1} and Γ_{2i+2}. For example, T_1 switches the left and right branches; thus $T_1(v_2) = v_3$, $T_1(v_3) = v_2$, $T_1(v_4) = v_6$, $T_1(v_5) = v_7$, $T_1(v_6) = v_4$, $T_1(v_7) = v_5$, and in general

$$T_1(v_{2^{m+1}+k}) = v_{3\cdot 2^m+k}, \quad 0 \le k < 2^m.$$

Grigorchuk's group G is generated by three transformations a, b, and c, where $a = T_1$ whereas b and c are defined by infinite sequences of transformations:

$$b = (T_3, T_5, T_{17}, T_{33}, \dots, T_{2^{3i+1}+1}, T_{2^{3i+2}+1}, \dots);$$

$$c = (T_3, T_9, T_{17}, T_{65}, \dots, T_{2^{3i+1}+1}, T_{2^{3i+3}+1}, \dots).$$

It is convenient to define

$$d = bc = (T_5, T_9, T_{33}, T_{65}, \dots, T_{2^{3i+2}+1}, T_{2^{3i+3}+1}, \dots).$$

(Grigorchuk formulated his original example in terms of transformations of the unit interval; the outcome is the same.)

5. Show that the subgroup of G generated by b and c is Abelian; also, $a^2 = b^2 = c^2 = 1$.
6. Show that any element g of G can be written as an alternating product from $\{a\}$ and $\{b, c, d\}$.)
7. Let H be the subgroup of G generated by words in which a appears an even number of times. (Thus, the even numbers and the odd numbers are invariant subsets.) Show that H is generated by b, c, aba, and aca, and is isomorphic to the subgroup of $G \times G$ generated by (c, a), (d, a), $(1, ac)$, and $(1, ad)$. Hence, projecting onto the second component shows that G is a homomorphic image of H. Conclude that G is infinite. (Hint: $\Gamma \setminus \{v_1\}$ is the disjoint union of Γ_2 and Γ_3.)
8. Prove that every element of G has order a power of 2.

Chapter 18

1. How many simple tensors are there in $F^{(2)} \otimes_F F^{(2)}$, for $F = \mathbb{Z}/p$?
2. Prove that $(\bigoplus_{i\in I} M_i) \otimes N \cong \bigoplus_{i\in I}(M_i \otimes N)$. (Hint: Proof of Proposition 18.11.)

3. Suppose M is an R-module, and $f\colon N \to P$ is an onto map of right R-modules. Prove that $\ker(f \otimes 1_M) = \ker f \otimes M$, viewed as a submodule of $N \otimes_R M$, yielding

$$(N \otimes M)/(\ker f \otimes M) \cong P \otimes M.$$

Put another way, if $K \to N \to P \to 0$ is exact, then so is

$$K \otimes M \to N \otimes M \to P \otimes M \to 0.$$

(Hint: Write $\hat{K} = \ker f \otimes M$. Then $\hat{K} \subseteq \ker(f \otimes 1_M)$, so $f \otimes 1_M$ induces a homomorphism $\bar{f}\colon (N \otimes M)/\hat{K} \to P \otimes M$ whose inverse comes from the balanced map $(f(a), b) \mapsto (a \otimes b) + \hat{K}$.)

4. If $I \lhd R$ and M is an R-module, show that $M/IM \cong (R/I) \otimes_R M$ as R/I-modules. (Hint: In Exercise 3, take $N = R$ and $P = R/I$.)
5. Under the identification of Example 18.14, show that $\mathrm{tr}(v \otimes f) = f(v)$. (Hint: $\mathrm{tr}(e_i \otimes e_j^*) = \delta_{ij} = \mathrm{tr}(e_{ij})$.)

Tensor products of algebras

6. Show that $M_n(R) \cong M_n(C) \otimes_C R$ for any C-algebra R.
7. Show that $\ker(f_1 \otimes f_2) = \ker f_1 \otimes R_2 + R_1 \otimes \ker f_2$ for any algebra homomorphisms $f_i\colon R_i \to T_i$. (Hint: $(\supseteq)$ is clear. Thus, letting $A_j = \ker f_j$ and $I = A_1 \otimes R_2 + R_1 \otimes A_2$, one has a natural algebra homomorphism $\varphi\colon (R_1 \otimes R_2)/I \to R_1/A_1 \otimes R_2/A_2$. In the other direction, the balanced map

$$R_1/A_1 \times R_2/A_2 \to (R_1 \otimes R_2)/I$$

given by $(r_1 + A_1, r_2 + A_2) \mapsto r_1 \otimes r_2 + I$, yields the inverse of φ.)

8. Suppose R is an algebra with base $\{b_i : i \in I\}$ over a field K. Write $b_i b_j = \sum_k \alpha_{ijk} b_k$ for $\alpha_{ijk} \in K$. Suppose F is a subfield of K containing all the structure constants α_{ijk}; cf. Remark 5.15′ of Volume 1. Then show that $R_0 = \sum_i F b_i$ is an F-algebra, and $R \cong R_0 \otimes_F K$.

Clifford algebras

9. Suppose R is algebraic of degree 2 over the field F. Define $Q\colon R \to F$ as follows: $Q(\alpha) = \alpha^2$ for $\alpha \in F$; if $a \in R \setminus F$ satisfies a polynomial $a^2 + \alpha_1 a + \alpha_2 = 0$ for $\alpha_i \in F$, then $Q(a) = -\alpha_2$. Show that Q is a quadratic form, and we thus have a functor from algebraic algebras of degree 2 to quadratic forms. The Clifford algebra construction $C(V, Q)$ is the universal from the quadratic form Q to this functor.
10. Show that $C(V, Q)$ is $\mathbb{Z}/2$-graded. (Hint: The ideal $\mathcal{I}$ of relations is generated by elements of even grade.)
11. In the Clifford algebra $C(V, Q)$, show that

$$v_1 v_2 = -v_2 v_1 + Q(v_1 + v_2) - Q(v_1) - Q(v_2)$$

for any v_1, v_2. Ordering the index set I and taking a base $\{e_i : i \in I\}$ of V, show (either using the regular representation or the word reduction procedure) that $\{1\} \cup \{e_{i_1} \cdots e_{i_m} : i_1 < i_2 < \cdots < i_m\}$ is a base for $C(V,Q)$. In particular, this holds for the Grassmann algebra $E(V)$.

12. Show that $C(V,Q)$ has an involution $(*)$ satisfying $v^* = v$, $\forall v \in V$. If the bilinear form of Q is nondegenerate, then $(C(V,Q), *)$ is simple as an algebra with involution. (In fact, it is isomorphic to a tensor product of generalized quaternion F-algebras, perhaps also with with a quadratic field extension of F.)

Tensor products and field extensions

13. Suppose K and L are fields containing F. Then, for any maximal ideal P of $K \otimes_F L$, show that $(K \otimes_F L)/P$ is a field containing K and F, called the **compositum** of K and F. Conversely, any such field can be obtained in this way.
14. Prove that the tensor product of f.d. separable field extensions is semisimple.
15. For any purely inseparable field extension R of F and any field $K \supseteq F$, show that $R \otimes_F K$ is not split semisimple. (Hint: Use Example 18.29.)
16. For any separable field extension K of F, show that $K \otimes_F K$ has a simple idempotent e with $(a \otimes b)\, e = (b \otimes a)\, e$ for all $a, b \in K$. (Hint: $K \otimes K$ is commutative semisimple, so the kernel of the multiplication map $\mu\colon K \otimes K \to K$ is generated by an idempotent e_0; take $e = 1 - e_0$, noting that $\mu(e) = 1$.)
17. For K/F Galois with Galois group G, show that $(\sigma \otimes \sigma)(e) = e$ for each $\sigma \in G$, with e as in Exercise 16.

Wedderburn's Principal Theorem over perfect fields

18. (Wedderburn's Principal Theorem.) For any a finite-dimensional algebra R over a perfect field F, prove that $R = S \oplus J$ as a vector space over F for a suitable semisimple subalgebra $S \cong R/J$ of R, where J is the Jacobson radical of R. (Hint: The algebraic closure of F is separable over F and splits R/J. Hence, some f.g. extension field K splits R/J, so apply Theorem 15.26.)

Applications to algebraic geometry

19. Prove that the tensor product of two reduced algebras over an algebraically closed field F is reduced. (Hint: As in Proposition 18.41.)
20. Prove Proposition 18.42 for algebraic sets instead of varieties. (Hint: Use Exercise 19.)

21. If V is an affine variety $R = F[V]$ and $J = \ker \mu\colon R \otimes R \to R$, show that $\mathcal{Z}(J)$ is the diagonal of V in $V \times V$, identifying $F[V] \otimes F[V]$ with $F[V \times V]$.

Tensor products of graded algebras

22. If A, B are F-algebras graded over a monoid $(\mathcal{G}, +)$, show that $A \otimes B$ is graded, where $(A \otimes B)_k = \bigoplus_{g+h=k} A_g \otimes B_h$. In particular, if A, B are superalgebras, then so is $A \otimes B$, with $(A \otimes B)_0 = A_0 \otimes B_0 \oplus A_1 \otimes B_1$ and $(A \otimes B)_1 = A_0 \otimes B_1 \oplus A_1 \otimes B_0$.

The Jacobson radical and tensor products

23. (Amitsur) If F is a field and R is an F-algebra without nonzero nil ideals, prove that $\mathrm{Jac}(R \otimes_F F(\lambda)) = 0$. (Hint, due to Bergman: First show that $L = F(\lambda)$ can be $\mathbb{Z}/n$-graded by $L_i = \lambda^i F(\lambda^n)$. Using an argument analogous to that of Exercise 16.41, one still can show that if $1 - r\lambda$ is invertible, then r is nilpotent, as follows: Compute its inverse inside the formal power series ring $R[[\lambda]]$, cf. Exercise 17.12, and show that $F[r]$ is finite-dimensional, which implies $\mathrm{Jac}(F[r])$ is nilpotent.)
24. Prove that $K \otimes_F \mathrm{Jac}(R) \subseteq \mathrm{Jac}(K \otimes_F R)$ whenever $K \supseteq F$ are fields and R is an algebra over F, equality holding if K/F is separable. (Hint: Let $J = \mathrm{Jac}(R)$. First, one needs to show that any element a of $K \otimes J$ is quasi-invertible. But $a \in L \otimes J$ for some subfield L of K finitely generated over F, so use Exercise 15A.4.

    For the last assertion, pass to R/J and assume that $J = 0$; now, passing to the Galois closure of K, assume that K/F is Galois. Let $R_1 = K \otimes_F R$. Then $G = \mathrm{Gal}(K/F) \otimes 1$ acts on R_1, with fixed subring $F \otimes_F R \cong R$. Take a base $a_1, \ldots, a_n$ of K over F and suppose $r = \sum_i a_i \otimes r_i \in \mathrm{Jac}(K \otimes R)$. Then for any a_j,

$$\sum_i \mathrm{tr}_{K/F}(a_i a_j) \otimes r_i = \sum_{\sigma \in G} \sigma(r(a_j \otimes 1)) \in R \cap \mathrm{Jac}(K \otimes R) = J = 0.$$

    Conclude that each $r_i = 0$, using Cramer's rule.)

Part V

Representations of Groups and Lie Algebras

Introduction to Representations of Groups and Lie Algebras

Noncommutative algebra often involves the investigation of a mathematical structure in terms of matrices. Having described a finite-dimensional representation of an algebra R as an algebra homomorphism $\hat{\rho}: R \to M_n(F)$, for suitable n, we would like to utilize tools from matrix theory (such as the trace) to study the original algebra.

The point of Part V is that the same ideas hold for other algebraic structures. In Chapter 19, we establish the basic framework for studying group representations, which in Chapter 20 leads us to **character theory**, the study of representations through traces. One of the fundamental results is Maschke's Theorem, which implies that every linear representation is a direct sum of irreducible representations. The same idea of proof enables us also to prove the parallel result for compact topological groups, leading us to the study of Lie groups (Appendix 19A) and algebraic groups (Appendix 19B).

Both of these structures can be explored using a derivative structure, that of Lie algebra, which is treated together with its own representation theory in Chapter 21. The corresponding associative structure, the enveloping algebra, is discussed briefly in Appendix 21A. (Later, in Appendix 23B,

we consider an interplay between groups and Lie algebras, which leads to a sketch of Zelmanov's solution of the Restricted Burnside Problem.)

Since multiplication in Lie algebras is not associative, we take the occasion in Appendix 21B to study other nonassociative structures such as alternative algebras and Jordan algebras.

Part of the Lie algebra theory (namely, the theory of Dynkin diagrams) spills over into Chapter 22, because of its connections to other structures. In Chapter 22, we also study Coxeter groups, which are certain groups of reflections whose study is motivated by the Weyl group of a semisimple Lie algebra.

One idea that pervades the various approaches is the correspondence of Remark 16.37, which transforms representation theory into the theory of modules; it lurks beneath the surface throughout Chapter 19 (especially in Proposition 19.12) and Chapter 21 and finally breaks through in Appendix 25C.

Chapter 19

Group Representations and Group Algebras

In this chapter we define group representations over a field F, and study their basic properties in terms of the structure of the group algebra, featuring Maschke's Theorem. G always denotes a group, with the operation written as multiplication, and the identity element written as 1. We focus on finite groups, with emphasis at the end on the symmetric group S_n, leaving representations of infinite groups for Appendices 19A and B.

Group representations

$\mathrm{GL}_F(V)$ denotes the group of 1:1 linear transformations of a vector space V over a field F. We write $\mathrm{GL}(V)$ when F is understood. (Thus, $\mathrm{GL}(V) = \mathrm{GL}(n, F)$ when $V = F^{(n)}$.)

Definition 19.1. A **representation** of a group G is a group homomorphism $\rho : G \to \mathrm{GL}(V)$ for a suitable vector space V over the field F. In the next two chapters we consider only the case when $\dim_F V = n < \infty$; then we say that the representation ρ is **finite-dimensional** (f.d.) of **degree** n. The representation ρ is called **complex** if $F = \mathbb{C}$; ρ is called **real** if $F = \mathbb{R}$. (More generally, one could study homomorphisms to the group of units of $\mathrm{End}_C M$, for any f.g. module M over a commutative ring C; the case when $C = \mathbb{Z}$ is of particular interest in the literature.)

Representations of groups are among the most important tools in mathematics. They enable us to study elements of the groups in terms of matrix techniques, viewing $\mathrm{GL}(n, F) \subset M_n(F)$. The case when the base field F is algebraically closed is especially important, since F then contains the eigenvalues of the matrices, and the canonical forms from Chapter 2 of Volume 1 become available. Accordingly, we concentrate on the case when $F = \mathbb{C}$ although, as we shall see, sometimes it is easier to distinguish the underlying group when F is not algebraically closed.

The kernel of the group representation ρ, denoted $\ker \rho$, measures the amount of information lost in passing to $\mathrm{GL}(n, F)$. A representation with kernel $\{1\}$ is called **faithful**. Here is the main example.

Example 19.2. (The regular representation.) The symmetric group S_n has a faithful representation sending the permutation π to the permutation matrix $\sum e_{\pi(i),i} \in GL(n, F)$. Combined with Cayley's Theorem, which injects any group G of order n into S_n, this shows that any group G of order n has a faithful representation ρ into $\mathrm{GL}(n, F)$, called the **regular representation** ρ_{reg}, sending an element $g \in G$ to the permutation matrix corresponding to the left multiplication map ℓ_g.

Note that $gg_i = g_i$ iff $g = 1$. Thus, all the diagonal entries of the permutation matrix $\rho_{\mathrm{reg}}(g)$ are 0 unless $g = 1$, and $\rho(1) = I$.

Since ρ_{reg} is faithful, one should expect it to contain all the information about G, and indeed ρ_{reg} is a key tool in representation theory, to be referred to repeatedly. (Following Remark 19.16, we tie this in with the regular representation of an algebra.)

More generally, a **permutation representation** is a representation ρ for which $\rho(g)$ is a permutation matrix, $\forall g \in G$. Explicitly, for the regular representation, we write $G = \{g_1 = 1, g_2, \ldots, g_n\}$ and send g to the permutation matrix having 1 in the i, j position precisely when $gg_i = g_j$.

Degree 1 representations.

The **degree 1 representations** are just the group homomorphisms $G \to \mathrm{GL}(1, F) = F^\times$, and they play a special role in the theory. Sometimes these representations are called "linear," but we defer this terminology since itconflicts with our usage above and with "linear group" of Appendix 19B. The **unit representation**, or **trivial representation**, denoted as $\mathbf{1}$, is the degree 1 representation given by $\rho(g) = 1$ for all g in G; it exists for any group G and any field F. Note that $\ker \mathbf{1} = G$. All other representations are called **nontrivial**.

We can already determine all degree 1 representations of finite Abelian groups.

LEMMA 19.3.

(i) *The number of degree* 1 *representations of the cyclic group* C_n *of order* n *equals the number of distinct* n*-th roots of* 1 *contained in* F*, which is* n *precisely when* F *contains a primitive* n*-th root of* 1.

(ii) *Suppose* $A = C_1 \times \cdots \times C_m$ *is a finite Abelian group, written as a direct product of cyclic groups* C_i *of order* n_i*. If* F *contains primitive* n_i*-roots of* 1 *for* $1 \le i \le m$*, then* A *has precisely* $n_1 \cdots n_m = |A|$ *distinct degree* 1 *representations.*

Proof. (i) Suppose ρ is a degree 1 representation of $C_n = \langle a \rangle$. Since $\rho(a)^n = \rho(a^n) = 1$, $\rho(a)$ must be one of the n-th roots of 1 in F. Conversely, if ζ is an n-th root of 1 in F, then there is a degree 1 representation $\rho\colon G \to F^\times$ given by $\rho(a^j) = \zeta^j$.

(ii) Given representations $\rho_i\colon C_i \to F^\times$ of degree 1, we can define $\rho\colon A \to F^\times$ by

$$\rho(c_1, \dots, c_m) = \rho_1(c_1) \cdots \rho_m(c_m). \tag{19.1}$$

Conversely, any degree 1 representation $\rho\colon A \to F^\times$ is determined by its restrictions $\rho_i\colon C_i \to F^\times$. □

Thus, we usually want to work in a field that contains "enough" roots of unity, which is why we focus on complex representations in this chapter. A famous example is the Klein group $K_4 = \langle a, b : a^2 = b^2 = (ab)^2 = 1 \rangle$. Any degree 1 representation ρ of K_4 must satisfy $\rho(a) = \pm 1$ and $\rho(b) = \pm 1$, and each of these four possibilities yields a representation, so K_4 has exactly four degree 1 representations, including the trivial representation.

These considerations become more important in light of the following observation:

Remark 19.4. For any representation $\rho\colon G/N \to \mathrm{GL}(V)$ of G/N, where $N \triangleleft G$, the composite $G \to G/N \xrightarrow{\rho} \mathrm{GL}(V)$ defines a representation $\hat{\rho}$ of G, given explicitly by

$$\hat{\rho}(g) = \rho(Ng).$$

We say that the representation ρ **lifts** to $\hat{\rho}$. Note that $\ker \rho = \ker \hat{\rho}/N$, and in particular $\ker \hat{\rho}(g) \supseteq N$.

Conversely, given a representation $\rho\colon G \to \mathrm{GL}(V)$ and $N \subseteq \ker \rho$, we have a natural representation $\bar{\rho}\colon G/N \to \mathrm{GL}(V)$ by Noether's first isomorphism theorem. In particular, any representation ρ of G is lifted from a faithful representation of the group $G/\ker \rho$.

When the representation ρ has degree 1, the group $G/\ker \rho$ is injected into $F^\times$ and thus is Abelian; in fact $G/\ker \rho$ is cyclic, by Proposition 4.1 of Volume 1. Thus, the degree 1 representations of any finite group G can be obtained by finding those normal subgroups such that G/N is cyclic, and then lifting their faithful representations. We can thereby extend Lemma 19.3 to arbitrary groups by means of the commutator subgroup G'. Recall that $G' \triangleleft G$, G/G' is Abelian, and G' is contained in every $N \triangleleft G$ such that G/N is Abelian.

PROPOSITION 19.5. *Suppose* m *is the exponent of the group* G/G'*, and* F *contains a primitive* m*-th root of* 1*. Then the number of distinct complex degree* 1 *representations of* G *is* $[G : G']$.

Proof. Let $n = [G : G']$. G/G', being Abelian, has n complex degree 1 representations, each of which provides a degree 1 representation of G, by Remark 19.4. Conversely, for ρ of degree 1, clearly $G/\ker \rho$ is Abelian (even cyclic), implying that $\ker \rho \supseteq G'$, and thus ρ is lifted from a degree 1 representation of G/G'. □

Here are some explicit examples of degree 1 representations. In view of Lemma 19.3, we consider only nonabelian groups.

Example 19.6. (i) Suppose $G = S_n$. Since the alternating subgroup A_n has index 2, the cyclic group S_n/A_n has exactly two degree 1 representations (assuming $\mathrm{char}(F) \neq 2$). One of them is the unit representation, and the other, denoted **sgn**, is given by sending each permutation π to $\mathrm{sgn}(\pi) = \pm 1$. **1** and **sgn** are the only degree 1 representations of S_n; cf. Exercise 1.

(ii) $|G| = 8$. There are two nonabelian examples:

(1) $G = D_4 = \langle a, b : a^4 = b^2 = 1,\ bab^{-1} = a^{-1} \rangle$, the dihedral group. Then $N = \langle a^2 \rangle \triangleleft G$, and from Lemma 19.3, $G/N \cong K_4$ has four degree 1 representations, which lift to four degree 1 representations of G.

(2) $G = Q_8 = \langle a, b : a^4 = b^4 = 1,\ a^2 = b^2,\ bab^{-1} = a^{-1} \rangle$, the quaternion group. Again, $N = \langle a^2 \rangle \triangleleft G$, and $G/N \cong K_4$ has four degree 1 representations, which lift to four degree 1 representations of G.

F.d. representations of degree greater than 1.

We now turn to representations of degree > 1. Some representations can be constructed directly from representations of smaller degree.

Definition 19.7. Given two representations ρ, φ of respective degrees m, n, we define their **direct sum** $\rho \oplus \varphi$ of degree $m + n$, by

$$(\rho \oplus \varphi)(g) = \begin{pmatrix} \rho(g) & 0 \\ 0 & \varphi(g) \end{pmatrix}$$

(written as blocks of sizes $m \times m$, $m \times n$, $n \times m$, and $n \times n$.).

Are there any other examples? To gain intuition, let us see what we can say about the matrix $\rho(g)$ at this early stage.

Remark 19.8. (i) If $a, b \in G$ are conjugates, i.e., $a = hbh^{-1}$ for some $h \in G$, then

$$\rho(a) = \rho(h)\rho(b)\rho(h)^{-1},$$

so $\rho(a)$ and $\rho(b)$ have the same minimal polynomial and also the same characteristic polynomial.

(ii) Suppose $g \in G$ with $g^m = 1$, and $F \subseteq \mathbb{C}$ contains a primitive m-th root of 1. Then $\rho(g)^m = \rho(g^m) = \rho(1) = I$ for any representation ρ of G. Hence, the minimal polynomial p_g of $\rho(g)$ divides $\lambda^m - 1$ whose m roots are all distinct; by Theorem 2.65 of Volume 1, the matrix $\rho(g)$ is diagonalizable. Thus, with respect to a suitable base we may assume that

$$\rho(g) = \operatorname{diag}\{\zeta_1, \ldots, \zeta_n\},$$

where $n = \deg \rho$ and each ζ_i is an m-th root of 1. Furthermore,

$$\rho(g^{-1}) = \rho(g)^{-1} = \operatorname{diag}\{\zeta_1^{-1}, \ldots, \zeta_n^{-1}\} = \operatorname{diag}\{\overline{\zeta_1}, \ldots, \overline{\zeta_n}\},$$

where $\bar{}$ denotes complex conjugation.

It follows at once that the minimal polynomial of $\rho(g^{-1})$ is the complex conjugate of p_g. If g^{-1} is conjugate to g, then by (i), the coefficients of p_g are all real.

For example, given a complex representation ρ of $C_2 = \{1, a\}$ of degree 2, we may assume after a change of base that the matrix $\rho(a)$ is a diagonal matrix whose eigenvalues are ± 1; since $\rho(1) = I$, we see that (up to change of base) the only possibilities for ρ are direct sums of degree 1 representations. Likewise, we shall see that every complex representation of any finite Abelian group becomes a direct sum of degree 1 representations after a suitable change of base.

The situation is more complicated for nonabelian groups.

Example 19.9. (i) A faithful complex representation ρ of S_3 having degree 2. Clearly, $\rho((12))$ must have order 2 and $\rho((123))$ must have order 3. We may assume that $\rho((123))$ is diagonal, with eigenvalues in $\{1, \zeta, \zeta^2\}$, where ζ is a primitive cube root of 1. But $(123)^{-1} = (132) = (12)(123)(12)^{-1}$ is conjugate to (123), so $\rho(123)^{-1}$ has the same eigenvalues as $\rho((123))$, implying that $\rho((123)) = \operatorname{diag}\{\zeta, \zeta^2\}$ (or $\operatorname{diag}\{\zeta^2, \zeta\}$). We take $\rho(12) = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, since then $\rho((12))\rho((123))\rho((12))^{-1} = \rho((123))^{-1}$; i.e., ρ "preserves" the defining relation of the group and thus is a group representation. Clearly, ρ is faithful. Actually, we obtain this representation by viewing S_3 as the dihedral group D_3 of order 6, and this example can be generalized easily to any dihedral group.

(ii) A faithful representation ρ of S_3 over $\mathbb{Q}$, having degree 2. Now we have to be a bit more subtle. Since $\lambda^3 - 1$ factors as $(\lambda^2 + \lambda + 1)(\lambda - 1)$ in $\mathbb{Q}[\lambda]$, and since $\rho((123)) \neq I$, the minimal polynomial of $\rho((123))$ must be $\lambda^2 + \lambda + 1$. We may put $\rho((123))$ in rational canonical form with respect to a suitable base, i.e., $\rho((123)) = \left(\begin{smallmatrix} 0 & 1 \\ -1 & -1 \end{smallmatrix}\right)$. Coincidentally, $\rho((12)) = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ again defines the desired representation.

(iii) A faithful representation of the dihedral group D_4 (notation as in Example 19.6(ii)), having degree 2:

$$a \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \qquad b \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

One has $\rho(b)\rho(a)\rho(b)^{-1} = \rho(a)^{-1}$ by direct computation.

These considerations lead us to several major questions:

1. When are two representations "essentially the same"?
2. When is a representation "essentially the same" as a suitable direct sum of other representations?
3. Can we utilize representations without guessing or computing each one in gory detail?

As we shall see, these questions become much easier to handle when we bring in structure theory.

Modules and vector spaces over groups

Although historically the theory of group representations preceded the theory of modules, we motivate our study with the structure of modules, since it gives us a ready framework on which to build the theory. Toward this end, we start with a monoid-theoretic version of modules. The connection to groups is in the following observation:

Remark 19.10. If $f\colon G \to M$ is a homomorphism of monoids and G is a group, then $f(G)$ is also a group. Indeed $f(g)f(g^{-1}) = f(1) = 1 = f(g^{-1})f(g)$, so $f(g^{-1}) = f(g)^{-1}$.

Our next definition (and subsequent theory) is modelled on the familiar notion of a module over a ring, but forgetting that the ring is endowed with addition.

Definition 19.11. A **module over** a monoid G, or G**-module**, is an Abelian group V together with an action $G \times V \to V$ satisfying the following laws for all g, h in G and v, w in V:

(i) $1v = v$;

(ii) $(gh)v = g(hv)$;

(iii) $g(v + w) = gv + gw$.

A G**-module map** is a group homomorphism $f\colon V \to W$ such that $f(gv) = gf(v)$, $\forall g \in G, v \in V$.

When a G-module V is a vector space over a given field F, we say that V is a G**-space** if it also satisfies the condition

(iv) $g(\alpha v) = \alpha(gv)$, $\forall \alpha \in F,\ g \in G, v \in V$.

Here is a basic example.

Example 19.11′. Any Abelian group V becomes a G-module via the **trivial action** $gv = v$, $\forall g \in G,\ v \in V$.

Although the terminology G-module is more common, we work in this chapter only with vector spaces over a given field F; hence, this stronger notion is more appropriate. Thus, G**-subspace** means a G-submodule that is also a vector subspace over F.

In our application of Definition 19.11, the monoid G is almost always a group. Note that the first two axioms are those of a group action on a set, whereas (iii) and (iv) show that left multiplication by g on V is a linear transformation of V. Thus, we get the following basic correspondence.

PROPOSITION 19.12. *For any group G, there is a* 1:1 *correspondence (described explicitly in the proof) between*

$$\{\textit{Group representations } G \to \mathrm{GL}(V)\} \textit{ and } \{G\textit{-space structures on } V\}.$$

Proof. ($\Rightarrow$) Given a group representation $\rho\colon G \to \mathrm{GL}(V)$, define multiplication $G \times V \to V$ by

$$gv = \rho(g)(v).$$

Conditions (i) and (ii) hold because ρ is a group homomorphism, and (iii) and (iv) hold because $\rho(g) \in \mathrm{GL}(V)$.

($\Leftarrow$) We reverse the argument, given a G-space structure on V. For each $g \in G$, define $\rho(g)\colon V \to V$ to be the map given by $v \mapsto gv$. Then $\rho(g) \in \mathrm{End}_F(V)$ by (iii) and (iv), and ρ is a monoid homomorphism by (i) and (ii), implying that $\rho(G) \leq GL(V)$ by Remark 19.10, so ρ is a group representation. □

Besides freeing us from matrix coefficients, this result transports us to the structure theory of G-spaces. Unfortunately, at this stage we know nothing about G-spaces, whereas we have studied modules over algebras. So we would like to find some algebra whose modules correspond naturally to the G-spaces. There is a perfect candidate.

Group algebras

Since no extra effort is involved, we work over an arbitrary commutative ring C.

Definition 19.13. For any set X, we define the **free C-module** CX with base indexed by the elements of X, a typical element of CX being $\sum_{g\in X} \alpha_g g$, where each $\alpha_g \in C$ and almost all $\alpha_g = 0$.

Now suppose X is a group G. The **group algebra** $C[G]$ is defined to be the free C-module CG, also endowed with multiplication given by

$$\left(\sum_{g\in G} \alpha_g g\right)\left(\sum_{h\in G} \beta_h h\right) = \sum_{s\in G}\left(\sum_{gh=s} \alpha_g \beta_h\right) s, \tag{19.2}$$

which makes $C[G]$ an algebra over C. (This is not to be confused with the structure in Definition 10.23 of Volume 1, which is always a commutative algebra!)

Although the notation in (19.2) might seem awkward, it is reminiscent of the polynomial algebra $C[\lambda]$, and (19.2) follows naturally from the requirement that the product of elements g and h in $C[G]$ be the same as their product gh in G, and that distributivity holds. (In fact this construction is a special case of the monoid algebra of Exercise 17.1.) $C[G]$ provides us a method of passing from group homomorphisms to algebra homomorphisms and back again.

Remark 19.14. There is a natural monoid homomorphism $G \to C[G]$ sending $g \mapsto 1_C g$ (where 1_C is the unit element of C); identifying g with $1_C g$ we can view G as a multiplicative subgroup of $C[G]$.

LEMMA 19.15. *Suppose $U = \mathrm{Unit}(R)$ is the group of invertible elements of a C-algebra R.*

(i) *Any C-algebra homomorphism $f\colon C[G] \to R$ restricts to a group homomorphism $f|_G : G \to U$.*

(ii) *Conversely, for any group homomorphism $f\colon G \to U$, there is a unique C-algebra homomorphism $\hat{f}\colon C[G] \to R$ whose restriction to G is f.*

(iii) *Parts* (i) *and* (ii) *are inverse correspondences, thereby yielding a* 1:1 *correspondence between* {*group homomorphisms* $G \to U$} *and* {*algebra homomorphisms* $C[G] \to R$}.

Proof. (i) By hypothesis, $f|_G$ preserves multiplication, and $f(G) \leq U$ by Remark 19.10.

(ii) Any such homomorphism $\hat{f}$ must be unique, satisfying

$$\hat{f}\left(\sum \alpha_g g\right) = \sum \alpha_g \hat{f}(g) = \sum \alpha_g f(g),$$

and we check that this formula indeed defines a homomorphism:

$$\begin{aligned}
\hat{f}\left(\sum \alpha_g g\right)\hat{f}\left(\sum \beta_h h\right) &= \left(\sum \alpha_g f(g)\right)\left(\sum \beta_h f(h)\right) \\
&= \sum_{g,h} \alpha_g \beta_h f(g) f(h) \\
&= \sum_{g,h} \alpha_g \beta_h f(gh) \\
&= \sum_{s \in G} \sum_{gh=s} \alpha_g \beta_h f(s) \\
&= \hat{f}\Big(\sum_{s \in G} \sum_{gh=s} \alpha_g \beta_h s\Big) \\
&= \hat{f}\left(\left(\sum \alpha_g g\right)\left(\sum \beta_h h\right)\right).
\end{aligned}$$

(iii) Follows at once from (i) and (ii). □

Put in a more categorical perspective, Lemma 19.15 yields

Remark 19.16. There is a functor C-**Alg** $\to$ **Grp** sending an algebra R to Unit(R). Recalling the terminology "universal" from Definition 8.5 of Volume 1, we can restate Lemma 19.15(ii) as follows: *The group algebra* $C[G]$ *(together with the natural inclusion map* $G \hookrightarrow \mathrm{Unit}(C[G])$*) is the universal from the group* G *to this functor.*

Thus, for any field F, the theory of group representations of a group G corresponds precisely to the theory of algebra representations of the group algebra $F[G]$. In the case when $R = F[G]$, in view of Remark 19.16, the regular representation of Example 13.49 is the algebra representation that corresponds to the regular group representation ρ_{reg} of G. (In retrospect, we could have used Remark 19.16 together with Lemma 19.15 to prove Proposition 19.12.)

Example 19.17. The **augmentation map** is the algebra representation $\epsilon\colon C[G] \to C$ corresponding to the trivial group representation; thus $\sum_{g \in G} \alpha_g g \mapsto \sum_{g \in G} \alpha_g$. The kernel, called the **augmentation ideal**, is $\{\sum_g \alpha_g g : \sum \alpha_g = 0\}$, clearly a nonzero ideal if $G \neq (1)$.

In particular, $C[G]$ is never a simple ring when $|G| > 1$.

Let us summarize the connections among the various structures.

Proposition 19.18. *Given a vector space* V *over a field* F*, we have a* 1:1 *correspondence between:*

(i) *group representations* $\rho\colon G \to \mathrm{GL}(V)$;
(ii) *algebra representations* $F[G] \to \mathrm{End}_F V$;
(iii) G*-space structures on the vector space* V*; and*
(iv) $F[G]$*-module structures on* V.

Proof. (i) $\Leftrightarrow$ (ii) By Lemma 19.15, with $R = \mathrm{End}_F V$ and $U = \mathrm{GL}(V)$.

(i) $\Leftrightarrow$ (iii) By Proposition 19.12.

(iv) $\Rightarrow$ (iii) Restrict the action to the group G.

(iii) $\Rightarrow$ (iv) Define the module operation as

$$\sum_g (\alpha_g g)v = \sum_g \alpha_g (gv),$$

where $\alpha_g \in F$, $g \in G$, and $v \in V$, and verify distributivity. □

We were searching for the correspondence from (i) to (iv), which enables us to bring in the module theory intact; we received (ii) as a bonus. All in all, we have four languages in which to develop the theory, and our main strategy is to translate results that we already have in one language (usually modules over $F[G]$) to the others. Let us proceed to carry out this program. First an easy observation.

Remark 19.19. $F[G]$ is a commutative ring iff G is an Abelian group. (Indeed, either condition holds iff $ab = ba$ for each a, b in G.)

Definition 19.20. Two representations ρ and τ of G over F are **equivalent** if their corresponding $F[G]$-modules are isomorphic.

Remark 19.20′. Let us interpret the definition for group representations. Suppose representations ρ and τ provide isomorphic $F[G]$-module structures on $V = F^{(n)}$. Thus, there is a linear transformation $T\colon V \to V$ that also is a G-module map; i.e.,

$$T(\rho(g)v) = \tau(g)T(v)$$

for all g in G and all v in V.

Writing A for the matrix corresponding to the linear transformation T, so that $T(v) = Av$ for all $v \in V$, we see that

$$A\rho(g)v = \tau(g)Av$$

for all $v \in V$; thus $A\rho(g) = \tau(g)A$ for all $g \in G$, implying

$$\tau(g) = A\rho(g)A^{-1}, \quad \forall g \in G. \tag{19.3}$$

(In the original terminology, A was called an **intertwining map**.)

Conversely, if representations ρ and τ of degree n satisfy (19.3) for some matrix $A \in \mathrm{GL}(n, F)$, then the corresponding $F[G]$-modules are isomorphic. Viewing A as a 'change of base' matrix, we see that two representations of degree n are equivalent iff they become the same after a suitable change of base of $F^{(n)}$.

Next let us interpret for a representation ρ what it means for its $F[G]$-module V to be simple; i.e., when V has no proper nonzero G-subspaces.

Definition 19.21. A representation is **irreducible** iff it corresponds to a simple $F[G]$-module. A representation is **reducible** if it is not irreducible.

Remark 19.21′. The simple $F[G]$-modules can all be recovered as the factors in any composition series $\mathcal{C}$ of $F[G]$. This is seen at once from following argument based on the Jordan-Hölder Theorem (Theorem 3.13 of Volume 1): Any simple $F[G]$-module M can be written as $F[G]/L$ for a maximal left ideal L, and the chain $F[G] \supset L \supset 0$ can be refined to a composition series equivalent to $\mathcal{C}$; hence M appears as a composition factor of $\mathcal{C}$. In particular, over any field F, each finite group has only finitely many irreducible representations (up to equivalence).

Reducibility of a given representation may depend on the field F; cf. Exercise 2. Let us see what reducibility means in matrix terms. Suppose a representation $\rho\colon G \to \mathrm{GL}(n, F)$ is reducible. In other words, $V = F^{(n)}$ has a proper nonzero G-subspace W. Let $m = \dim_F W < n$ and extend an F-base $b_1, \ldots, b_m$ of W to an F-base $b_1, \ldots, b_n$ of V. Taking the equivalent representation τ with respect to this new base $b_1, \ldots, b_n$, we can partition any $n \times n$ matrix as $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, where A is $m \times m$, B is $m \times (n-m)$, C is $(n-m) \times m$, and D is $(n-m) \times (n-m)$. Thus, writing

$$\tau(g) = \begin{pmatrix} A(g) & B(g) \\ C(g) & D(g) \end{pmatrix},$$

we have $\tau(g)(w) \in W$ for all w in W; i.e., $C(g)W = 0$, so $C(g) = 0$. We conclude for all g in G that $\tau(g)$ is of the form

$$\tau(g) = \begin{pmatrix} A(g) & B(g) \\ 0 & D(g) \end{pmatrix}. \tag{19.4}$$

But then

$$\begin{aligned} \begin{pmatrix} A(gh) & B(gh) \\ 0 & D(gh) \end{pmatrix} &= \tau(gh) = \tau(g)\tau(h) \\ &= \begin{pmatrix} A(g) & B(g) \\ 0 & D(g) \end{pmatrix} \begin{pmatrix} A(h) & B(h) \\ 0 & D(h) \end{pmatrix} \\ &= \begin{pmatrix} A(g)A(h) & A(g)B(h) + B(g)D(h) \\ 0 & D(g)D(h) \end{pmatrix}. \end{aligned}$$

Matching components in the corners shows that

$$A(gh) = A(g)A(h) \qquad \text{and} \qquad D(gh) = D(g)D(h).$$

In this manner, our setup gives rise to two more representations:

1. τ restricts to a representation $\tau|_W$ sending $g \mapsto A(g)$, which has degree m and corresponds to the G-subspace W;
2. The representation $g \mapsto D(g)$ has degree $n - m$ and corresponds to the factor G-space V/W. (This is seen intuitively by identifying the vectors in W with 0.)

Conversely, if $\tau(g)$ has this form $\left(\begin{smallmatrix} A(g) & B(g) \\ 0 & D(g) \end{smallmatrix}\right)$ for all g in G (where $A(g)$ has size $m \times m$, etc.), then the subspace W generated by the first m elements of the base of V satisfies $\tau(g)W \subseteq W$ for all g in G; i.e., W is a G-subspace of V, implying that τ is reducible. Thus, we have proved:

Proposition 19.22. *A group representation ρ of degree n is reducible, iff there is a representation τ equivalent to ρ for which each matrix $\tau(g)$, $g \in G$, has the form (19.4) (for suitable $1 \le m < n$).*

Corollary 19.23. *If $\rho\colon G \to \mathrm{GL}(n, F)$ is a representation such that the corresponding algebra homomorphism $\hat{\rho}\colon F[G] \to M_n(F)$ is onto, then ρ is irreducible.*

Proof. Otherwise, replacing ρ by a suitable equivalent representation, we could assume that $\rho(g)$ always has the form (19.4). But then $\hat{\rho}(\sum \alpha_g g) = \sum \alpha_g \rho(g)$ also has this form (since the bottom left corner, being the sum of 0's, is also 0), contrary to $\hat{\rho}$ being onto. □

(The converse actually was proved already in slightly different language in Corollary 15A.4; a more intrinsic argument not relying on the density theorem is given in Remark 19.35 below.)

Let us carry this discussion further, and see what happens in (19.4) when also $B(g) = 0$ for all $g \in G$. In this case W' is also a G-subspace of V, so we have $V = W \oplus W'$ as $F[G]$-modules; writing τ, τ' for the representations corresponding to W, W', we clearly have $\rho = \tau \oplus \tau'$.

Definition 19.24. A representation is called **completely reducible** if it is the direct sum of irreducible representations.

Remark 19.24′. A representation is completely reducible iff its corresponding module is semisimple.

Maschke's Theorem.

Having linked the theory of group representations to modules over group algebras, our next task is to examine such modules. Suppose the group algebra $F[G]$ is a semisimple ring. Then by Proposition 14.23, every $F[G]$-module is a direct sum of simple submodules, and thus every representation is completely reducible. This often turns out to be the case. To see why, we consider a method of turning linear transformations (over F) into module maps (over $F[G]$).

LEMMA 19.25. *Let M_i be $F[G]$-modules for a finite group G, and suppose $\psi\colon M_1 \to M_2$ is an arbitrary linear transformation over F. Then we have an $F[G]$-module map $\overline{\psi}\colon M_1 \to M_2$ defined by*

$$\overline{\psi}(v) = \sum_{g\in G} g^{-1}\psi(gv), \qquad \forall v \in M_1.$$

Proof. Clearly $\overline{\psi}$ is an F-linear transformation. Furthermore, for any element $h \in G$, we have the following computation:

$$\begin{aligned}
\overline{\psi}(hv) &= \sum_{g\in G} g^{-1}\psi(g(hv)) \\
&= \sum_{g\in G} g^{-1}\psi((gh)v) \\
&= \sum_{gh\in G} h(gh)^{-1}\psi((gh)v) \\
&= \sum_{g\in G} hg^{-1}\psi(gv) \\
&= h\overline{\psi}(v).
\end{aligned}$$

This proves that $\overline{\psi}\colon M_1 \to M_2$ also preserves the G-module structure, as desired. □

The map $\Phi\colon \mathrm{Hom}_F(M_1, M_2) \to \mathrm{Hom}_{F[G]}(M_1, M_2)$ given by $\psi \mapsto \overline{\psi}$ is called the **averaging procedure**. If ψ already is an $F[G]$-module map, then

$$\overline{\psi}(v) = \sum_{g\in G} g^{-1}\psi(gv) = \sum_{g\in G} \psi(g^{-1}gv) = |G|\psi(v), \tag{19.5}$$

so we see that Φ is onto when $|G|^{-1} \in F$.

THEOREM 19.26 (MASCHKE'S THEOREM). *$F[G]$ is a semisimple ring, for any finite group G whose order is not divisible by* $\mathrm{char}(F)$. *(This condition is automatic when* $\mathrm{char}(F) = 0$.*)*

Proof. By Theorem 14.13, it suffices to show that any left ideal L of $F[G]$ has a complement as an $F[G]$-module. Certainly L has a complement as a vector space over F, so define the corresponding projection $\pi\colon F[G] \to L$ and take $\overline{\pi}$ as in Lemma 19.25. Clearly, $\bar{\pi}(F[G]) \subseteq L$, so $g\bar{\pi}(a) \in gL \subseteq L$, implying $\pi(g\bar{\pi}(a)) = g\bar{\pi}(a)$ for all $g \in G$ and $a \in F[G]$. Consequently, we have the $F[G]$-module map $\hat{\pi} = \frac{1}{|G|}\overline{\pi}\colon F[G] \to L$.

For each $a \in L$ we have $ga \in L$; hence, $\pi(ga) = ga$ and

$$\hat{\pi}(a) = \frac{1}{|G|}\sum_{g\in G} g^{-1}\pi(ga) = \frac{1}{|G|}\sum_{g\in G} g^{-1}(ga) = a.$$

Hence, $\hat{\pi}(F[G]) = L$, and $\hat{\pi}$ is a projection as $F[G]$-modules. □

COROLLARY 19.27. *Under the hypotheses of Maschke's Theorem, every representation of G is completely reducible.*

Proof. Apply Remark 19.24′ and Proposition 14.23. □

As we shall see, the method of averaging applies to many other situations, in particular to (infinite) topological groups (in Appendix 19A). An alternate structural proof of Maschke's Theorem (over $\mathbb{R}$ or $\mathbb{C}$) is given in Exercise 13, using the theory of bilinear forms. A third approach (Exercise 32) provides information about the group algebra $F[G]$ for every group G.

Maschke's Theorem gives us quite explicit information about the ring-theoretic structure of $F[G]$, which we record for further use.

Remark 19.28. Suppose $n = |G|$ is not divisible by $\mathrm{char}(F)$. By Maschke's Theorem,

$$F[G] \cong R_1 \times \cdots \times R_t,$$

where R_i are simple Artinian algebras and $|G| = \dim_F F[G] = \sum_{i=1}^{t} \dim_F R_i$. Recall from Remark 14.20 that the t maximal ideals of $F[G]$ have the form

$$P_i = R_1 \times \cdots \times R_{i-1} \times 0 \times R_{i+1} \times \cdots \times R_t$$

and $R_i \cong F[G]/P_i$. Rearranging the components, we take P_1 to be the augmentation ideal (Example 19.17), and then $R_1 = F$. We have projections $\pi_i \colon F[G] \to R_i$, for $1 \leq i \leq t$.

By the Wedderburn-Artin Theorem (14.24), each $R_i \cong M_{n_i}(D_i)$, where D_i is a division algebra over F. By Corollary 15.10, we see that each R_i has a unique simple module L_i up to isomorphism, which we take to be a minimal left ideal of R_i. In particular, $\dim_F L_i \leq \dim_F F[G] = n$.

Remark 19.29. $F[G]$ has nonzero nilpotent ideals when $\mathrm{char}(F) = p \neq 0$ divides $|G|$, by an easy argument given in Exercise 23. Thus, for G finite, $F[G]$ is semisimple iff $\mathrm{char}(F)$ does not divide $|G|$.

Even when $\mathrm{char}(F) \neq 0$ and Maschke's Theorem is no longer available, we can turn to indecomposable modules and the Krull-Schmidt Theorem (Exercise 16.29). The ensuing theory, which involves an interplay with $\mathbb{Z}[G]$, is called **modular group representation theory**.

Digression: Group algebras of infinite groups.

Having obtained Maschke's Theorem, we are led to ask what happens for infinite groups. In Appendices 19A and B we consider theories for important classes of infinite groups, but we would like to drop all restrictions for the moment. How do these results generalize to arbitrary group algebras $F[G]$? If F is uncountable of characteristic 0, then $\mathrm{Jac}(F[G]) = 0$ by Exercise 34. In particular, $\mathrm{Jac}(\mathbb{C}[G]) = 0$. Perhaps the major open question remaining for group algebras is whether necessarily $\mathrm{Jac}(\mathbb{Q}[G]) = 0$. Group algebras have been studied intensively; cf. Passman [**Pas**].

Group algebras over splitting fields

Since our ultimate object is to learn about the group G, we usually want to choose the field F so that the structure of $F[G]$ as an algebra is as nice as possible, but also with the theory of F as manageable as possible. Maschke's Theorem motivates us to choose F such that $\mathrm{char}(F) \nmid |G|$ so that $F[G]$ is semisimple. This leads us to the following variant of Definition 18.30.

Definition 19.30. A field F is a **splitting field** of a group G if

$$F[G] \cong M_{n_1}(F) \times \cdots \times M_{n_t}(F)$$

for suitable n_i (where $n_1 = 1$).

It follows that $F[G]$ is split semisimple, with each simple component having center F.

Remark 19.31. By Theorem 14.27 applied to Maschke's Theorem, any algebraically closed field of characteristic 0 is a splitting field of every finite group, so we could just take $F = \mathbb{C}$ or better yet, $F = \overline{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$.

This observation can be improved.

Remark 19.32. If K/F is a field extension and G is a group, then

$$K[G] \cong K \otimes_F F[G].$$

(Indeed, the balanced map $K \times F[G] \mapsto K[G]$ given by $(k, a) \mapsto ka$ yields an algebra homomorphism $\Phi \colon K \otimes_F F[G] \to K[G]$. But $K \otimes_F F[G]$ and $K[G]$ are both free K-modules with base G (cf. Remark 18.27), which is fixed by Φ, so Φ is an isomorphism.)

Proposition 19.33. *Any finite group G has a splitting field that is a finite extension of $\mathbb{Q}$.*

Proof. Apply Remarks 19.31 and 19.32 to Proposition 18.31. □

Thus, tensor products are a powerful tool in the theory of group representations. Surprisingly, when $|F|$ is finite and relatively prime to $|G|$, $F[G]$ must be split, because of another theorem of Wedderburn; cf. [**Row3, Appendix B**]) or Theorem 24.42. (But F still may not be a splitting field of G; cf. Exercise 21.) For arbitrary fields, by a theorem of Brauer [**Jac5, Theorem 5.25**], if $\exp(G) = m$, then any field F containing a primitive m-th root of 1 is a splitting field of G.

Hypothesis 19.34. We assume until Example 19.44 that the F is a splitting field of the finite group G. Following the notation of Remark 19.28, we write $F[G] \cong R_1 \times \ldots \times R_t$, where each $R_i = M_{n_i}(F)$ and $n_1 = 1$. L_i is a minimal left ideal of R_i.

Remark 19.35. Since $\dim_F L_i = n_i$ and $L_1, \dots, L_t$ are all nonisomorphic as $F[G]$-modules, there are precisely t inequivalent irreducible representations of G, to be denoted throughout as $\rho_1, \dots, \rho_t$, having respective degrees $n_1, \dots, n_t$. But in view of Corollary 19.23, each projection $\pi_i\colon F[G] \to R_i$ restricts to an irreducible representation of G, so these are precisely the t inequivalent irreducible representations $\rho_1, \dots, \rho_t$, and we may take ρ_i to be the restriction of π_i to G, $1 \le i \le n$.

The interchange between π_i, ρ_i, and the minimal left ideal L_i of R_i is crucial in the sequel. Note that $\{\rho_i(g) : g \in G\}$ spans $\pi_i(F[G]) = R_i = M_{n_i}(F)$, so in conjunction with Corollary 19.23, we have the following important characterization of irreducible representations:

PROPOSITION 19.36. *For any splitting field F of the group G, a representation ρ of degree n is irreducible iff $\{\rho(g) : g \in G\}$ spans $M_n(F)$.*

For example, for ρ irreducible, if $g \in Z(G)$, then $\rho(g)$ commutes with all of $M_n(F)$ and thus is a scalar matrix.

Remark 19.36′. Since $\dim_F M_n(F) = n^2$, we see for ρ irreducible that there are elements $g_1, \dots, g_{n^2}$ in G such that $\{\rho(g_1), \dots, \rho(g_{n^2})\}$ is a base of $M_n(F)$ over F. Although this argument was presented for splitting fields, we get it for arbitrary fields F with $\mathrm{char}(F) \nmid |G|$ by passing to the algebraic closure $\overline{F}$. Indeed, if $\{\rho(g_1), \dots, \rho(g_{n^2})\}$ are independent over $\overline{F}$, they certainly are independent over F. (One can get this result for any field F by using Corollary 15A.4.)

Remark 19.37. The number of components of $F[G]$ isomorphic to F is precisely $[G : G']$, in view of Proposition 19.5.

We can summarize the situation quite concisely for Abelian groups:

PROPOSITION 19.38. *The following are equivalent (for F a splitting field of a finite group G):*

(i) *G is Abelian.*
(ii) *The group algebra $F[G]$ is commutative.*
(iii) *$F[G] \cong F \times F \times \cdots \times F$.*
(iv) *Every irreducible representation of G has degree* 1.

Proof. (i) $\Leftrightarrow$ (ii) By Remark 19.19.

(ii)$\Leftrightarrow$ (iii) $\Leftrightarrow$ (iv) $F[G]$ is commutative iff each of its components $M_{n_i}(F)$ is commutative, iff each $n_i = 1$. □

Remark 19.39. The following formula follows from Remark 19.35:

$$|G| = \dim_F F[G] = \sum_{i=1}^{t} \dim_F M_{n_i}(F) = \sum_{i=1}^{t} n_i^2 = 1 + \sum_{i=2}^{t} n_i^2. \tag{19.6}$$

For low values of n, this formula often enables us to determine t with hardly any other prior knowledge of the group structure of G. For example, although it is well known that any nonabelian group of order 6 (resp. of order 8) is isomorphic to the symmetric group S_3 (resp. the dihedral group or the quaternion group of order 8), let us forget these facts for the moment.

For $n = 6$, note that $6 = 1 + 1 + 2^2$ is the only way of writing 6 as a sum of positive squares (not all 1), so G is nonabelian iff $t = 3$, in which case $F[G] \cong F \times F \times M_2(F)$.

For $n = 8$, likewise $8 = 1 + 1 + 1 + 1 + 2^2$, so G is nonabelian iff $t = 5$. In this case G has precisely one complex representation of degree 2 (up to equivalence) to go along with the four complex representations of degree 1, so $\mathbb{C}[G] \cong \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times M_2(\mathbb{C})$.

It is more difficult to describe the natural embedding $G \hookrightarrow F[G]$ in terms of the algebra structure of $F[G]$. Any element g corresponds to the t-tuple $(\rho_1(g), \rho_2(g), \dots, \rho_t(g))$, where $\rho_1(g) = 1$. Let us see what this means in the easiest nontrivial example:

Example 19.40. $|G| = 2$. Thus, $G = \{1, a\}$ with $a^2 = 1$. Then, for any splitting field F of G, $F[G] \cong F \times F$. This isomorphism sends $1 \mapsto (1, 1)$ and $a \mapsto (1, -1)$; cf. Example 19.3. In the other direction, we can express the idempotents $(1, 0)$ and $(0, 1)$ of $F[G]$ as linear combinations of 1 and a:

$$(1, 0) = \frac{1 + a}{2}; \qquad (0, 1) = \frac{1 - a}{2}. \tag{19.7}$$

This calculation can be used to prove from scratch that any field F of characteristic $\neq 2$ is a splitting field for C_2, since Equation (19.7) yields the two simple central idempotents for $F[C_2]$. (See Lemma 20.9 for a general result along these lines.)

Splitting fields have a downside. Our eagerness to obtain a good structure for $F[G]$ has come at the expense of losing some information about G. Read pessimistically, Proposition 19.38 says that $\mathbb{C}[G]$ has the same structure for all Abelian groups G of order n. Likewise, the complex representation theories of the dihedral and quaternion groups of order 8 are identical. For this reason, one might want to look at fields F that are not necessarily splitting fields, cf. Exercises 19 and 21, but we do not pursue this direction.

The center of the group algebra.

We have seen in Remark 19.35 that when F is a splitting field of a finite group G, the number t of simple components of $F[G]$ is also the number of irreducible representations of G. There is another important description of this number. The center, $\mathrm{Cent}(F[G])$, is $F \times \cdots \times F$, taken t times, so $\dim_F \mathrm{Cent}(F[G]) = t$.

We would like a base of $\mathrm{Cent}\, F[G]$ that is given explicitly in terms of the elements of G. For future reference, let us study this problem more generally for group algebras as free modules over an arbitrary commutative ring C.

If G is Abelian, then $C[G]$ is commutative, $t = |G|$, and G itself is the base. For G nonabelian, we want to enlarge $Z(G)$ to a base of $\mathrm{Cent}(C[G])$. Toward this end, we recall the equivalence relation on a group G, given by $a \sim b$ iff $b = gag^{-1}$ for some g in G; the equivalence classes are called "conjugacy classes."

Remark 19.41. Given a conjugacy class $\mathcal{C}$ of G, define $z_{\mathcal{C}} = \sum_{g\in\mathcal{C}} g$. Then $z_{\mathcal{C}} \in \mathrm{Cent}(C[G])$. Indeed, for any a in G we have

$$az_{\mathcal{C}}a^{-1} = \sum_{g\in\mathcal{C}} aga^{-1} = \sum_{aga^{-1}\in\mathcal{C}} aga^{-1} = z_{\mathcal{C}}.$$

Note that for any $z \in Z(G)$, its conjugacy class $\mathcal{C}$ is just $\{z\}$ since $aza^{-1} = z$; thus $z_{\mathcal{C}} = z$. However, for any conjugacy class $\mathcal{C}$ containing a noncentral element, $z_{\mathcal{C}} \notin G$.

Theorem 19.42. *For any commutative ring C, $\mathrm{Cent}(C[G])$ is free as a C-module having base $\{z_{\mathcal{C}} : \mathcal{C}$ is a conjugacy class of $G\}$.*

Proof. Let $Z = \mathrm{Cent}(C[G])$. We just saw that each $z_{\mathcal{C}} \in Z$. To prove that the $z_{\mathcal{C}}$ are independent, write $\mathcal{C}(g)$ for the conjugacy class of g. Suppose

$$0 = \sum_{\mathcal{C}} c_{\mathcal{C}} z_{\mathcal{C}} = \sum_{\mathcal{C}} c_{\mathcal{C}} \sum_{g\in\mathcal{C}} g = \sum_{\mathcal{C}} \sum_{g\in\mathcal{C}} c_{\mathcal{C}} g = \sum_{g\in G} c_{\mathcal{C}(g)} g;$$

thus each $c_{\mathcal{C}} = 0$.

It remains to show that the $z_{\mathcal{C}}$ span Z. Suppose $z = \sum \alpha_g g \in Z$. For any $h \in G$,

$$\sum \alpha_g g = z = hzh^{-1} = \sum_{g\in G} \alpha_g hgh^{-1} = \sum_{g\in G} \alpha_{h^{-1}gh} g,$$

where in the last sum we replaced g by $h^{-1}gh$. Matching coefficients shows that $\alpha_g = \alpha_{h^{-1}gh}$ for all h in G; i.e., the coefficient is the same for conjugate

elements. In particular, we can define $\alpha_{\mathcal{C}}$ to be α_g for any g in $\mathcal{C}$, and rewrite

$$z = \sum_{g\in G} \alpha_g g = \sum_{\mathcal{C}} \left(\sum_{g\in\mathcal{C}} \alpha_{\mathcal{C}} g \right) = \sum_{\mathcal{C}} \alpha_{\mathcal{C}} \sum_{g\in\mathcal{C}} g = \sum_{\mathcal{C}} \alpha_{\mathcal{C}} z_{\mathcal{C}}. \qquad \square$$

Corollary 19.43. *Suppose F is a splitting field for a finite group G. The following numbers are equal (and are each denoted as t in this chapter):*

(i) *The number of conjugacy classes of G.*
(ii) *The number of inequivalent irreducible representations of G.*
(iii) *The number of simple components of $F[G]$.*
(iv) $\dim_F \mathrm{Cent}(F[G])$.

Proof. (i) equals (iv) by the theorem, and we already showed that (ii), (iii), and (iv) are equivalent. $\square$

Example 19.44. Let us determine t for $G = S_n$. We recall from [**Row3, Remark 5.20**] that two permutations are conjugate iff they can be written as products of disjoint cycles of the same respective lengths. Thus, S_3 has three conjugacy classes, represented by the permutations (1), (12), and (123); hence, S_3 has precisely three inequivalent irreducible representations, two of which are of degree 1 and the third (of degree 2) given in Example 19.9(ii). (We also obtained $t = 3$ via Formula (19.6).)

Likewise, S_4 has five conjugacy classes, represented by (1), (12), (123), (1234), and (12)(34), so S_4 has five inequivalent irreducible representations.

Soon we investigate the representations of S_n in far greater depth, following Definition 19.49.

The case when F is not a splitting field

At times, the ground field F may not be a splitting field for G. For example, one might want F to be an ordered field, in particular $F = \mathbb{R}$.

Remark 19.45. Assume that $\mathrm{char}(F) \nmid |G|$, so $F[G] \cong R_1 \times \cdots \times R_t$, where the $R_i = M_{n_i}(D_i)$ are the simple components. In view of Proposition 19.18, the irreducible representations of G correspond to the minimal left ideals L_i of $F[G]$ (one isomorphism class for each component) and have degree $\dim_F L_i = \dim_F R_i e_{11} = n_i \dim_F D_i$. Thus, (19.6) now becomes

$$\sum_{i=1}^{t} n_i^2 \dim_F D_i = n, \tag{19.8}$$

which is less convenient. (For F a splitting field, $D_i = F$, so $\dim_F D_i = 1$, and we are back to our previous considerations.)

Fortunately, for $F = \mathbb{R}$, Frobenius' theorem (cf. Remark 14.31) says that the only noncommutative f.d. division algebra over $\mathbb{R}$ is Hamilton's algebra of quaternions $\mathbb{H}$; cf. Example 14.29. Hence, in this case, D_i must be $\mathbb{R}, \mathbb{C}$, or $\mathbb{H}$, so $\dim_F D_i = 1, 2$, or 4.

Example 19.46. We describe every possible irreducible real representation for some groups that we have already encountered. The Chinese Remainder Theorem (CRT) is useful in determining the structure of their group algebras.

(i) The cyclic group $C_n = \langle a \rangle$ of order n. Since the only real roots of 1 are ± 1, we have two representations of degree 1 when n is even, and only one representation (the trivial representation) when n is odd. To wit:

$$\mathbb{R}[C_2] \cong \mathbb{R}[\lambda]/\langle \lambda^2 - 1 \rangle \cong \mathbb{R}[\lambda]/\langle \lambda - 1 \rangle \times \mathbb{R}[\lambda]/\langle \lambda + 1 \rangle \cong \mathbb{R} \times \mathbb{R}.$$

When $n = 3$, we also have the irreducible real representation of degree 2 given by $a \mapsto \left(\begin{smallmatrix} 0 & 1 \\ -1 & -1 \end{smallmatrix}\right)$; cf. Exercise 2.

$$\mathbb{R}[C_3] \cong \mathbb{R}[\lambda]/\langle \lambda^3 - 1 \rangle \cong \mathbb{R}[\lambda]/\langle \lambda - 1 \rangle \times \mathbb{R}[\lambda]/\langle \lambda^2 + \lambda + 1 \rangle \cong \mathbb{R} \times \mathbb{C}.$$

When $n = 4$, we have the irreducible representation of degree 2 given by $a \mapsto \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

$$\mathbb{R}[C_4] \cong \mathbb{R}[\lambda]/\langle \lambda^4 - 1 \rangle \cong \mathbb{R}[\lambda]/\langle \lambda - 1 \rangle \times \mathbb{R}[\lambda]/\langle \lambda + 1 \rangle \times \mathbb{R}[\lambda]/\langle \lambda^2 + 1 \rangle \cong \mathbb{R} \times \mathbb{R} \times \mathbb{C}.$$

(ii) G is the dihedral group D_4 of order 8. As noted in Examples 19.6 and 19.9(iii), all of the complex irreducible representations are already defined over $\mathbb{Q}$ (and thus over $\mathbb{R}$). In fact, $\mathbb{Q}$ is a splitting field for D_4.

(iii) $G = \langle a, b : a^4 = b^4 = 1,\ a^2 = b^2,\ bab^{-1} = a^{-1} \rangle$, the quaternion group Q. There is a group injection from Q to the multiplicative group of Hamilton's quaternion algebra $\mathbb{H}$, given by $a \mapsto i$ and $b \mapsto j$. (Note that $a^2 \mapsto -1$.) This extends to a projection $\mathbb{R}[Q] \to \mathbb{H}$, implying that $\mathbb{H}$ is one of the simple components of $\mathbb{R}[Q]$. Thus, besides its four representations of degree 1, Q has an irreducible real representation of degree $4 = \dim_{\mathbb{R}} \mathbb{H}$.

One natural way to study the irreducible representations of G over an arbitrary field F is to take a splitting field $K \supset F$ and define the **extension** of ρ to K as the composite representation

$$\rho\colon G \to \mathrm{GL}(n, F) \hookrightarrow \mathrm{GL}(n, K).$$

Thus, we can study representations over F in terms of representations over K. When $F = \mathbb{R}$ and $K = \mathbb{C}$, this process is called the **complexification** of ρ; cf. Exercise 14.

Remark 19.47. If $F[G] \cong R_1 \times \cdots \times R_t$ where R_i are simple Artinian F-algebras, then for any splitting field $K \supseteq F$ by Remark 19.32,

$$K[G] \cong F[G] \otimes_F K \cong \left(\prod_{i=1}^{t} R_i \right) \otimes_F K \cong \prod_{i=1}^{t} (R_i \otimes_F K). \tag{19.9}$$

Writing $R_i = M_{n_i}(D_i)$ for suitable division algebras D_i, we also know that

$$R_i \otimes_F K = M_{n_i}(D_i) \otimes_F K \cong M_{n_i}(D_i \otimes_F K),$$

so this raises the question of the structure of D_i and how it relates to $D_i \otimes_F K$. A thorough investigation of this question requires knowledge of f.d. division algebras, cf. Chapter 24, but becomes much easier when $F = \mathbb{R}$. Then, as we just noted, D_i must be $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$, so $D_i \otimes_{\mathbb{R}} \mathbb{C}$ is respectively:

(1) $\mathbb{R} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C}$;
(2) $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \times \mathbb{C}$ (using Example 18.29(ii));
(3) $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{C} \cong M_2(\mathbb{C})$ (which can be seen computationally, but is also an immediate consequence of Proposition 24.8).

In (1), an irreducible representation over $\mathbb{R}$ remains irreducible over $\mathbb{C}$ of the same degree; for example, the trivial representation of $\mathbb{R}$ extends to the trivial representation of $\mathbb{C}$.

In (2), an irreducible representation of $\mathbb{R}$ of degree $2n_i$ splits into two irreducible representations of degree n_i over $\mathbb{C}$, which can be viewed as complex conjugates of each other. For example, this happens with the degree 2 representation of the cyclic group C_4 in Example 19.46(i).

In (3), an irreducible representation corresponding to a minimal left ideal of $M_{n_i}(\mathbb{H})$ has degree $4n_i$, and shrinks to an irreducible representation of degree $2n_i$ over $\mathbb{C}$, as seen with the quaternion group in Example 19.46(iii).

Working backwards, we see that Remark 19.47 yields

Proposition 19.48. *Any complex irreducible representation of G of degree n_i either is extended from a real irreducible representation or corresponds to a real irreducible representation of degree $2n_i$.*

This result is similar to the fact from field theory that any polynomial over $\mathbb{R}$ splits into a product of linear and quadratic factors. But here we have collapsed together two cases (2) and (3), which are intrinsically different.

Supplement: Group algebras of symmetric groups

Since, by Cayley's Theorem, any finite group is isomorphic to a subgroup of the symmetric group S_n, the representations of S_n are of fundamental

importance in the general theory. The structure of $\mathbb{Q}[S_n]$ was determined independently by Young and Frobenius, who showed in particular that $\mathbb{Q}$ is a splitting field of S_n for all n. We present Young's explicit method (perfected by von Neumann), which introduces some ideas basic to the theory of combinatorial algebra. In Exercises 36ff. we see how to view the combinatorics in terms of the module theory. In what follows, we work over an arbitrary field F of characteristic 0 or of characteristic $> n$.

Definition 19.49. A **(descending) partition** of n, of **length** k, is a sequence of positive integers $(\nu_1, \nu_2, \ldots, \nu_k)$ for which $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_k$ and $\sum \nu_i = n$. $\mathrm{Par}(n)$ denotes the set of partitions of n.

$\mathrm{Par}(n)$ is totally ordered via the lexicographic order of Example 0B.2 of Volume 1. We follow the custom of denoting a partition of n by the symbol λ, although everywhere else we use λ for an indeterminate.

Remark 19.50. Any permutation $\pi \in S_n$ can be written as a product of disjoint cycles $C_1 \cdots C_k$; since these cycles commute with each other, we write the product in order of descending length. Letting ν_j be the length of C_j, we thereby partition n as $(\nu_1, \ldots, \nu_k)$. On the other hand, by [**Row3, Remark 5.20**], two permutations are conjugate iff they can be written as products of disjoint cycles of the same respective lengths. Thus, the descending partitions of n are in 1:1 correspondence with the conjugacy classes of S_n.

For example, the conjugacy class in S_9 of

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 1 & 4 & 9 & 5 & 6 & 8 & 7 & 2 \end{pmatrix} = (13492)(78)(5)(6)$$

corresponds to the partition $\lambda = (5, 2, 1, 1)$ of 9.

Definition 19.51. Any partition $\lambda = (\nu_1, \ldots, \nu_k)$ yields a **(Ferrers) shape for** λ, which is a left-rectified system of k rows of boxes, the i-th row having ν_i boxes. (Note that ν_1 is the number of columns in the shape.) A **(Young) diagram**, or **tableau** of n, is a shape in which each box is filled with a different **entry** from 1 to n.

Thus, any partition gives rise to $n!$ tableaux. For example, the shape for (5,2,1,1) could have tableaux

1	3	5	7	9
2	4			
6				
8				

or

8	4	5	7	9
2	3			
1				
6				

,

altogether 9! possibilities.

We write T_λ to emphasize that the tableau T comes from the partition λ. For each tableau T, we aim to find the corresponding idempotent e_T satisfying

$$\dim_F(e_T F[S_n] e_T) = 1.$$

This will show by Digression 14.21(iv) that $F[S_n]e_T$ is a minimal left ideal of $F[S_n]$, and thus the corresponding group representation ρ_T is irreducible. Furthermore, we want all of the representations from tableaux of distinct partitions to be inequivalent; i.e., $e_{T_\lambda} F[S_n] e_{T'_{\lambda'}} = 0$ for all $\lambda \neq \lambda'$. Together, these facts will imply that F is a splitting field for S_n.

Definition 19.52. Given any subgroup H of S_n, define the following two elements of $F[S_n]$:

$$r_H = \sum_{\sigma \in H} \sigma; \quad s_H = \sum_{\tau \in H} \mathrm{sgn}(\tau)\tau.$$

Remark 19.53. (i) $\pi r_H = r_H = r_H \pi, \forall \pi \in H$.

(ii) $\pi s_H = \mathrm{sgn}(\pi) s_H = s_H \pi, \forall \pi \in H$.

Here is the key observation of the discussion.

LEMMA 19.54. *If subgroups H and K of S_n have a common transposition, then $r_H s_K = s_K r_H = 0$.*

Proof. Suppose $(i\,j) \in H \cap K$. Then

$$r_H s_K = (r_H(i\,j))s_K = r_H((i\,j)s_K) = -r_H s_K,$$

implying $r_H s_K = 0$; $s_K r_H = 0$ is proved analogously. □

We obtain subgroups of S_n by defining an action of S_n on a given tableau T, whereby any permutation π acts on the entries of T.

Definition 19.55. A permutation π is a **row permutation** of a tableau T if each row of T is invariant under π; π is a **column permutation** of T if each column of T is invariant under π. We write $P(T)$ for the subgroup of row permutations of T, and $Q(T)$ for the subgroup of column permutations of T.

In our example following Definition 19.51, the permutation sending the first tableau to the second is the column permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 8 & 2 & 4 & 3 & 5 & 1 & 7 & 6 & 9 \end{pmatrix}.$$

Let us record some facts about $P(T)$ and $Q(T)$.

Remark 19.56. Let $P = P(T)$ and $Q = Q(T)$.

(i) $P \cap Q = (1)$, since any row and column permutation must leave each entry both in the same row and the same column, and thus be the identity. It follows at once that if $\sigma_1\tau_1 = \sigma_2\tau_2$ for $\sigma_i \in P$ and $\tau_i \in Q$, then $\sigma_1 = \sigma_2$ and $\tau_1 = \tau_2$. (Indeed, $\sigma_1^{-1}\sigma_2 = \tau_1\tau_2^{-1} \in P \cap Q = (1)$.)

(ii) $r_P s_Q = \sum_{\sigma\in P,\tau\in Q} \operatorname{sgn}(\tau)\sigma\tau$, which is nonzero in view of (i).

(iii) $\pi P\pi^{-1} = P(\pi T)$ and $\pi Q\pi^{-1} = Q(\pi T)$, $\forall \pi \in S_n$, in view of [**Row3, Remark 5.20**]. In particular, $P(\sigma T) = P$, $\forall \sigma \in P$; likewise $Q(\tau T) = Q$, $\forall \tau \in Q$.

(iv) $\pi r_P \pi^{-1} = r_{P(\pi T)}$ and $\pi s_Q \pi^{-1} = s_{Q(\pi T)}$, $\forall \pi \in S_n$.

Taking $H = P(T)$ and $K = Q(T')$, we want to apply Lemma 19.54 by showing that H and K have a common transposition; this is clearly the case iff some row of T has two entries i, j that appear in a single column of T' (for this means $(ij) \in P(T) \cap Q(T')$).

LEMMA 19.57. *Suppose T and T' are tableaux arising respectively from partitions $\lambda = (\nu_1, \dots)$ and $\lambda' = (\nu'_1, \dots)$. If $\nu_1 > \nu'_1$, then the first row of T contains two entries that appear in a single column of T'.*

Proof. Obvious, since ν'_1 is the number of columns of T'. □

PROPOSITION 19.58. *Suppose T and T' are tableaux arising respectively from partitions $\lambda = (\nu_1, \dots)$ and $\lambda' = (\nu'_1, \dots)$.*

(i) *If $\lambda > \lambda' \in \operatorname{Par}(n)$, then $P(T) \cap Q(T')$ have a common transposition.*

(ii) *If $\lambda = \lambda' \in \operatorname{Par}(n)$ and $P(T) \cap Q(T')$ does not contain any transposition, then $T' = \sigma\tau T$ for suitable $\sigma \in P(T)$ and $\tau \in Q(T)$.*

Proof. (i) We want to obtain two entries ℓ, ℓ' both in the first row of T and some column of T', implying that $(\ell\,\ell') \in P(T) \cap Q(T')$. By Lemma 19.57, we are done unless $\nu_1 = \nu'_1$. Thus, T and T' have the same number of columns and, by the pigeonhole principle, we are done unless each column of T' contains exactly one element of the first row of T. Applying a suitable column permutation to T' moves all of these entries to the first row of T'. Now we can erase the first rows of T and T' and apply induction.

(ii) First, we claim that $\tilde\sigma T = \tilde\tau T'$ for suitable $\tilde\sigma \in P(T)$ and $\tilde\tau \in Q(T')$. Arguing as in (i), we may assume that each column of T' contains exactly one entry from the first row of T, and applying a suitable column permutation τ_1 to T' brings these entries to the first row, so T and $\tau_1 T'$ have the same entries in the first row, implying that $\sigma_1 T$ and $\tau_1 T'$ have the same first row for some $\sigma_1 \in P(T)$. Erasing the first row and applying induction shows

that $\sigma_2(\sigma_1 T) = \tau_2(\tau_1 T')$ for suitable $\sigma_2 \in P(T)$, $\tau_2 \in Q(T')$, so we have the claim, taking $\tilde\sigma = \sigma_2\sigma_1$ and $\tilde\tau = \tau_2\tau_1$.

Now $\tilde\tau \in Q(T') = Q(\tilde\tau T') = Q(\tilde\sigma T)$, and furthermore

$$T = \tilde\sigma^{-1}\tilde\tau(T') = (\tilde\sigma^{-1}\tilde\tau\tilde\sigma)(\tilde\sigma^{-1}(T'));$$

taking $\sigma = \tilde\sigma$ and $\tau = \tilde\sigma^{-1}\tilde\tau^{-1}\tilde\sigma \in \tilde\sigma^{-1}Q(\tilde\sigma T)\tilde\sigma = Q(\tilde\sigma^{-1}\tilde\sigma T) = Q(T)$, we have $T = \tau^{-1}\sigma^{-1}T'$, so $T' = \sigma\tau T$. □

We are ready for a key computation.

LEMMA 19.59. *Let $a_T = r_{P(T)}s_{Q(T)}$ and assume that $\operatorname{char}(F) \nmid n!$*

(i) *$a_T \rho a_T \in F a_T$ for any $\rho \in S_n$.*

(ii) *$a_T^2 = \alpha_T a_T$ for suitable $0 \neq \alpha_T \in F$.*

(iii) *$a_{T'} F[S_n] a_T = 0$ for any tableaux T, T' arising from distinct partitions.*

Proof. Write P for $P(T)$ and Q for $Q(T)$. (i) Let $a = a_T \rho a_T$, written $a = \sum_\pi \alpha_\pi \pi$. We claim that $a = \alpha_1 a_T$; i.e., $\alpha_\pi = \operatorname{sgn}(\tau)\alpha_1$ whenever $\pi = \sigma\tau$ ($\sigma \in P$, $\tau \in Q$), and all other $\alpha_\pi = 0$. So first note

$$\sigma a \tau = \sigma a_T \rho a_T \tau = \operatorname{sgn}(\tau) r_P s_Q \rho r_P s_Q = \operatorname{sgn}(\tau) a_T \rho a_T = \operatorname{sgn}(\tau) a. \tag{19.10}$$

Comparing coefficients of $\sigma\tau$ yields $\alpha_1 = \operatorname{sgn}(\tau)\alpha_{\sigma\tau}$; i.e., $\alpha_{\sigma\tau} = \operatorname{sgn}(\tau)\alpha_1$.

When π is not of the form $\sigma\tau$ then, by the contrapositive of Proposition 19.58(ii), there is a transposition $\sigma \in P \cap Q(\pi T)$. But $Q(\pi T) = \pi Q \pi^{-1}$, so $\sigma = \pi\tau\pi^{-1}$ for some transposition $\tau \in Q$. Thus, $\pi = \sigma\pi\tau$ and, by (19.10),

$$\alpha_\pi = \alpha_{\sigma\pi\tau} = \operatorname{sgn}(\tau)\alpha_\pi = -\alpha_\pi,$$

implying that $\alpha_\pi = 0$ for any such π, as desired.

(ii) By (i) (taking $\rho = 1$), we have ${a_T}^2 = \alpha_T a_T$ for some $\alpha_T \in F$. If α_T were 0, then $\operatorname{tr}(a_T) = 0$, viewing a_T as a nilpotent matrix under the regular representation (Example 19.2). But $a_T = \sum_{\sigma\in P,\tau\in Q}(\operatorname{sgn}\tau)\sigma\tau$. If $\sigma \neq 1$ or $\tau \neq 1$, then $\sigma\tau$ is a nonidentity permutation matrix and thus has trace 0 in the regular representation. Hence, $\operatorname{tr}(a_T) = \operatorname{tr}(1 \cdot 1) = \operatorname{tr}(1) \neq 0$, a contradiction.

(iii) Either $\lambda > \lambda'$ or $\lambda' > \lambda$. First assume that $\lambda > \lambda'$. By Proposition 19.58, for any $\pi \in S_n$ there is some transposition in $P(\pi T) \cap Q(T')$, implying $s_{Q(T')} r_{P(\pi T)} = 0$ by Lemma 19.54. Thus,

$$a_{T'} \pi a_T = r_{P(T')} s_{Q(T')} \pi r_{P(T)} s_{Q(T)} = r_{P(T')} s_{Q(T')} r_{P(\pi T)} \pi s_{Q(T)} = 0$$

for each $\pi \in S_n$, yielding $a_{T'} F[S_n] a_T = 0$.

If $\lambda' > \lambda$, then the same argument shows that $a_T F[S_n] a_{T'} = 0$. Hence $(a_{T'} F[S_n] a_T F[S_n])^2 = 0$, implying that $a_{T'} F[S_n] a_T F[S_n] = 0$ since $F[S_n]$ is semiprime; therefore $a_{T'} F[S_n] a_T = 0$. □

THEOREM 19.60. *Suppose* $\mathrm{char}(F) = 0$ *or* $\mathrm{char}(F) > n$. *For any tableau* T, *let* $e_T = \alpha_T^{-1} a_T$, *an idempotent of* $F[S_n]$. *Then* $e_T F[S_n] e_T = F e_T$, *so* $F[S_n]e_T$ *is a minimal left ideal of* $F[S_n]$ *and corresponds to an irreducible representation of* S_n. *In particular,* F *is a splitting field for* S_n.

If tableaux T *and* T' *come from different partitions, then their corresponding representations are inequivalent. Hence, we have a* 1:1 *correspondence between* $\mathrm{Par}(n)$ *and the classes of irreducible representations of* S_n.

Proof. Let $R = F[S_n]$ and $L_i = Re_T$. By Lemma 19.59(ii), the e_T are idempotents of R; hence, by Lemma 19.59(i), $e_T R e_T = F e_T$. Since R is semisimple, L_i is a minimal left ideal of R by Digression 14.21 and thus corresponds to an irreducible representation of S_n. Furthermore, L_i is a minimal left ideal of some simple component R_i of R, which by the Wedderburn-Artin Theorem has the form $M_{n_i}(D_i) = \mathrm{End}_{D_i} L_i$. But then $D_i \cong e_T R_i e_T = F e_T$, a copy of F, so $R_i \cong M_{n_i}(F)$ for each i, and we conclude that F is a splitting field for S_n.

If T and T' come from different partitions, then e_T and $e_{T'}$ are in different components since $e_T R e_{T'} = 0$ by Lemma 19.59(iii), so we have inequivalent irreducible representations for each element of $\mathrm{Par}(n)$. But by Remark 19.50, $|\mathrm{Par}(n)|$ also is the number of conjugacy classes of S_n, which equals the number of inequivalent irreducible representations of S_n, so we have them all. □

Let I_λ be the simple component of R corresponding to the partition λ. The idempotent $e_T \in I_\lambda$ depends on the choice of tableau $T = T_\lambda$ to fill the shape corresponding to λ. A Young tableau T is called **standard** if each row is an increasing sequence from left to right and each column is an increasing sequence from top to bottom. We write a tableau $T = (m_{i,j})$ to indicate that $m_{i,j}$ is in the i, j position. Young and Frobenius proved

THEOREM 19.61. *Suppose* $\mathrm{char}(F) = 0$ *or* $\mathrm{char}(F) > n$. *Then*

$$I_\lambda = \bigoplus_{T_\lambda \text{ standard}} F[S_n] e_{T_\lambda}.$$

Proof. We prove Theorem 19.61 in two stages. To demonstrate independence of the $F[S_n]e_{T_\lambda}$, we utilize the lexicographic order on the subscripts $(i, j) \in \mathbb{N} \times \mathbb{N}$; cf. Example 0B.2. We say that a tableau $T_\lambda = (m_{i,j})$ is

greater than $T'_\lambda = (m'_{i,j})$ if $m_{i,j} > m'_{i,j}$ at the first (i, j) for which they differ.

We claim for any two standard tableaux $T'_\lambda > T_\lambda$ that $e_{T'_\lambda} e_{T_\lambda} = 0$. Indeed, notation as in the previous paragraph, take the first pair (i_0, j_0) at which $m'_{i_0,j_0} \neq m_{i_0,j_0}$. Let $q = m_{i_0,j_0}$; then $m'_{i_0,j_0} > q$. There is some position (i_1, j_1) at which q appears in T'_λ, i.e., $m'_{i_1,j_1} = q$. Clearly $(i_1, j_1) > (i_0, j_0)$, but since T'_λ is standard, its entries increase along the rows, implying $i_1 \neq i_0$; hence $i_1 > i_0$, and since the entries of T'_λ also increase along the columns, we must have $j_1 < j_0$.

Clearly q and m_{i_0,j_1} both appear in the i_0 row of T_λ as well as in the j_1 column of T'_λ. It follows that $s_{Q(T'_\lambda)} r_{P(T_\lambda)} = 0$ by Lemma 19.54, which implies $e_{T'_\lambda} e_{T_\lambda} = 0$, as claimed.

Now it is easy to prove that the e_{T_λ} are independent (for a given partition λ). Otherwise, there would be some dependence $\sum_{i=1}^{t} a_i e_i = 0$ where $a_i \in F[S_n]$, $a_1 e_1 \neq 0$, and $e_i = e_{T_i}$; here we order the standard tableaux T_i in increasing order. But the claim shows that each $e_i e_1 = 0$, so

$$0 = \sum (a_i e_i) e_1 = a_1 e_1,$$

contrary to assumption.

To complete the proof of Theorem 19.61, it remains to show that the e_{T_λ} (for standard T_λ) span I_λ. This can be done using an elementary but tricky argument discovered by Garnir, given in Sagan [**Sag**]. This time we order tableaux by lexicographic order via the transpose of the order used above. Namely, we say that a tableau $T_\lambda = (m_{i,j})$ is greater than $T'_\lambda = (m'_{i,j})$ if $m_{i,j} > m'_{i,j}$ at the first (j, i) for which they differ. (First we look at j and then at i.)

We need to show that for any nonstandard tableau $T_\lambda = (m_{i,j})$, e_{T_λ} can be written as a linear combination of idempotents of "smaller" tableaux. First note that $e_{\tau T} = \tau e_T \tau^{-1} = \pm \tau e_T$ for any column permutation $\tau \in Q(T)$; hence, $F[S_n] e_{\tau T} = F[S_n] e_T$, and we could replace T by τT. In this way we may assume that each column of T is arranged in increasing order. Thus, for T nonstandard, we must have some row i with two consecutive decreasing entries, i.e., $m_{i,j} > m_{i,j+1}$. Then we have the increasing sequence

$$m_{1,j+1} < m_{2,j+1} < \cdots < m_{i,j+1} < m_{i,j} < m_{i+1,j} < m_{i+2,j} < \cdots < m_{\mu_j, j},$$

where μ_j is the length of the j column. This sequence contains i entries from the $j+1$ column and $\mu_j + 1 - i$ entries from the j column.

Let H_1 denote the group of permutations of the set

$$M_1 = \{m_{1,j+1}, m_{2,j+1}, \ldots, m_{i,j+1}\},$$

let H_2 denote the group of permutations of $M_2 = \{m_{i,j}, m_{i+1,j}, \dots, m_{\mu_j,j}\}$, and let K denote the group of permutations of the set $M_1 \cup M_2$. We identify $H_1 \times H_2$ with $Q(T) \cap K$; thus, we can write K as a disjoint union of left cosets

$$K = \bigcup_{\pi \in \mathcal{S}} (H_1 \times H_2)\pi, \tag{19.11}$$

where $\mathcal{S}$ denotes a transversal of $H_1 \times H_2$ in K containing (1).

For any $\tau \in Q(T)$, we see that the $\mu_j + 1$ elements $M_1 \cup M_2$ are all in the $j, j+1$ columns of $\tau^{-1}T$, so two of these must appear in the same row. It follows from Lemma 19.54 that

$$0 = r_{P(\tau^{-1}T)} s_K = \tau^{-1} r_{P(T)} \tau s_K,$$

implying $r_{P(T)} \tau s_K = 0$, and thus

$$0 = \sum_{\tau \in Q(T)} \operatorname{sgn}(\tau) r_{P(T)} \tau s_K = r_{P(T)} s_{Q(T)} s_K = \alpha_T e_T s_K.$$

On the other hand, for any $\tau \in Q(T)$, Remark 19.53(ii) yields $\operatorname{sgn}(\tau) e_T \tau = \operatorname{sgn}(\tau)^2 e_T = e_T$, implying that

$$e_T s_{H_1 \times H_2} = |H_1 \times H_2| e_T.$$

Now applying (19.11) yields

$$0 = e_T s_K = \sum_{\pi \in \mathcal{S}} \operatorname{sgn}(\pi) e_T s_{H_1 \times H_2} \pi,$$

implying that

$$\begin{aligned} e_T &= -\sum_{1 \neq \pi \in \mathcal{S}} \operatorname{sgn}(\pi) e_T s_{H_1 \times H_2} \pi = -|H_1 \times H_2| \sum_{1 \neq \pi \in \mathcal{S}} \operatorname{sgn}(\pi) e_T \pi \\ &= -\sum_{1 \neq \pi \in \mathcal{S}} |H_1 \times H_2| \operatorname{sgn}(\pi) \pi e_{\pi^{-1}T} \end{aligned} \tag{19.12}$$

by Remark 19.56(iv). But each such π^{-1} is by definition not in $H_1 \times H_2$, and thus replaces some numbers in the j column of T by lower numbers from the $j+1$ column of T, thereby reducing T in our ordering. By induction, each of these $e_{\pi^{-1}T}$ is spanned by the idempotents of standard tableaux, so by (19.12), the same holds for e_T, concluding the proof of the theorem. □

Since each $F[S_n]e_{T_\lambda}$ is a minimal left ideal, Theorem 19.61 means $I_\lambda \cong M_{n_\lambda}(F)$, where n_λ is the number of standard Young tableaux for the partition λ; in other words, n_λ is the degree of the corresponding irreducible representation, and $\sum n_\lambda^2 = [F[S_n] : F] = n!$. In the combinatorics literature, this number n_λ is customarily written f^λ. Thus, a formula for computing the f^λ should be of considerable significance. To get such a formula we need another notion.

Definition 19.62. The **hook** $H_{i,j}$ is the set of those boxes in the shape for λ that lie either to the right of or underneath (i,j), including (i,j) itself; explicitly,

$$H_{i,j} = \{(i,j') : j' > j\} \cup \{(i',j) : i' \geq i\}.$$

The **corner** of $H_{i,j}$ is the box in the i,j position. The **hook number** $h_{i,j}$ is the number of boxes in $H_{i,j}$; in other words, writing ν_i for the length of the i row and μ_j for the length of the j column,

$$h_{i,j} = (\nu_i - j) + (\mu_j - i) + 1.$$

For example, taking $\lambda = \{4,3,1\}$ and writing the hook number $h_{i,j}$ in the corner of the hook $H_{i,j}$ for each (i,j) yields the following array of hook numbers:

$$\begin{array}{cccc} 6 & 4 & 3 & 1 \\ 4 & 2 & 1 & \\ 1 & & & \end{array} \quad . \tag{19.13}$$

Remark 19.63. Suppose $i' \geq i$ and let $j' = \nu_{i'}$. If it happens that also $i' = \mu_{j'}$, then $h_{i,j} = h_{i,j'} + h_{i',j} - 1$. This is obvious, since one simply replaces the hook by the path going along the opposite sides of the rectangle $[i,i'] \times [j,j']$, but counting the (i',j') box twice. More explicitly,

$$\begin{aligned} h_{i,j'} + h_{i',j} &= (\nu_i - j') + (\mu_{j'} - i) + 1 + (\nu_{i'} - j) + (\mu_j - i') + 1 \\ &= (\nu_i - j) + (\mu_j - i) + 2 = h_{i,j} + 1. \end{aligned}$$

Theorem 19.64 (Frame, Robinson, and Thrall). $f^\lambda = \frac{n!}{\prod h_{i,j}}$.

For example, in (19.13), $f^\lambda = \frac{8!}{6 \cdot 4 \cdot 4 \cdot 3 \cdot 2} = 7 \cdot 5 \cdot 2 = 70$.

Proof (Greene[GreNW], Nijenhuis, and Wilf). Defining $f_\lambda = \frac{n!}{\prod h_{i,j}}$, we need to show that $f^\lambda = f_\lambda$. We proceed by induction on n, assuming that the formula holds for all partitions of $n-1$.

For any shape λ, we define a **reduced shape** for λ to be a shape obtained by erasing a box which is at the end of both a row and a column.

For example, the three reduced shapes obtained from (19.13), together with their hook numbers, would be

$$\begin{array}{|c|c|c|}\hline 5&3&2\\\hline 4&2&1\\\hline 1\\\cline{1-1}\end{array}\quad,\qquad \begin{array}{|c|c|c|c|}\hline 6&4&2&1\\\hline 3&1\\\cline{1-2} 1\\\cline{1-1}\end{array}\quad,\qquad \begin{array}{|c|c|c|c|}\hline 5&4&3&1\\\hline 3&2&1\\\cline{1-3}\end{array}\quad.$$

If T is a standard Young tableau, then the entry n must appear at the end of some row and column, and there the hook number is 1; erasing that box leaves us with a standard Young tableau on a partition of $n-1$. In our example, the number of standard tableaus for each of our reduced shapes is respectively 21, 35, and 14, and each of these comes from a different standard tableau of λ.

Conversely, for any given shape and any box at the end of a row and column, there is a standard Young tableau having n in that box. (Just remove the box, fill in a standard Young tableau for the reduced shape of $n-1$, and then redraw the box and write in n.)

Consequently, f^λ is the sum of the $f^{\lambda'}$ taken over all partitions λ' corresponding to reduced shapes for λ. But, by induction, each $f^{\lambda'} = f_{\lambda'}$. Thus, it remains to prove:

Claim. $f_\lambda = \sum f_{\lambda'}$, summed over all λ' corresponding to reduced shapes for λ. (In our example, $70 = 21 + 35 + 14$.)

Equivalently, we need to show that $\sum p_{\lambda'} = 1$, where $p_{\lambda'} = \frac{f_{\lambda'}}{f_\lambda}$. Suppose that the shape λ' is obtained by erasing the box at (v, ν_v). The only hook numbers $h_{i,j}$ that have been affected are for $i = v$ or $j = \nu_v$; these have been reduced by 1. Hence

$$\begin{aligned} p_{\lambda'} = \frac{f_{\lambda'}}{f_\lambda} &= \frac{(n-1)!}{n!} \prod_{i=1}^{v-1} \frac{h_{i,\nu_v}}{h_{i,\nu_v}-1} \prod_{j=1}^{\nu_v-1} \frac{h_{v,j}}{h_{v,j}-1} \\ &= \frac{1}{n} \prod_{i=1}^{v-1} \left(1 + \frac{1}{h_{i,\nu_v}-1}\right) \prod_{j=1}^{\nu_v-1} \left(1 + \frac{1}{h_{v,j}-1}\right) \\ &= \frac{1}{n} \left(\sum_{I,J} \underbrace{\prod_{i\in I} \left(\frac{1}{h_{i,\nu_v}-1}\right)}_{q_I} \underbrace{\prod_{j\in J} \left(\frac{1}{h_{v,j}-1}\right)}_{q'_J} \right), \end{aligned} \tag{19.14}$$

where I ranges over the subsets of $\{1, \dots, v–1\}$ and J ranges over the subsets of $\{1, \dots, \nu_v - 1\}$ (where by convention for $I = \emptyset$, we write 1 for $\prod_{i\in I} \frac{1}{h_{i,\nu_v}-1}$, and likewise for $J = \emptyset$.

To prove that $\sum p_{\lambda'} = 1$, we show that each $p_{\lambda'}$ is the probability for one of an exhaustive set of mutually exclusive occurrences. Consider the following process. We randomly pick a box (i_1, j_1) of the n boxes of the shape, and then we randomly pick another box (i_2, j_2) in the hook H_{i_1,j_1}, continuing in this way (each time picking a different box (i_{u+1}, j_{u+1}) in the hook H_{i_u,j_u} of the previous selection) until we reach the right end of some row, at which point we stop. (Note that each selection shrinks the size of the hook, so the process does terminate.)

We aim to show that the probability of finishing at (v, ν_v) is precisely $p_{\lambda'}$, where λ' is the corresponding reduced shape for λ when we erase the box (v, ν_v). To see this, we fix v (and thus ν_v) and consider the sequence of choices $(i_1, j_1), \dots, (i_t, j_t)$ that we make, where $i_t = v$ and $j_t = \nu_v$. Let $I = \{i_1, \dots, i_t\}$ and $J = \{j_1, \dots, j_t\}$, and $\mathcal{I} = \{(i_1, j_1), \dots, (i_t, j_t)\}$. Let $p(\mathcal{I}|i_1, j_1)$ denote the probability that we start at (i_1, j_1), continue with (i_2, j_2), and traverse the set $\mathcal{I}$ in reaching (v, ν_v).

Let $\tilde{p} = \prod_{i\in I} \left(\frac{1}{h_{i,\nu_v}-1}\right) \prod_{j\in J} \left(\frac{1}{h_{v,j}-1}\right)$. Since the probability of starting with any specific box (i_1, j_1) is $\frac{1}{n}$, it remains to prove that $p(\mathcal{I}|i_1, j_1) = \tilde{p}$. If $I = J = \emptyset$, then this is clear since $\tilde{p} = 1$, whereas we already started at (v, ν_v). In general, write $\mathcal{I}_1 = \mathcal{I} \setminus \{(i_1, j_1)\}$. Our second pick must be one of the $h_{i_1,j_1}-1$ boxes on the hook H_{i_1,j_1}, so $i_2 = i_1$ or $j_2 = j_1$, implying that

$$p(\mathcal{I}|i_1, j_1) = \frac{1}{h_{i_1,j_1}-1} \big(p(\mathcal{I}_1|i_2, j_1) + p(\mathcal{I}_1|i_1, j_2)\big).$$

By induction, $p(\mathcal{I}_1|i_2, j_1) = (h_{i_1,\nu_v}-1)\tilde{p}$ and $p(\mathcal{I}_1|i_1, j_2) = (h_{v,j_1}-1)\tilde{p}$. But Remark 19.63 shows that $h_{i_1,j_1}-1 = h_{i_1,\nu_v}-1 + h_{v,j_1}-1$, so we get

$$p(\mathcal{I}|i_1, j_1) = \frac{(h_{i_1,\nu_v}-1) + (h_{v,j_1}-1)}{h_{i_1,j_1}-1} \tilde{p} = \tilde{p},$$

as desired. □

(Pak [**Pak**] provides a proof based on geometric combinatorics that also yields other important information such as the Robinson-Schensted-Knuth correspondence.)

Appendix 19A. Representations of infinite groups

In this appendix, we write the identity element of a group as e (rather than as 1, which we did in the body of Chapter 19). Our main goal is to generalize the main results of Chapter 19 to infinite subgroups in $\mathrm{GL}(n, F)$. But first we recall some basic properties of finitely generated (f.g.) groups, and obtain finiteness results for related subgroups (such as the commutator subgroup G'), in preparation for later results about solvable groups. We recall that any subgroup H of finite index in a f.g. group G is f.g., by Remark 00.1. We also recall the notation of Definition 17.17 for conjugates and group commutators.

Remark 19A.1. $(a, b)^c = (a^c, b^c)$. Thus, any conjugate of a product of commutators is itself a product of commutators of the same length.

LEMMA 19A.2. *Suppose $A, B \triangleleft G$ and let $T = \{(a,b) : a \in A,\ b \in B\}$. Let $k = |T|$. If, for each $c \in T$, c^{m+1} is a product of $\leq m$ commutators in T, then $|(A,B)| \leq k^{km}$.*

Proof. For $c \in T$, $c(a,b) = (a^c, b^c)c$, which means that we can rearrange any product of commutators to move a specified commutator c to the right. Any product w of $km+1$ commutators has one commutator (a,b) occurring at least $m+1$ times; moving these occurrences to the right enables us to rewrite the commutator as $(k-1)m$ commutators times $(a,b)^{m+1}$, which, by hypothesis, is a product of $\leq m$ commutators, so we rewrite w as a product of $\leq km$ commutators in all. By induction, any element of (A,B) can be rewritten as a product of at most km commutators. Thus, the number of different possible elements in (A,B) is at most k^{km}. □

LEMMA 19A.3. *If the center Z of a group G is of finite index m in G, then the commutator subgroup G' is finite.*

Proof. Let T be the set of commutators in G. By definition, T generates G'. If $G = \bigcup_{i=1}^{m} Zg_i$, then $(z_1 g_i, z_2 g_j) = (g_i, g_j)$, so G has at most m^2 distinct commutators.

We claim that $(a,b)^{m+1}$ is a product of m commutators, for any $a, b \in G$. Indeed by Lagrange's theorem, $(a,b)^m \in Z$, so

$$(a,b)^{m+1} = b^{-1}(a,b)^m b(a,b) = b^{-1}(a,b)^m baba^{-1}b^{-1} = \big((a,b)^{m-1}(a,b^2)\big)^{b^{-1}},$$

a product of m commutators, by Remark 19A.1.

We apply Lemma 19A.2 to conclude the proof. □

(This argument shows that $|G'| \leq m^{2m^3}$, but with care one can improve the bound.)

PROPOSITION 19A.4. *If $A, B \triangleleft G$ and $T = \{(a,b) : a \in A,\ b \in B\}$ is finite, then (A,B) is finite.*

Proof. Let $L = (A,B) \triangleleft G$. Conjugation by any element of G permutes T, so, letting $n = |T|$, we get a homomorphism $\varphi\colon G \to S_n$, sending an element $g \in G$ to the map $u \mapsto u^g$ $(u \in T)$; $\ker\varphi$ is of finite index $(\leq n!)$ and centralizes L. $\ker\varphi \cap L$ is a central subgroup of L having finite index, so, by Lemma 19A.3, L' is finite as well as normal in G. Passing to G/L', we may assume that $L' = \{e\}$, so L is Abelian.

Now if $a \in A$ and $c \in L$, we thus have $(a,c) \in L$ commuting with c, so

$$(a,c)^2 = c^a c^{-1}(a,c) = c^a (a,c) c^{-1} = (a,c^2) \tag{19A.1}$$

is a single commutator. Hence (A,L) is finite by Lemma 19A.2, so passing to $G/(A,L)$, we may assume that A centralizes L. The symmetrical argument permits us to assume that B centralizes L.

Taking $a \in A$ and $b \in B$, we see that $(a,b) \in L$ commutes with b, and repeating the computation of Equation (19A.1), $(a,b)^2 = (a,b^2)$ is a single commutator; hence L is finite, as desired. □

Linear groups

We are ready to introduce the principal subject of this appendix.

Definition 19A.5. A group is called **linear** if it has a faithful f.d. representation, i.e., if it can be viewed as a subgroup of $\mathrm{GL}(n,F)$ for suitable n.

The theory of representations of infinite groups starts with linear groups. Our object here is to explore linear groups, often from a topological viewpoint. Of course, $\mathrm{GL}(n,F)$ itself is linear. Various representations of $\mathrm{GL}(n,F)$ describe some of its basic structure.

Example 19A.5′. The **adjoint representation** Ad of $\mathrm{GL}(n,F)$ on $M_n(F)$ is given by taking conjugates of matrices; namely, $\mathrm{Ad}(g)\colon a \mapsto a^g$. We write Ad_g for the transformation $\mathrm{Ad}(g)$.

This avenue is pursued in Exercises A2 and A3.

Example 19A.6. Some other famous linear groups:

(i) The **special linear group** $\mathrm{SL}(n,F)$, i.e., the set of matrices having determinant 1.

(ii) The group $\mathrm{D}(n,F)$ of diagonal matrices $\{\sum_{i=1}^{n} \alpha_i e_{ii} : \forall \alpha_i \neq 0\}$.

(iii) The group $\mathrm{T}(n,F)$ of upper triangular matrices whose diagonal entries are nonzero.

(iv) The subgroup $\mathrm{UT}(n,F)$ of **unipotent** upper triangular matrices, consisting of those matrices in $T(n,F)$ having 1 on each diagonal entry.

(v) Suppose $(*)$ is an involution on $M_n(F)$; cf. Definition 15.27. A matrix a is called $(*)$-**unitary** if $a^* = a^{-1}$, i.e., $aa^* = I$. The set G of $(*)$-unitary matrices is a group, since $(ab)(ab)^* = abb^*a^* = aa^* = I$.

In the case when $(*)$ is the transpose, G is called the **orthogonal group** $\mathrm{O}(n,F) = \{a \in \mathrm{GL}(n,F) : aa^t = I\}$.

When $F = \mathbb{C}$ and $(*)$ is the Hermitian transpose, cf. Example 14.33(iv), G is called the **unitary group** $\mathrm{U}(n,\mathbb{C})$. Note that $\mathrm{O}(n,\mathbb{R}) \subset \mathrm{U}(n,\mathbb{C})$, since the Hermitian transpose restricts to the transpose on $M_n(\mathbb{R})$.

When $n = 2m$ and (*) is the symplectic involution of $M_n(F)$, cf. Example 14.33(iii), G is called the **symplectic group** $\mathrm{Sp}(n, F)$.

(vi) The **special orthogonal group** $\mathrm{SO}(n, F) = \mathrm{O}(n, F) \cap \mathrm{SL}(n, F)$.

(vii) The **projective general linear group** $\mathrm{PGL}(n, F) = \mathrm{GL}(n, F)/F^\times$. Although not a subgroup of $\mathrm{GL}(n, F)$, $\mathrm{PGL}(n, F)$ can be viewed as a linear group via Exercise A4.

These are all examples of linear algebraic groups, to be defined in Appendix B.

Remark 19A.7. Kolchin's Theorem (Theorem 15B.4) implies that, for F algebraically closed, any unipotent subgroup of $M_n(F)$ is conjugate to a subgroup of $\mathrm{UT}(n, F)$, and in particular is a nilpotent group of nilpotence index $\leq n$.

Remark 19A.8. (i) Suppose G is a linear group, viewed as a subgroup of $\mathrm{GL}(n, F)$, and $F^{(n)}$ is simple as a G-space. Then G generates $M_n(F)$ as an F-algebra, by Corollary 15A.4. In the older terminology, G is then called **irreducible**, although this conflicts with the usage "irreducible varieties," which is needed in Appendix 19B.

(ii) Hypotheses as in (i), if N is a normal subgroup of G, then $F^{(n)}$ is a finite sum of simple N-subspaces of equal dimension over F. (Indeed, taking a simple N-subspace V, one clearly has $F^{(n)} = \sum_{g\in G} gV$, and moreover each gV is a simple N-space, since any N-subspace gU of gV yields an N-subspace U of V. But then one gets the desired assertion by taking a minimal sum.) In fact, this sum is direct, by Exercise 14.4.

Burnside's problem revisited.

Burnside's problem has a positive solution for periodic linear groups.

THEOREM 19A.9. *Suppose F is a field of characteristic* 0.

(i) *(Burnside) Every subgroup $G \subseteq \mathrm{GL}(n, F)$ of finite exponent is a finite group.*

(ii) *(Schur) Any f.g. periodic subgroup G of $\mathrm{GL}(n, F)$ is finite.*

Proof. Passing to the algebraic closure, we may assume that F is algebraically closed (of characteristic 0), and thus is a splitting field for G.

(i) Let $m = \exp(G)$. Each element g of G satisfies the polynomial $\lambda^m = 1$, so the eigenvalues of g are m-th roots of 1, and $\mathrm{tr}(g)$ is a sum of m-th roots of 1. First we consider the case where $F^{(n)}$ is simple as a G-space.

Taking $g_1, \dots, g_{n^2}$ as in Remark 19.36′ and writing an arbitrary element $g \in G$ as $g = \sum_{j=1}^{n^2} \gamma_j g_j$, one has the system of n^2 equations

$$\mathrm{tr}(g_i g) = \sum_{i,j} \gamma_j \,\mathrm{tr}(g_i g_j), \quad \text{for} \quad 1 \leq i \leq n^2. \tag{19A.2}$$

By Lemma 00.8, there is a unique solution of (19A.2) for each set of values of $\mathrm{tr}(g_i g)$. But for each $g \in G$, $\mathrm{tr}(g_i g)$ is a sum of m-roots of unity, so there are only finitely many possibilities for solving these equations, yielding finitely many solutions for the γ_j and thus for g.

In general, suppose $F^{(n)}$ has a proper G-space V; extending a base of V to a base of $F^{(n)}$ enables us to put each element of G in upper triangular form $g \mapsto \begin{pmatrix} \rho_1(g) & * \\ 0 & \rho_2(g) \end{pmatrix}$, where the representations ρ_1 and ρ_2 have degree $< n$; by induction, $\rho_1(G)$ and $\rho_2(G)$ are finite groups, so $\ker \rho_1$ and $\ker \rho_2$ have finite index, and we conclude by observing that $\ker \rho_1 \cap \ker \rho_2 = \{e\}$ by Remark 15B.2.

(ii) Let E be the (f.g.) subfield of F generated over $\mathbb{Q}$ by the entries of the generators of G. Each element g of G is defined over $M_n(E)$, so its characteristic polynomial is defined over E. On the other hand, the characteristic values of g are roots of 1, of which E has only finitely many. Hence there are only finitely many possible characteristic polynomials for the elements of G. In particular, there are only finitely many possible traces for elements of G, and thus the same proof as in the second paragraph of (i) shows that there are only finitely many possibilities for $g \in G$. □

Topological groups.

To continue our study of linear groups, we look closer at subgroups of $\mathrm{GL}(n, \mathbb{C})$. The matrices over $\mathbb{C}$ inherit a natural topological structure from $\mathbb{C}$; for example, we could define a metric, and thus a topology, on $M_n(\mathbb{C})$ by taking

$$||(\alpha_{ij})|| = \sup\{|\alpha_{ij}| : 1 \leq i, j \leq n\}. \tag{19A.3}$$

This leads us to

Definition 19A.10. A **topological group** is a group G that is also a topological space in which the multiplication map $G \times G \to G$ and the inverse map $G \to G$ given by $g \mapsto g^{-1}$ are continuous.

In particular, the left multiplication map by any element of G is continuous. We often designate a topological group in terms of its properties as a

topological space; **quasicompact** means every open cover has a finite subcover, and **compact** denotes quasicompact Hausdorff. A topological group G is called **locally compact** if every element is contained in an open set whose closure is compact (Hausdorff).

Example 19A.10′. (i) Any group is a topological group with respect to the **discrete topology**, in which each point is clopen (closed and open).

(ii) $\mathrm{GL}(n,\mathbb{C})$ is a topological group, using the topology of (19A.3), and thus its subgroups are topological with respect to the induced topology. They are locally compact but often not compact, and are used as the motivating examples in the theory.

(iii) The orthogonal group $\mathrm{O}(n,\mathbb{R})$ and unitary group $\mathrm{U}(n,\mathbb{C})$ are compact topological groups. Indeed, if $a = (a_{ij})$, then taking the i,i entry of $1 = aa^*$, we get

$$\sum_{j=1}^{n} |a_{ij}|^2 = 1,$$

implying that each $|a_{ij}| \leq 1$.

$\mathrm{U}(1,\mathbb{C})$ is just the complex unit circle S^1.

(iv) The **continuous dihedral group** consists of all rotations and reflections of the plane fixing the origin. It is a compact group, generated by S^1 and one other element b satisfying the relations $b^2 = 1$ and $bab^{-1} = a^{-1}$ for all $a \in S^1$. (Thus, taking $c = ab$ yields $c^2 = abab = aa^{-1}bb = 1$.)

(v) Any ordered group is topological, the open intervals comprising a base for the topology.

Recall that a topological space is **connected** if it is not the union of two disjoint open subsets. It is easy to see that the closure of a connected subspace is connected; also, by Zorn's lemma, any connected subspace is contained in a maximal connected subspace, called a **connected component**. Being maximal, any connected component is closed, and clearly any topological space is a disjoint union of its connected components. We write G_g for the connected component of $g \in G$. Thus, G_e is the connected component of the identity element e.

Proposition 19A.11. *G_e is a closed normal subgroup of G, the other connected components being the cosets of G_e.*

Proof. Since multiplication is continuous, $gG_e \subseteq G_g$ and $g^{-1}G_g \subseteq G_e$, implying $gG_e = G_g$.

For any $a, b \in G_e$, clearly $G_e = G_a = aG_e$, which contains ab, so $ab \in G_e$. Since $g \mapsto g^{-1}$ is continuous, $\{a^{-1} : a \in G_e\}$ is connected and contains e, implying that it is contained in G_e. Hence, G_e is a closed subgroup of G.

Finally, for each $g \in G$, $gG_eg^{-1} = G_{geg^{-1}} = G_e$; thus, $G_e \triangleleft G$. □

Proposition 19A.12. *Every open subgroup H of a quasicompact group G is closed, of finite index.*

Proof. G is the union of the cosets of H, which are disjoint. But these are open (since multiplication by an element is a homeomorphism) and thus there is a finite subcover; i.e., H has only finitely many cosets. The complement of H is the union of all cosets of H other than H itself and is open, so H is closed. □

When studying finite groups, our major tool was Maschke's Theorem, which was proved by means of the averaging procedure. Although the averaging procedure relies heavily on the finiteness of G, for (infinite) compact topological groups we can average using integration rather than sums.

Definition 19A.13. A **Haar measure** for a topological group G is a Borel measure μ (i.e., a real-valued measure defined on all countable unions and intersections of open subsets and closed subsets of G) that is invariant under the right action of G; i.e., $\mu(S) = \mu(Sh)$ for all measurable sets S and all h in G, and such that $\mu(H) < \infty$ for every compact subgroup H of G. The Haar measure is **normalized** if $\mu(G) = 1$.

Example 19A.14. If G is a finite group of order n, then G has a unique normalized Haar measure, given by $\mu(g) = \frac{1}{n}$ for all $g \in G$.

Haar [**Haa**] proved the existence of a Haar measure for any compact topological group. Von Neumann [**VN, pp. 33–40**] (really a reprint of his 1940-1941 lectures at the Institute for Advanced Study) presented a readable exposition of this fact, proving that every locally compact topological group G has a Haar measure. Given this result, and inspired by Example 19A.14, it is not difficult to generalize Maschke's Theorem to **continuous** f.d. representations of compact topological groups, i.e., group homomorphisms $G \to \mathrm{GL}(n,\mathbb{C})$ that are continuous as maps of topological spaces. We say that a G-space V is **topological** if V is endowed with a topology for which the action $G \times V \to V$ is continuous; these are the G-spaces corresponding to continuous representations. Thus, the key is the following modification of Lemma 19.25.

LEMMA 19A.15. *Suppose V_1 and V_2 are compact G-spaces over a compact (Hausdorff) topological group G, and $\psi\colon V_1 \to V_2$ is an arbitrary linear transformation over F. Then we have an $F[G]$-module map $\overline{\psi}\colon V_1 \to V_2$ defined by*

$$\overline{\psi}(v) = \int_{g\in G} g^{-1}\psi(gv)\nu(dg), \qquad \forall v \in V_1,$$

where ν is a Haar measure of G.

Proof. $\overline{\psi}$ is an F-linear transformation. Furthermore, for any h in G,

$$\begin{aligned}\overline{\psi}(hv) &= \int_{g\in G} g^{-1}\psi(g(hv))\nu(dg)\\ &= \int_{g\in G} g^{-1}\psi((gh)v)\nu(dg)\\ &= \int_{gh\in G} h(gh)^{-1}\psi((gh)v)\nu(d(gh))\\ &= \int_{g\in G} hg^{-1}\psi(gv)\nu(dg)\\ &= h\overline{\psi}(v).\end{aligned}$$

This proves that $\overline{\psi}\colon V_1 \to V_2$ preserves the G-space structure and thus is an $F[G]$-module map. □

PROPOSITION 19A.16. *Every continuous f.d. representation of a compact (Hausdorff) group is a finite direct sum of irreducible continuous representations.*

Proposition 19A.16 follows easily from Lemma 19A.15. Vinberg [**Vinb, pp. 27–29**] *gives a direct proof, outlined in Exercises A12–A18, by extracting the idea behind the proof of the existence of a Haar measure.*

Lie groups.

The remainder of this appendix can be considered as a bridge to Appendix 19B and Chapter 21.

Let us refine the notion of topological group even further. If $S \subset G$ is a neighborhood of some element g, then $g^{-1}S$ is a neighborhood of the identity element e, homeomorphic to S, so the local structure of G can be determined in terms of the neighborhoods of e.

Definition 19A.17. (Since we are introducing differential structures, we assume that the base field F is $\mathbb{R}$ or $\mathbb{C}$.) A **Lie group** is a topological linear group having a neighborhood of e diffeomorphic to Euclidean space, such

that the group operations $(g,h) \mapsto gh$ and $g \mapsto g^{-1}$ are analytic (resp. C^∞) when the base field is $\mathbb{C}$ (resp. $\mathbb{R}$).

A **Lie homomorphism** of Lie groups is a group homomorphism that also is a differential map. A **Lie subgroup** of a Lie group is a subgroup that is also a submanifold.

Remark 19A.18. Any differential map $\phi\colon G \to H$ of differential manifolds gives rise to the **tangent map** $d_g\phi\colon D_gG \to D_{\phi(g)}H$ of the tangent spaces, for each point $g \in G$. (This is obtained by differentiating the coordinate maps of ϕ.)

Now we bring in the group structure. We assume that G and H are Lie groups and that ϕ is a Lie homomorphism. Then $\phi(ag) = \phi(a)\phi(g)$ implies that the left multiplication map ℓ_a satisfies

$$\phi\ell_a = \ell_{\phi(a)}\phi,$$

and using the chain rule, we get:

$$d_a\phi\, d_e\ell_a = \ell_{\phi(a)}\, d_e\phi.$$

Thus, for any ξ in D_eG, writing $a\xi$ for $d_e\ell_a(\xi)$, we get

$$d_a\phi(a\xi) = \phi(a)\, d_e\phi(\xi). \tag{19A.4}$$

In particular, Equation (19A.4) enables us to recover d_a from d_e. We can push this further. For the remainder of this discussion, let $\mathfrak{I}$ denote the closed unit interval $[0,1]$.

PROPOSITION 19A.19. *Suppose G and H are Lie groups and G is connected. Any Lie homomorphism $\phi\colon G \to H$ is uniquely determined by its tangent map $d_e\phi$.*

Proof. Since G is connected, for any $g \in G$ there is a differential curve $f\colon \mathfrak{I} \to G$ given by $f(0) = e$ and $f(1) = g$. Then the tangent vector $f'(t) \in D_{f(t)}G$, and since $\ell_{f(t)}$ is a diffeomorphism, $f'(t) = f(t)\xi_t$ for some $\xi_t \in D_eG$. Putting $q(t) = \phi(f(t))$, we have its tangent $q'(t) = d_{f(t)}\phi(f'(t)) = d_{f(t)}\phi(f(t)\xi_t)$; hence, Equation (19A.4) (with $a = f(t)$ and $\xi = \xi_t$) yields

$$q'(t) = q(t)\, d_e\phi(\xi_t),$$

which could be viewed as a system of differential equations for the coordinates of q. Since $q(0) = \phi(e) = e$, this has a unique solution, so $\phi(g) = q(1)$ is uniquely determined. □

COROLLARY 19A.20. *Any continuous representation $\rho\colon G \to \mathrm{GL}(n,F)$ of a Lie group is determined by the tangent map $d_e\rho\colon D_eG \to D_e(\mathrm{GL}(n,F))$.*

This is a powerful result, but it needs to be interpreted. First of all, $D_e\,\mathrm{GL}(n,F)$ can be identified with $M_n(F)$, because a vector space is its own tangent space and an open subspace has the same tangent space. Let us compute an example.

Example 19A.21. The tangent representation for the adjoint representation Ad of Example 19A.5′. We take any differentiable curve $f\colon \mathfrak{I} \to G$ with $f(0) = e$. Let $f'(0) = \xi$. Differentiating $f(t)f(t)^{-1} = e$, we get

$$f(t)\frac{d}{dt}\left(f(t)^{-1}\right) + f(t)^{-1}\frac{d}{dt}f(t) = 0,$$

so evaluating at $t = 0$ yields $\frac{d}{dt}f(t)^{-1}|_{t=0} = -\frac{d}{dt}f(t)|_{t=0} = -\xi$. Hence, for $a \in M_n(F)$,

$$d_e\,\mathrm{Ad}_\xi(a) = \frac{d}{dt}\,\mathrm{Ad}_{f(t)}(a)\Big|_{t=0} = \frac{d}{dt}(f(t)af(t)^{-1})\Big|_{t=0} = \xi a - a\xi.$$

In other words, defining $\mathrm{ad}_b\colon a \mapsto ba - ab$, we see that

$$d_e\,\mathrm{Ad}_b = \mathrm{ad}_b\,.$$

We still lack an algebraic structure for D_eG on which to frame the theory. This could be built directly on D_eG; cf. Exercise A23. But there is a shortcut.

Remark 19A.22. If the Lie group G has a faithful continuous representation $\rho\colon G \to \mathrm{GL}(n,F)$, then, using Example 19A.21, we can transport the structure from $\mathrm{GL}(n,F)$ to G, defining the operation

$$[ba] = d_e\,\mathrm{Ad}_{\rho(b)}(\rho(a)) = \mathrm{ad}_{\rho(b)}(\rho(a)) = \rho(b)\rho(a) - \rho(a)\rho(b).$$

The ensuing algebra, called a **Lie algebra**, is one of the most fundamental structures in algebra, which we study in depth in Chapter 21; see also Appendix 21A for a development in terms of algebraic groups.

Another infinite group of considerable interest is the **braid group**, which arises in several different contexts and is treated in Exercises 24ff.

Appendix 19B: Algebraic groups

In Appendix A we studied linear Lie groups in terms of the differential structure of $\mathbb{C}$ and $\mathbb{R}$. We can free ourselves from these fields (and from the theory of differential manifolds), by turning to varieties as studied in Chapter 10 of Volume 1. What emerges is a vibrant topic drawing on algebra, analysis, geometry, and topology.

Note 19B.1. A slight problem with terminology: In Volume 1, we required affine varieties (and subvarieties) to be irreducible. Here it is more convenient to define an affine variety to be a closed (but not necessarily irreducible) subset of $\mathbf{A}^{(n)}$; then we may include the orthogonal groups, which are reducible. But anyway we reduce quickly to irreducible varieties in Proposition 19B.7.

Definition 19B.2. An **algebraic group** is a variety G endowed with a group structure $(G, \cdot, e)$ such that the inverse map (given by $g \mapsto g^{-1}$) and multiplication map $G \times G \to G$ (given by $(a,b) \mapsto a \cdot b$) are morphisms of varieties. A **morphism** $\varphi\colon G \to H$ of algebraic groups is a group homomorphism that is also a morphism of varieties.

The algebraic group G is **affine** when its underlying variety is affine.

In particular, for any $a \in G$ the maps $g \mapsto ag$ and $g \mapsto ga$ are automorphisms of G to itself as a variety, and we can view G as a topological group with respect to the Zariski topology; cf. Definition 19A.10. Thus, we can draw on results of Appendix 19A.

Subgroups of algebraic groups that are also closed subvarieties are themselves algebraic groups. Thus, the kernel of a morphism of algebraic groups, being the inverse image of a point, is itself an algebraic group.

The study of algebraic groups has developed apace with the development of algebraic geometry; the structure of algebraic varieties replaces the analytic hypotheses on Lie groups. (One key step is given in Exercise B1.)

Our main objective in this appendix is to use algebraic techniques to obtain some basic results of algebraic groups, including the celebrated Tits alternative. Results about solvable and nilpotent algebraic groups are given in Exercises B15ff. We do not consider certain basic issues such as the algebraic group structure of G/N for a closed normal subgroup N of G, and the "unipotent radical."

Let us start by showing that our earlier examples of Lie groups are algebraic groups.

Example 19B.3. (i) The base field F itself can be viewed as a group under $+$, called the **additive group** of F, whose coordinate algebra is $F[\lambda]$.

(ii) In Example 10.2(vii) of Volume 1, $\mathrm{GL}(n,F)$ was identified with an affine subvariety of $F^{(n+1)^2}$, with coordinate algebra $F[d,\lambda_{ij}\colon 1\le i,j\le n]$, where $d=\det(\lambda_{ij})^{-1}$. Thus, $\mathrm{GL}(n,F)$ is itself an affine algebraic group.

(iii) For $n=1$ in (ii), we get the **multiplicative group** $F^\times$ of F whose coordinate algebra is $F[\lambda,\lambda^{-1}]$.

(iv) $\mathrm{SL}(n,F)$ is a closed subvariety of $\mathrm{GL}(n,F)$ and thus is an affine algebraic group. Likewise, the other groups in Example 19A.6 also are displayed as varieties and thus are algebraic.

(v) Any elliptic curve (Chapter 11 of Volume 1) is a variety whose group structure is given in Definition 11.8.

Our first main goal is to prove that every affine algebraic group G is linear. We study G by means of the Zariski topology, with respect to which G is quasicompact; cf. Corollary 10.23 of Volume 1. Recall that G_e denotes the connected component of the identity element e. By Proposition 19A.11, G_e is closed and thus a subvariety of G.

Proposition 19B.4. *Suppose G is an algebraic group.*
(i) *Any open subgroup of G is closed of finite index.*
(ii) *Any closed subgroup H of G of finite index is open.*
(iii) *G_e is clopen of finite index in G.*

Proof. (i) By Proposition 19A.12.

(ii) The complement of H is the union of the finitely many cosets of H other than H itself and thus is closed, so H is open.

(iii) G, being a variety, has only finitely many connected components. In view of Proposition 19A.11, G_e is closed with only finitely many cosets, and thus is of finite index. Hence, G_e is open, by (ii). □

Corollary 19B.5. *Any closed subgroup H of an algebraic group G of finite index contains G_e.*

Proof. The connected component H_e in H is closed, of finite index in H and thus in G; hence, H_e is clopen, implying that $H_e=G_e$ since G_e is connected. □

Patently, any irreducible component of an algebraic group is connected. We aim for the converse.

Remark 19B.6. For any irreducible subset S of an algebraic group G, Sg and gS are irreducible sets for each $g\in G$, since left and right multiplication by g are automorphisms of the variety. It follows that if S is an irreducible component of G, then so are Sg and gS; in particular, all irreducible components have the same dimension.

Proposition 19B.7. *Every element $g\in G$ is in a unique irreducible component of G.*

Proof. Let $C_1,\dots,C_m$ denote the irreducible components of G. Then $\bigcup_{i\ne j}C_i\cap C_j$ has smaller dimension, so some $g_0\in G$ is not in the intersection of any two distinct maximal irreducible components. Now, if $g\in C_i\cap C_j$, then $g_0=(g_0g^{-1})g\in g_0g^{-1}C_i\cap g_0g^{-1}C_j$, a contradiction in view of Remark 19B.6. □

Corollary 19B.8. *G_e is also the irreducible component of G at e.*

Proof. Let H denote this irreducible component; then $H\subseteq G_e$. But for any $h\in H$, Hh^{-1} is an irreducible component containing e, so $H=Hh^{-1}$. Hence $h_1h_2^{-1}\in H$ for all $h_1,h_2\in H$, implying that H is a subgroup of G_e. But the same argument shows that each of the finitely many irreducible components of G_e is a coset of H, so, in view of Proposition 19B.7, G_e is a finite disjoint union of cosets of H, each of which are clopen, implying that $H=G_e$ since G_e is connected. □

Example 19B.9. An algebraic group G is connected iff $G=G_e$, iff G is irreducible. Of the algebraic groups given in Example 19A.6, $\mathrm{D}(n,F)$, $\mathrm{T}(n,F)$, and $\mathrm{UT}(n,F)$ are easily seen to be connected. $\mathrm{O}(n,F)$ is not connected when $\mathrm{char}(F)\ne 2$, since $\mathrm{O}(n,F)_e=\mathrm{SO}(n,F)$; cf. Exercise B4.

Here is a crucial but easy result.

Lemma 19B.10. *If U and V are open subsets of an algebraic group G with U dense, then $UV=G$.*

Proof. Take $g\in G$. Since gV^{-1} is open we have some $a\in U\cap gV^{-1}$; writing $a=gb^{-1}$ we see that $g=ab\in UV$. □

Proposition 19B.11. *Suppose $H\le G$. Then $\overline{H}\le G$; furthermore, if H contains a nonempty open subset U of $\overline{H}$, then H is closed.*

Proof. For any $a\in H$ we see that $a\overline{H}=\overline{aH}=\overline{H}$; thus $H\overline{H}\subseteq\overline{H}$. But then for any $b\in\overline{H}$ we have $\overline{H}b=\overline{Hb}\subseteq\overline{H}$, proving that $\overline{H}\le G$.

For the second assertion, $H=\cup\{aU: a\in H\}$ is open in $\overline{H}$, implying by Lemma 19B.10 that $\overline{H}=HH=H$. □

Corollary 19B.12. *If $\varphi\colon G\to H$ is a morphism of affine algebraic groups, then $\varphi(G)$ is a closed subgroup of H and $\varphi(G)_e=\varphi(G_e)$.*

Proof. $\varphi(G)$ contains a nonempty open subset of its closure in H, by Theorem 10.36 of Volume 1, and thus is closed by Proposition 19B.11. But likewise $\varphi(G_e)$ is a closed subgroup, of finite index in $\varphi(G)$, and therefore contains $\varphi(G)_e$ by Corollary 19B.8; the reverse inclusion is clear. □

Comultiplication in algebraic groups.

For convenience, we assume for the remainder of this discussion that the base field F is algebraically closed, although many of the results can be extended to arbitrary fields by passing to the algebraic closure.

The study of an algebraic group G relies heavily on the properties of G as an algebraic variety and its commutative coordinate algebra $F[G]$ (notation as in Definition 10.17 of Volume 1). Unfortunately, throughout this chapter $F[G]$ has denoted the group algebra, which usually is a different concept altogether; for example, the group algebra usually is **not** commutative. In this discussion, to avoid ambiguity, we utilize a different notation. $\mathcal{A}$ will always denote the coordinate algebra of the variety G.

Remark 19B.13. The group structure of an affine algebraic group G can be transferred to its coordinate algebra $\mathcal{A}$. First, as in Example 10.9 of Volume 1, we can identify the points of an affine algebraic group G with the maximal ideals of $\mathcal{A}$. In this way, the identity element e corresponds to some map $\epsilon\colon \mathcal{A} \to F$, given by $\epsilon(f) = f(e)$, and the group inverse to a map $S\colon \mathcal{A} \to \mathcal{A}$ given by $S(f)(g) = f(g^{-1})$. (Note that $S^2(f) = f$.)

To translate multiplication from G to $\mathcal{A}$, recall from Remark 10.26 of Volume 1 that $\mathcal{A}$ can be viewed as polynomials defined on G. Given any $f \in \mathcal{A}$, we can define $\tilde{f}\colon G \times G \to F$ by $\tilde{f}(g_1, g_2) = f(g_1 g_2)$. Proposition 18.42 enables us to view $\tilde{f}$ as an element of $\mathcal{A} \otimes \mathcal{A}$, which we denote as $\Delta(f)$. Thus, we have a map $\Delta\colon \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ satisfying the following remarkable property:

If $\Delta(f) = \sum f_{i1} \otimes f_{i2} \in \mathcal{A} \otimes \mathcal{A}$, then

$$f(g_1 g_2) = \Delta(f)(g_1, g_2) = \sum f_{i1}(g_1) f_{i2}(g_2) \tag{19B.1}$$

for all $g_1, g_2 \in G$.

Definition 19B.14. In the setup of Remark 19B.13, ϵ is called the **counit**, Δ is called **comultiplication**, and S is called the **antipode**. This terminology anticipates the theory of Hopf algebras given in Chapter 26.

Example 19B.15. (i) $G = \mathbb{G}_{\mathbf{a}}$. Here $\mathcal{A} = F[\lambda]$ and $\epsilon(\lambda) = 0$.

$$\Delta(\lambda)(g_1, g_2) = \lambda(g_1 + g_2) = \lambda g_1 + \lambda g_2,$$

yielding $\Delta(\lambda) = \lambda \otimes 1 + 1 \otimes \lambda$. $S(\lambda)(g) = \lambda(-g) = -\lambda(g)$, so $S(\lambda) = -\lambda$.

(ii) $G = \mathrm{GL}(n, F)$ whose coordinate algebra $F[d, \lambda_{ij} : 1 \le i, j \le n]$ is described in Example 19B.3(ii). Here $\epsilon(\lambda_{ij}) = \delta_{ij}$. $\Delta(\lambda_{ij})$ takes the i, j position of the matrix product, and thus is $\sum_{k=1}^{n} \lambda_{ik} \otimes \lambda_{kj}$. $S(\lambda_{ij})$ yields the i, j position of the inverse matrix, and thus is d times the adjoint matrix of (λ_{ij}).

Let us examine the role of the left multiplication map. Given any $a \in G$, define $\ell_a\colon G \to G$ by $\ell_a(g) = ag$; also define $\hat{\ell}_a\colon \mathcal{A} \to \mathcal{A}$ by $\hat{\ell}_a(f) = f\ell_{a^{-1}}$, i.e.,

$$\hat{\ell}_a(f)(g) = f(a^{-1}g). \tag{19B.2}$$

Thus, $\hat{\ell}$ defines a representation $\hat{\ell}\colon G \to \mathrm{GL}(\mathcal{A})$, by which we view $\mathcal{A}$ as a G-space. (Although the use of the inverse in this definition complicates the proof of Theorem 19B.19 below, it is standard in the literature.)

Remark 19B.16. If $\Delta(v) = \sum f_i \otimes v_i$ for $v, f_i, v_i \in \mathcal{A}$ then, for all $a, g \in G$,

$$\hat{\ell}_a v(g) = v(a^{-1}g) = \Delta(v)(a^{-1}, g) = \sum f_i(a^{-1}) v_i(g); \tag{19B.3}$$

i.e., $\hat{\ell}_a v = \sum f_i(a^{-1}) v_i$.

LEMMA 19B.17. *A subspace V of $\mathcal{A}$ is a G-subspace (under the action $gv = \hat{\ell}_g(v)$) iff $\Delta(V) \subseteq \mathcal{A} \otimes_F V$.*

Proof. Let W be a complementary space to V in $\mathcal{A}$. Then one can write

$$\Delta(v) = \sum f_i \otimes v_i + \sum h_j \otimes w_j$$

for $f_i, h_j \in \mathcal{A}$, $v_i \in V$, and F-independent $w_j \in W$. By Remark 19B.16, for any $a \in G$,

$$\hat{\ell}_a v = \sum f_i(a^{-1}) v_i + \sum h_j(a^{-1}) w_j.$$

Thus, $Gv \in V$ iff each $h_j(a^{-1}) = 0$ for all $a \in G$, i.e., iff each $h_j = 0$. □

Affine algebraic groups are linear.

$\mathcal{A}$ still denotes the coordinate algebra of the affine algebraic group G.

LEMMA 19B.18. *Any f.d. subspace of $\mathcal{A}$ is contained in a f.d. G-subspace of $\mathcal{A}$.*

Proof. Given a f.d. subspace $V = \sum_{j=1}^m Fv_j \subset A$, put

$$\Delta v_j = \sum_{i=1}^t f_{ij} \otimes v_{ij},$$

for $f_{ij}, v_{ij} \in \mathcal{A}$. Let $W = \sum_{i=1}^t \sum_{j=1}^m Fv_{ij}$, which by inspection is f.d. over F. By Remark 19B.16, $av_j = \hat{\ell}_a v_j = \sum f_{ij}(a^{-1})v_{ij} \in W$ for all $a \in G$, $1 \le j \le m$, implying that $Gv_j \subseteq W$, so $\sum_{j=1}^m Gv_j$ is a G-subspace that is a subspace of W and thus f.d. □

We are ready for a major result about affine algebraic groups.

THEOREM 19B.19. *Every affine algebraic group G is linear, i.e., isomorphic to a subgroup of* $\mathrm{GL}(n, F)$.

Proof. Write the coordinate algebra of G as $\mathcal{A} = F[v_1, \dots, v_n]$ and let $V = \bigoplus_{i=1}^n Fv_i$. Expanding the generating set v_i if necessary, we may assume by Lemma 19B.18 that V is a G-space. Thus, by Lemma 19B.17, we may write

$$\Delta v_i = \sum_{j=1}^n f_{ij} \otimes v_j,$$

for suitable $f_{ij} \in \mathcal{A}$; hence, by Remark 19B.16,

$$\hat{\ell}_a v_i = \sum f_{ij}(a^{-1})v_j. \tag{19B.4}$$

Define $\rho\colon G \to \mathrm{GL}(n, F)$ by taking $\rho(a)$ to be the matrix $(f_{ij}(a))$ acting on the base $\{v_1, \dots, v_n\}$; in other words,

$$\rho(a)(v_i) = \sum_j f_{ij}(a)v_j = \hat{\ell}_{a^{-1}} v_i. \tag{19B.5}$$

Then ρ is a group homomorphism , since for $a, b, g \in G$,

$$\rho(a)\rho(b)(v_i)(g) = \rho(a)\hat{\ell}_{b^{-1}}(v_i)(g) = \rho(a)(v_i(bg)) = v_i(abg) = \rho(ab)(v_i)(g);$$

i.e., $\rho(a)\rho(b) = \rho(ab)$. Moreover, hence ρ is 1:1; indeed, if $a \in \ker \rho$, then (19B.5) shows that $v_i = \hat{\ell}_{a^{-1}} v_i$ for all i, implying $a^{-1} = e$. □

Remark 19B.20. In the proof of Theorem 19B.19, ρ is a morphism of varieties and thus of algebraic groups. Indeed, denote the coordinate algebra of $\mathrm{GL}(n, F)$ as $\tilde{\mathcal{A}} = F[\lambda_{ij}, d : 1 \le i, j \le n]$. In view of Remark 10.35 of Volume 1, we want to show that the map $\tilde{\rho} : \tilde{\mathcal{A}} \to \mathcal{A}$ corresponding to ρ is an algebra homomorphism. But $\tilde{\rho}$ is given by the substitution map $\lambda_{ij} \mapsto f_{ij}$, $1 \le i, j \le n$.

It follows by Corollary 19B.10 that $\rho(G)$ is a closed subgroup of $\mathrm{GL}(n, F)$. In conclusion, we have displayed an arbitrary affine algebraic group as a closed algebraic subgroup of $\mathrm{GL}(n, F)$ for suitable n.

The Tits alternative

We turn to one of the most oft-quoted theorems about linear groups, due to Tits [**Tit**]. Recall that a group is called **virtually solvable** if it has a solvable subgroup of finite index.

THEOREM 19B.21 (THE TITS ALTERNATIVE). *Every f.g. linear group either is virtually solvable or contains a free subgroup.*

(In fact, Tits proved in characteristic 0 that every linear group satisfies the Tits alternative, but this fails in nonzero characteristic; cf. Exercise B5.) We already recorded an application in Theorem 17.66, that every f.g. linear group has either polynomial growth or exponential growth.

Concerning the proof, the free subgroup is generated using the pingpong lemma (Lemma 17.16). Dixon [**Di**] found a reasonably elementary modification of Tits' proof, which we provide now, after a bit of preparation.

Remark 19B.22. If $N \lhd G$ and Na, Nb generate a free subgroup of G/N, then a, b generate a free subgroup of G.

Thus, in verifying Tits' alternative, we may pass to homomorphic images modulo normal solvable subgroups.

Another ingredient is the exterior algebra $E(V)$ of Example 18.40(iv).

Remark 19B.23. Suppose $\rho\colon G \to \mathrm{GL}(V)$ is a group representation, where $V = F^{(n)}$. Then G acts naturally on $E(V)$ by

$$g(v_1 \wedge \cdots \wedge v_m) = \rho(g)v_1 \wedge \cdots \wedge \rho(g)v_m.$$

(We use the wedge product notation here to avoid ambiguity.)

Let $E_k(V)$ denote the subspace of $E(V)$ spanned by all words of length k. If $\{e_1, \dots, e_n\}$ is a base of V, then $\{e_{i_1} \wedge \cdots \wedge e_{i_k} : 1 \le i_1 < i_2 < \cdots < i_k \le n\}$ is a base of $E_k(V)$, which thus has dimension $\binom{n}{k}$. Hence, the restriction of the action of G to $E_k(V)$ yields a representation $\hat{\rho}_k$ of degree $\binom{n}{k}$.

We also want to apply the theory of algebraic groups, but with care, since finitely generated groups usually are not closed subvarieties of $\mathrm{GL}(n, F)$. (In fact, by an obvious counting argument, for $F = \mathbb{C}$ the only f.g. algebraic subgroups are finite!)

We are ready for the bulk of the proof of the Tits alternative in characteristic 0, leaving one computation (in Step V) to the reader. (The case for $\mathrm{char}(F) > 0$ is similar, as indicated in Exercise B7.) Suppose $\rho\colon G \to \mathrm{GL}(n, F)$ is a faithful representation for F algebraically closed. Let $V = F^{(n)}$, viewed naturally as a G-space, or equivalently, as an $F[G]$-module.

STEP I. V has a composition series $V = V_0 \supset V_1 \supset \cdots \supset V_t = 0$ of length $\leq n$, and, by definition, G acts naturally on each V_i/V_{i+1}. Let G_i denote the natural homomorphic image of G in $\mathrm{End}(V_i/V_{i+1})$. If each G_i is solvable, then G is solvable, so we are done unless some G_i is not solvable. In view of Remark 19B.22, it suffices to show that G_i contains a free subgroup; thus, replacing G by G_i, and V by V_i/V_{i+1}, we may assume that V is simple as a G-space. Thus, the faithful representation ρ is irreducible.

STEP II. We claim that G contains an element a such that some eigenvalue of $\rho(a)$ is not a root of 1. On the contrary, assume that every characteristic value of each element a of G is a root of 1. Then some power of a is unipotent. In fact, arguing as in Proposition 00.7, we claim that only finitely many roots of 1 can appear as eigenvalues of elements of G. To see this, first note that any element a satisfies its Hamilton-Cayley polynomial f_a of degree n. Since f_a is integral, the coefficient of λ^{n-i} for each $0 < i \leq n$, is the sum of $\binom{n}{i}$ products of characteristic values of a (which are roots of 1), and thus has absolute value $\leq 2^n$. This gives only finitely many possibilities for each coefficient, and thus for f_a, and thus for the characteristic values of a.

Now, the same argument as given in the last paragraph of the proof of Theorem 15B.4 shows that G is finite; namely, the elements of G are obtained by solving a simultaneous set of linear equations, of which there are only finitely many possibilities; each has a unique solution, by Lemma 00.8.

STEP III. We claim that there is an absolute value $|\ |$ on F, for which $\rho(a)$ has an eigenvalue α with absolute value $\neq 1$. This is seen by taking α not a root of 1, by Step II, extending absolute values to f.g. field extensions, and then passing to the algebraic closure by means of Theorem 12.43 of Volume 1 (and the subsequent discussion).

Namely, let F_1 be the subfield of F generated over $\mathbb{Q}$ by all entries of the generators of G (viewed as matrices). Clearly we may view $G \subseteq \mathrm{GL}(n, F_1)$. Also, we may replace F by the algebraic closure of F_1, so it suffices to prove the assertion of this step for F_1. Taking a transcendence base B of F_1 over F_0, we write F_1 as a finite extension of a purely transcendental field extension $K = F_0(B)$ of F_0. There are two cases to consider.

Case 1. α is transcendental over $\mathbb{Q}$. Hence, we may take our transcendence base B to contain α. We extend the α-adic absolute value (on $\mathbb{Q}(\alpha)$) naturally to a valuation of K, and then of F_1, and finally to the algebraic closure F.

Case 2. α is algebraic over $\mathbb{Q}$. Take the minimal polynomial f of α over $\mathbb{Z}$. If f is not monic, then some prime p divides the leading coefficient of f, and a is not integral over the local ring $\mathbb{Z}_p$. Hence, by Theorem 12.45

of Volume 1, there is some valuation ring V lying above C not containing a; i.e., $|\alpha| < 1$ under the corresponding absolute value.

On the other hand, for f monic, α is integral over $\mathbb{Z}$; hence, by Proposition 00.7, after an automorphism, we have $|\alpha| \neq 1$.

Having proved Step III, we note that by passing to the completion of F, we may assume that F is complete as well as algebraically closed (by Kurschak's Theorem, which is Exercise 12.28 of Volume 1). We say that $\rho(a)$ has a **dominant eigenvalue** (with respect to a given absolute value on F) if one of the eigenvalues of $\rho(a)$ has value greater than each of the others, and its eigenspace is one-dimensional.

STEP IV. There is a (different) representation $\hat{\rho}$ of G such that $\rho(a)$ has a dominant eigenvalue. Indeed, assume that $\rho(a)$ has eigenvectors $v_1, \ldots, v_n$, with respective eigenvalues $\alpha_1, \ldots, \alpha_n$, and arrange the v_i such that the $|\alpha_i|$ are in descending order; i.e., for suitable k,

$$|a_1| = |\alpha_2| = \cdots = |\alpha_k| > |a_{k+1}| \geq \cdots \geq |\alpha_n|.$$

Take $\hat{\rho}$ as in Remark 19B.23; the eigenvectors of $\hat{\rho}_k(a)$ are clearly the $v_{i_1} \wedge \cdots \wedge v_{i_k}$, and the eigenvector $v_1 \wedge \cdots \wedge v_k$ has eigenvalue $\alpha = \alpha_1 \cdots \alpha_k$, which is the single eigenvalue of greatest absolute value (namely $|\alpha_1|^k$). Thus, we replace ρ by $\hat{\rho}_k$, which has the dominant eigenvalue α. Although $\hat{\rho}_k$ may be reducible, we could replace it by the irreducible component for which α remains the dominant eigenvalue; note that if $V = W_1 \oplus W_2$ as G-spaces and $gv = \alpha v$, then writing $v = w_1 + w_2$ we have

$$gw_1 + gw_2 = g(w_1 + w_2) = gv = \alpha v = \alpha w_1 + \alpha w_2,$$

so both $w_1 \in W_1$ and $w_2 \in W_2$ are eigenvectors with respect to α.

STEP V. One can further modify a such that both $\rho(a)$ and $\rho(a)^{-1}$ have dominant eigenvalues; this is an elaborate but elementary argument described in Exercise B6, based on conjugating by high enough powers of a such that the dominant eigenvalue is much larger than all other eigenvalues. (This is also contained in [**Tit, Proposition 3.11**].)

STEP VI. Let us construct a free subgroup of rank 2. We aim to find a conjugate gag^{-1} of a and an integer $m > 0$ such that a^m and $ga^m g^{-1}$ satisfy the hypotheses of the pingpong lemma (Lemma 17.16) and thus generate a free group.

To unify notation, write $a_1 = a$ and $a_2 = a^{-1}$. Let α_i be the dominant eigenvalue of a_i, and let v_i be the eigenvectors for $i = 1, 2$. Thus, $a_i v_i = \alpha_i v_i$. Writing $V_i = Fv_i$ for $i = 1, 2$, take an a-invariant complement V_3 of $V_1 \oplus V_2$.

Let $\overline{G}$ denote the closure of G in $M_n(F)$ with respect to the Zariski topology. Then $\overline{G}$ is an algebraic group (although not necessarily f.g.). We

introduce the notation

$$H(S,T) = \{g \in \overline{G} : gS \cap T \neq 0\}$$

for any subspaces S, T of $F^{(n)}$.

In other words, considering transformations $\varphi_g \colon S \oplus T \to V$ given by $(s,t) \mapsto gs - t$, we see that

$$H(S,T) = \{g \in \overline{G} : \varphi_g \text{ does not have full rank}\},$$

so $H(S,T)$ is Zariski closed, of smaller dimension than $\overline{G}$ when $S \cap T = 0$, since then $e \notin H(S,T)$. Also $\dim H(V_i, V_i) < \dim \overline{G}$ for $i = 1, 2$, since the representation ρ is irreducible. Since $\dim V_i = 1$, there is some $c \in G$ missing $H(V_1, V_1)$, $H(V_2, V_2)$ and $H(V_i + V_j, V_k)$ and $H(V_i, V_j + V_k)$, for all permutations (i, j, k) of $(1, 2, 3)$. In particular,

$$(cV_1 \cup c^{-1}V_1) \cap (V_1 \cup (V_2 + V_3)) = 0; \quad (cV_2 \cup c^{-1}V_2) \cap (V_2 \cup (V_1 + V_3)) = 0.$$

Let $S = V_3 \cup c(V_2+V_3) \cup c^{-1}(V_2+V_3) \cup c(V_2+V_3) \cup c^{-1}(V_1+V_3)$; we see that $\{v_1, v_2\} \cap S = \emptyset$. Now we work with the topology induced from the absolute value on F (obtained in Step III). Since every subspace of $F^{(n)}$ is closed in the product topology, there is a compact neighborhood N_i of $\{v_i\}$, $i = 1, 2$, with $N_i \cap S = \emptyset$. Let $N_0 = N_1 \cup N_2$ and $W_0 = \{\alpha v : \alpha \in F, v \in N_0\}$, which is closed in this induced topology; clearly, $W_0 \cap S = \emptyset$, so $W_0 \cap \{cv_1, cv_2\} = \emptyset$. Hence N_0 contains a compact neighborhood N of $\{v_1, v_2\}$ with $W_0 \cap cN = \emptyset$. Put

$$\Gamma = \{\alpha r : 0 \neq \alpha \in F,\ r \in N\}.$$

Clearly $\Gamma \cap c\Gamma \subseteq W_0 \cap cN = \emptyset$. Now let $\tilde{N} = cN \cup c^{-1}N$, which is compact and disjoint from $V_i \oplus V_3$ for $i = 1, 2$ (since $N \cap S = \emptyset$). Take any $v \in \tilde{N}$. Writing $v = \beta_1 v_1 + \beta_2 v_2 + v_3$ for $\beta_i \in K$ and $v_3 \in V_3$, and recalling that α_i is the dominant eigenvalue for a_i, we see that

$$\lim_{k \to \infty} \alpha_i^{-k} a_i^k v = \beta_i v_i, \qquad i = 1, 2.$$

Γ contains $\beta_1 v_1$ and $\beta_2 v_2$; hence, there is k such that $\alpha_i^{-m} a_i^m v \in \Gamma$ for all $m > k$. It follows that $a^{ku}\tilde{N} \subseteq \Gamma$ for all $u \in \mathbb{Z}$; thus, $a^{ku}c\Gamma \subseteq \Gamma$ and $a^{ku}c^{-1}\Gamma \subseteq \Gamma$ for all $u \in \mathbb{Z}$.

Now, taking $A = \langle ca^k c^{-1} \rangle$, $B = \langle a^k \rangle$, $\Gamma_A = c\Gamma$, and $\Gamma_B = \Gamma$, we see that

$$A\Gamma_B = A\Gamma = \bigcup_{u \in \mathbb{Z}} \{ca^{ku}c^{-1}\Gamma\} = \bigcup_{u \in \mathbb{Z}} \{c(a^{ku}c^{-1}\Gamma)\} \subseteq c\Gamma = \Gamma_A;$$

$$B\Gamma_A = \bigcup_{u \in \mathbb{Z}} \{a^{ku}c\Gamma\} \subseteq \Gamma = \Gamma_B;$$

$$v_1 \in a^{ku}\Gamma_B \cap \Gamma_B, \quad \forall u \in \mathbb{Z} \quad (\text{since } Fv_1 \subset \Gamma = \Gamma_B).$$

Thus, the hypotheses of Lemma 17.16 are satisfied, from which we conclude that $ca^k c^{-1}$ and a generate a free subgroup of G, as desired. This concludes the proof of Theorem 19B.21 in characteristic 0. □

In summary, the proof of the Tits alternative consists of finding an element a with a dominant eigenvalue, modifying the choice so that both a and a^{-1} have dominant eigenvalues, and then showing by means of the pingpong lemma that a suitably high power of a and its conjugate generate a free group.

Recently Breuillard and Gelanter [**BreG**] have strengthened the Tits alternative as follows:

THEOREM 19B.24. *Any f.g. linear group contains either a free subgroup that is Zariski dense (in the relative topology), or a Zariski open solvable subgroup.*

Tits' original method of taking high powers of semisimple elements does not seem to suffice here, so an alternate approach is needed. The following arithmetic version is the main result of Breuillard-Gelanter:

LEMMA 19B.25. *Suppose F is a local field and G is a subgroup of $\mathrm{GL}(n, F)$, endowed with the topology from the absolute value on F. Any f.g. linear group contains either a dense free subgroup or an open solvable subgroup.*

Chapter 20

Characters of Finite Groups

In this chapter we introduce group characters, the key notion in group representations. Various structural tools, such as bilinear forms and tensor products, enhance their role. Group characters have a wide range of applications, including Burnside's celebrated theorem that every group of order p^iq^j (p, q prime) is solvable. At the end of the chapter, we treat induced characters, culminating with a theorem of E. Artin showing that all characters are induced from characters of cyclic subgroups.

Assume throughout that G is a group, $F \subseteq \mathbb{C}$ is a splitting field for G, and $\rho\colon G \to \mathrm{GL}(n, F)$ is a representation.

Definition 20.1. The **character afforded** by the representation ρ is the function $\chi_\rho\colon G \to F$ given by

$$\chi_\rho(g) = \mathrm{tr}(\rho(g));$$

by definition, χ_ρ has **degree** $\deg \chi_\rho = \deg \rho$.

Remark 20.2. (i) $\chi_\rho(1) = \mathrm{tr}(I) = n = \deg \rho$.

(ii) $\chi_\rho(aga^{-1}) = \mathrm{tr}(\rho(aga^{-1})) = \mathrm{tr}(\rho(a)\rho(g)\rho(a)^{-1}) = \mathrm{tr}(\rho(g)) = \chi_\rho(g)$ for all $g \in G$. Thus, $\chi_\rho(g)$ depends only on the conjugacy class of g, and likewise, $\chi_\rho = \chi_{\rho'}$ for any equivalent representations ρ and ρ'.

(iii) Taking the sum of characters as functions, we have $\chi_\rho + \chi_\tau = \chi_{\rho\oplus\tau}$ for all representations ρ and τ. Thus, the sum of characters is a character. We write $m\chi$ for $\chi + \cdots + \chi$, taken m times.

(iv) Any character χ extends to a linear transformation $\hat{\chi}\colon F[G] \to F$ given by $\hat{\chi}(\sum \alpha_g g) = \sum \alpha_g \chi(g)$.

By (ii), each character actually is afforded by an equivalence class of representations. (Fortunately, Corollary 20.6 below shows that inequivalent representations always afford distinct characters.) Before we continue, here are a few basic examples.

Example 20.3. (i) The **unit character** χ_1, the character of the unit representation $\mathbf{1}$, satisfies $\chi_1(g) = 1$ for all $g \in G$.

(ii) More generally, if $\deg \rho = 1$, then $\chi_\rho(g) = \rho(g)$.

(iii) Suppose $\chi = \chi_{\rho_{\mathrm{reg}}}$; cf. Example 19.2. Then $\chi(g) = 0$ unless $g = 1$, in which case $\chi(1) = \mathrm{tr}(I) = \deg \rho_{\mathrm{reg}} = |G|$. It follows that $\hat{\chi}(\sum \alpha_g g) = |G|\alpha_1$.

We also need some basic arithmetic properties of characters.

Proposition 20.4.

(i) *If the element $g \in G$ has order m, then $\chi_\rho(g)$ is a sum of m-roots of unity and thus is integral over $\mathbb{Z}$ (in the sense of Definition 5.19 of Volume 1).*
(ii) $|\chi_\rho(g)| \le \deg \rho$, *equality holding iff $\rho(g)$ is a scalar matrix.*
(iii) $\chi_\rho(g) = \deg \rho$ *iff* $g \in \ker \rho$.
(iv) $\chi_\rho(g^{-1}) = \overline{\chi_\rho(g)}$.
(v) *If g is conjugate to g^{-1}, then $\chi_\rho(g)$ is real.*
(vi) *If $g^2 = 1$, then $\chi_\rho(g)$ is an integer of the same parity as* $\deg \rho$.

Proof. Let $n = \deg \rho$. By Remark 19.8, we may assume that $\rho(g)$ is diagonal whose entries $\zeta_1, \ldots, \zeta_n$ are roots of 1.

(i) All roots of 1 are integral over $\mathbb{Z}$, and thus so is $\sum \zeta_i = \chi_\rho(g)$.

(ii) $|\chi_\rho(g)| \le |\zeta_1| + \cdots + |\zeta_n| = 1 + \cdots + 1 = n$. In case equality holds, all the ζ_i must be equal, so the matrix is scalar.

(iii) All of the ζ_i equal 1 iff $\rho(g) = I$.

(iv) $\chi_\rho(g^{-1}) = \mathrm{tr}(\rho(g)^{-1}) = \sum \zeta_i^{-1} = \sum \overline{\zeta_i} = \overline{\mathrm{tr}(\rho(g))} = \overline{\chi_\rho(g)}$, since the inverse of any root of 1 is its complex conjugate.

(v) By (iv), $\overline{\chi_\rho(g)} = \chi_\rho(g^{-1}) = \chi_\rho(g)$.

(vi) The eigenvalues of $\rho(g)$ are all square roots of 1, i.e., ± 1. □

Schur's orthogonality relations

The character χ_ρ is called **irreducible** if the representation ρ is irreducible. Fixing a set $\rho_1 = \mathbf{1}, \rho_2, \ldots, \rho_t$ of inequivalent irreducible representations of G (of respective degrees $n_1 = 1, n_2, \ldots, n_t$), we write χ_i for χ_{ρ_i}. (Thus, $\chi_1(g) = 1$, $\forall g \in G$.) These are all the irreducible characters of G. Since any linear representation is a finite direct sum of irreducible representations,

any character χ is a finite sum $u_1\chi_1 + \cdots + u_t\chi_t$ of irreducible characters for suitable $u_1, \ldots, u_t \in \mathbb{N}$, where μ_j is called the **multiplicity** of χ_j in χ. In this manner, much character theory reduces to studying $\chi_1, \ldots, \chi_t$, which we do by means of an inner product.

Recall that t is also the number of conjugacy classes of our given finite group G. We write $\mathcal{R}$ for $F^{(t)}$, viewed as an F-algebra where addition and multiplication are taken componentwise. Writing the conjugacy classes of G as $\mathcal{C}_1, \ldots, \mathcal{C}_t$, we define a **class function** to be a function $\{\mathcal{C}_1, \ldots, \mathcal{C}_t\} \to F$. The set of class functions is identified with $\mathcal{R}$, by sending a class function f to the t-tuple $(f(\mathcal{C}_1), \ldots, f(\mathcal{C}_t))$.

We want a base of $\mathcal{R}$ that reflects the structure of G. Remark 20.2(ii) shows that any character is a class function and thus belongs to $\mathcal{R}$; the natural candidate for our base is $\{\chi_1, \ldots, \chi_t\}$. In fact, we get an even stronger result. (The proof is given after a short discussion.)

THEOREM 20.5. *The characters $\chi_1, \ldots, \chi_t$ comprise an orthonormal base of $\mathcal{R}$ with respect to the Hermitian inner product $\langle\ ,\ \rangle$ given by*

$$\langle \phi, \psi \rangle = \frac{1}{|G|} \sum_{1 \le k \le t} |\mathcal{C}_k| \phi(\mathcal{C}_k) \overline{\psi(\mathcal{C}_k)}. \tag{20.1}$$

In particular, the characters $\chi_1, \ldots, \chi_t$ are distinct! Hence, t is also the number of distinct irreducible characters of G.

We call this inner product (20.1) the **Schur inner product**. Viewing a class function ϕ as a function $\phi\colon G \to F$ for which $\phi(a) = \phi(b)$ whenever a and b belong to the same conjugacy class, one could rewrite (20.1) as

$$\langle \phi, \psi \rangle = \frac{1}{|G|} \sum_{1 \le k \le t} \sum_{g \in \mathcal{C}_k} \phi(g)\overline{\psi(g)} = \frac{1}{|G|} \sum_{g \in G} \phi(g)\overline{\psi(g)}, \tag{20.2}$$

which perhaps is more intuitive. Saying $\chi_1, \ldots, \chi_t$ is an orthonormal base thus means:

$$\delta_{ij} = \frac{1}{|G|} \sum_{g \in G} \chi_i(g)\bar{\chi}_j(g). \tag{20.3}$$

This formulation has some immediate dividends.

COROLLARY 20.6.

(i) *Writing an arbitrary class function ϕ as $\sum_{i=1}^t u_i\chi_i$ for suitable u_i in F, we have $u_i = \langle \sum_j u_j\chi_j, \chi_i \rangle = \langle \phi, \chi_i \rangle$. Thus, a class function is a character iff the multiplicity $\langle \phi, \chi_i \rangle \in \mathbb{N}$, for each $1 \le i \le t$.*

(ii) *Two inequivalent irreducible representations cannot afford the same character.*

(iii) *If $\chi = \sum_{i=1}^t u_i\chi_i$, for $u_i \in \mathbb{N}$, then $\langle \chi, \chi \rangle = \sum u_i^2 \in \mathbb{N}$. Consequently $\langle \chi, \chi \rangle = 1$ iff $\chi = \chi_i$ for some i; i.e., iff the character χ is irreducible.*

Example 20.7. Recall that $n_i = \deg \rho_i = \chi_i(1)$. If $\chi = \chi_{\text{reg}}$, then $\langle \chi, \chi_i \rangle = \frac{|G|n_i}{|G|} = n_i$ in view of Example 20.3(iii), so $\chi = \sum n_i\chi_i$. From this vantage point, the regular representation yields all the character theory of G.

We prove Theorem 20.5 by means of the ring-theoretic structure of the group algebra. Write $F[G] = \prod_{i=1}^t M_{n_i}(F)$, and let π_i be the projection onto the i-th component. Since ρ_i is the restriction of π_i to G, we see that

$$\chi_i(g) = \operatorname{tr}(\pi_i(g)).$$

Remark 20.8. Let $e_i = \pi_i(1)$, the idempotent that is the unit element of the i-th component of $F[G]$; thus, $\pi_i(F[G]) = F[G]e_i$. Also $\pi_i(e_j) = \delta_{ij}e_j$. Thus, for any $g \in G$, in the notation of Remark 20.2(iv),

$$\begin{aligned} \hat{\chi}_j(ge_i) &= \operatorname{tr}(\pi_j(ge_i)) \\ &= \operatorname{tr}(\pi_j(g)\pi_j(e_i)) = \operatorname{tr}(\pi_j(g)\delta_{ij}e_j) = \delta_{ij}\operatorname{tr}(\pi_j(g)) = \delta_{ij}\chi_j(g). \end{aligned}$$

Our first task is to compute e_i in terms of the structure of G.

LEMMA 20.9. $|G|e_i = \sum_{g \in G} n_i\chi_i(g^{-1})g$.

Proof. Writing $e_i = \sum_{h \in G} \alpha_h h$, we have $g^{-1}e_i = \alpha_g + \sum_{h \ne g} \alpha_h g^{-1}h$, so, in view of Examples 20.3(iii) and 20.7 and Remark 20.8,

$$|G|\alpha_g = \chi_{\text{reg}}(g^{-1}e_i) = \sum_{j=1}^t n_j\chi_j(g^{-1}e_i) = \sum_{j=1}^t \delta_{ij}n_j\chi_j(g^{-1}) = n_i\chi_i(g^{-1}).$$

Hence

$$|G|e_i = \sum_{g \in G} |G|\alpha_g g = \sum_{g \in G} n_i\chi_i(g^{-1})g. \qquad \square$$

With the preliminaries in hand, we get to the heart of our discussion.

THEOREM 20.10. *For each $a \in G$ and any $1 \le i, j \le t$, we have*

$$\sum_{g \in G} \chi_i(ga)\overline{\chi_j(g)} = \frac{\delta_{ij}|G|\chi_i(a)}{n_i}.$$

Proof. The idea is to compute $|G|^2 e_i e_j$ in two ways: first multiplying $e_i e_j$ before applying Lemma 20.9, and secondly applying Lemma 20.9 before multiplying. First,

$$|G|^2 e_i e_j = \delta_{ij}|G|^2 e_i = \delta_{ij}|G| \sum_{a\in G} n_i \chi_i(a^{-1})a.$$

On the other hand,

$$\begin{aligned} |G|^2 e_i e_j &= (|G|e_i)(|G|e_j) = \sum_{h\in G} n_i\chi_i(h^{-1})h \sum_{g\in G} n_j\chi_j(g^{-1})g \\ &= \sum_{g\in G,\ h\in G} n_i n_j \chi_i(h^{-1})\overline{\chi_j(g)}hg = \sum_{g,a\in G} n_i n_j \chi_i(ga^{-1})\overline{\chi_j(g)}a, \end{aligned}$$

where we substituted $h = ag^{-1}$. Matching coefficients of a in these two expressions yields

$$\delta_{ij}|G| n_i \chi_i(a^{-1}) = n_i n_j \sum_{g\in G} \chi_i(ga^{-1})\overline{\chi_j(g)},$$

or (using a instead of a^{-1}),

$$\sum_{g\in G} \chi_i(ga)\overline{\chi_j(g)} = \frac{\delta_{ij}|G|\chi_i(a)}{n_j} = \frac{\delta_{ij}|G|\chi_i(a)}{n_i}. \qquad \square$$

Finally, we are ready for

Proof of Theorem 20.5. Take $a = 1$ in Theorem 20.10, yielding $\langle \chi_i, \chi_j\rangle = \delta_{ij}$ (via (20.2)) since $\chi_i(1) = n_i$. $\square$

Another proof of Theorem 20.5 is motivated by the interaction of inequivalent irreducible representations ρ_i and ρ_j, which in turn correspond to nonisomorphic simple $F[G]$-modules L_i and L_j; thus we are led to calculate all the $F[G]$-module homomorphisms from L_i to L_j in two ways: structurally (Exercise 19.6, which in a sense is the "true" Schur's Lemma), and via the "averaging procedure" applied to the F-linear transformations from L_i to L_j (cf. Exercise 3).

Although this chapter is limited to finite groups, the reader who has perused Appendix 19A probably has already guessed that a generalization of Theorem 20.5 should hold for compact (infinite) groups, using the inner product

$$\langle f_1, f_2\rangle = \int_G f_1(g)\overline{f_2(g)}dg.$$

This is described concisely in Vinberg [**Vinb,** §**III.8**].

The character table

Having established the critical role of irreducible characters, we recapitulate our notation: $n_i = \deg \rho_i = \chi_i(1)$; also, $\mathcal{C}_1, \dots, \mathcal{C}_t$ denote the conjugate classes of G, and we put $m_j = |\mathcal{C}_j|$. We pick a **conjugacy representative** g_j from $\mathcal{C}_j$ for each j, taking $g_1 = 1$.

We also recall some group theory. Let $C(g) = \{a \in G : ag = ga\}$, the centralizer of g in G. Then $m_j = \frac{|G|}{|C(g_j)|}$ by [**Row3, Proposition 19.7**]; in particular, each m_j divides $|G|$.

Definition 20.11. The **character table** is the $t \times t$ matrix $X = (\chi_{ij})$, where $\chi_{ij} = \chi_i(g_j)$.

Thus, rows correspond to the irreducible characters $\chi_1 = \mathbf{1}$, $\dots$, χ_t, and columns correspond to the conjugacy representatives. Our object is to learn to construct the character table from a given group G, utilizing intrinsic numerical properties. First note that $\chi_{1j} = \chi_1(g_j) = 1$ for all j, and $\chi_{i1} = \chi_i(1) = n_i$ for all i; hence, the general character table of a finite group looks like

Conjugacy Representatives:

	$g_1 = 1$	g_2	$\dots$	g_t
χ_1	1	1	$\dots$	1
χ_2	n_2	$\chi_2(g_2)$	$\dots$	$\chi_2(g_t)$
χ_3	n_3	$\chi_3(g_2)$	$\dots$	$\chi_3(g_t)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
χ_t	n_t	$\chi_t(g_2)$	$\dots$	$\chi_t(g_t)$

Remark 20.12. Let us interpret Proposition 20.4 as information about the character table:

(i) If $o(g_j) = m$, then χ_{ij} is a sum of m-th roots of 1, and thus is integral over $\mathbb{Z}$.

(ii) $|\chi_{ij}| \le n_i$, equality holding iff $\rho_i(g_j)$ is a scalar matrix.

(iii) $\chi_{ij} = n_i$ iff $g_j \in \ker \rho_i$.

(iv) The column corresponding to the class of g is the complex conjugate of the column corresponding to the class of g^{-1}, and is in $\mathbb{Z}$ if $g^2 = 1$.

Example 20.13. (i) Suppose G is a cyclic group $\langle g\rangle$ of order n. There are n conjugacy classes $\{1\}, \{g\}, \dots \{g^{n-1}\}$, so there are n irreducible characters,

each of degree 1, determined in Example 19.3. After the characters are suitably renumbered, the character table must be

Conjugacy Representatives:

	1	g	g^2	$\dots$	g^{n-1}
χ_1	1	1	1	$\dots$	1
χ_2	1	ζ	ζ^2	$\dots$	ζ^{n-1}
χ_3	1	ζ^2	ζ^4	$\dots$	$\zeta^{2(n-1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
χ_n	1	ζ^{n-1}	$\zeta^{2(n-1)}$	$\dots$	ζ

(ii) The Klein group $G = \{1, a, b, ab\}$. Again, there are four irreducible characters, all of degree 1; there are also four possibilities for choosing χ_{i2} and χ_{i3} from ± 1, so the character table can be written as:

Conjugacy Representatives:

	1	a	b	ab
χ_1	1	1	1	1
χ_2	1	-1	1	-1
χ_3	1	1	-1	-1
χ_4	1	-1	-1	1

(iii) $G = S_3$. Thus, $t = 3$, with $n_1 = n_2 = 1$ and $n_3 = 2$ by Remark 19.39. $\chi_1 = 1$ and χ_2 is the sign representation, so we start with the character table

Conjugacy Representatives:

	(1)	(12)	(123)
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	?	?

How can we fill in the last row? Let $N = \langle (123) \rangle$, the only nontrivial normal subgroup of G. If $\ker \rho_3 \supseteq N$, then ρ_3 would correspond to an irreducible representation of $G/N \cong C_2$, which is impossible since we already have ρ_1 and ρ_2. Thus ρ_3 is faithful, and we appeal to Example 19.9(ii) to

produce the last row, yielding the character table

	(1)	(12)	(123)
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

(iv) If $\mathbb{Q}$ is a splitting field for G, then the entries of the character table of G are all in $\mathbb{Z}$. Indeed, the characters are all elements of $\mathbb{Q}$ integral over $\mathbb{Z}$, and thus belong to $\mathbb{Z}$.

(v) For any n, the entries of the complex character table of S_n are all in $\mathbb{Z}$, since $\mathbb{Q}$ is a splitting field; cf. Theorem 19.60 and (iv).

Schur's orthogonality relations applied to the character table.

One can obtain the missing numbers in Example 20.13(iii) merely by utilizing numerical properties of the character table discovered by Schur, which are immediate consequences of Theorem 20.5.

Remark 20.14 (Schur I). Theorem 20.5 also can be viewed as a weighted orthogonality relationship between any two rows of the character table. Explicitly, notation as above, (20.1) becomes

$$\delta_{ik}|G| = \sum_{j=1}^{t} m_j \chi_{ij} \overline{\chi_{kj}}.$$

In the matrix algebra $M_t(F)$, if $AB = \alpha I$, then $B = \alpha A^{-1}$ (and so $BA = \alpha I$). This basic observation enables us to switch from rows to columns and prove another relation.

THEOREM 20.15 (SCHUR II). $\sum_{i=1}^{t} \chi_{ij} \overline{\chi_{ik}} = \delta_{jk} \frac{|G|}{m_k}$.

Proof. χ_{kj} is the j,k entry of the transpose matrix X^t, so letting D denote the diagonal matrix $\operatorname{diag}\{m_1, \dots, m_t\}$ we see that

$$\sum_{j=1}^{t} m_j \chi_{ij} \overline{\chi_{kj}} = \sum_{j=1}^{t} \overline{\chi_{kj}} m_j \chi_{ij}$$

is the k,i entry of the matrix $\overline{X} D X^t$. Thus, Schur I yields the matrix equation

$$|G| I = \overline{X} D X^t,$$

i.e., $X^t = |G|(\overline{X}D)^{-1} = |G| D^{-1} \overline{X}^{-1}$, implying

$$|G| D^{-1} = X^t \overline{X},$$

yielding the desired relations since $D^{-1} = \operatorname{diag}\{m_1^{-1}, \dots, m_t^{-1}\}$. □

Example 20.16. Let G be any nonabelian group of order 8. By Equation (19.6), we write $8 = 4+1+1+1+1$ and conclude that $t = 5$. To build the character table, we note that since G is a 2-group, there is a central element c of order 2, and $G/\langle c\rangle$ has order 4 and exponent 2; thus we start the table from Example 20.13(ii):

Conjugacy Representatives:

	1	a	b	ab	c
χ_1	1	1	1	1	1
χ_2	1	−1	1	−1	1
χ_3	1	1	−1	−1	1
χ_4	1	−1	−1	1	1
χ_5	2	?	?	?	?

($\chi_{15} = \chi_{25} = \chi_{35} = \chi_{45} = 1$, since the image of c in $G/\langle c\rangle$ is 1.) We fill in the table by using Schur II for the first and j-th columns, for each $j \geq 2$ in turn:

Conjugacy Representatives:

	1	a	b	ab	c
χ_1	1	1	1	1	1
χ_2	1	−1	1	−1	1
χ_3	1	1	−1	−1	1
χ_4	1	−1	−1	1	1
χ_5	2	0	0	0	−2

Note that D_4 and the quaternion group Q both are nonabelian of order 8, so this gives an example of two nonisomorphic groups having the same character tables. Also note that although its character table is defined over $\mathbb{Z}$, Q is not split by $\mathbb{Q}$ (or even by $\mathbb{R}$), by Example 19.46(iii). Other character tables are worked out in Exercises 6 and 10–16.

Arithmetic properties of characters

Surprisingly, we get even more arithmetic information about the X_{ij} by returning to the center of the group algebra. Since $F \subseteq \mathbb{C}$ has characteristic 0, we can view $\mathbb{Z} \subset F$, and thus $\mathbb{Z}[G] \subset F[G]$. Let $Z = \mathrm{Cent}(\mathbb{Z}[G])$. By Theorem 19.42, Z is a free $\mathbb{Z}$-module with base $z_1, \ldots, z_t$, where $z_j = z_{\mathcal{C}_j} = \sum_{g\in\mathcal{C}_j} g$.

Clearly, $Z \subseteq \mathrm{Cent}(F[G])$ (since each z_j commutes with each $g \in G$). Hence Remark 19.35 implies $\hat{\rho}_i(z_j) \in \mathrm{Cent}(M_{n_i}(F_i))$; i.e., $\hat{\rho}_i(z_j)$ is a scalar matrix $\alpha_{ij} I$ for some $\alpha_{ij} \in F$. These numbers α_{ij} play a central role, so we compute them:

$$n_i\alpha_{ij} = \mathrm{tr}(\hat{\rho}_i(z_j)) = \mathrm{tr}\left(\sum_{g\in\mathcal{C}_j} \rho_i(g)\right) = \sum_{g\in\mathcal{C}_j} \mathrm{tr}\,\rho_i(g) = \sum_{g\in\mathcal{C}_j} \chi_{ij} = m_j\chi_{ij}.$$

Thus,

$$\alpha_{ij} = \frac{m_j\chi_{ij}}{n_i}. \tag{20.4}$$

But Z is a ring, so $z_jz_k \in Z$ for all j, k, and thus

$$z_jz_k = \sum u_{jk\ell}\, z_\ell \tag{20.5}$$

for suitable $u_{jk\ell}$ in $\mathbb{Z}$. Applying $\hat{\rho}_i$ to (20.5) yields

$$\alpha_{ij}\alpha_{ik} I = \alpha_{ij} I\, \alpha_{ik} I = \sum u_{jk\ell}\, \alpha_{i\ell} I. \tag{20.6}$$

Equation (20.6) leads directly to the following important observation.

LEMMA 20.17. *Each number $\alpha_{ij} = \frac{m_j\chi_{ij}}{n_i}$ is integral over $\mathbb{Z}$.*

Proof. Let $M_i = \sum_{k=1}^{t} \alpha_{ik}\mathbb{Z}$, a f.g. $\mathbb{Z}$-module containing $\alpha_{i1}, \ldots, \alpha_{it}$. Then, for each j, comparing coefficients of the scalar matrices in (20.6) yields

$$\alpha_{ij}M_i = \sum_k \alpha_{ij}\alpha_{ik}\mathbb{Z} \subseteq \sum_\ell \alpha_{i\ell}\mathbb{Z} = M_i,$$

implying that α_{ij} is integral over $\mathbb{Z}$, by Theorem 5.21 of Volume 1. □

Recalling that the only elements of $\mathbb{Q}$ integral over $\mathbb{Z}$ are in $\mathbb{Z}$ itself, we are ready for some startling applications.

THEOREM 20.18 (FROBENIUS). *n_i divides $|G|$ for each i.*

Proof. By Schur I, $\frac{|G|}{n_i} = \sum_k \frac{m_k\chi_{ik}\overline{\chi_{ik}}}{n_i} = \sum_k \alpha_{ik}\overline{\chi}_{ik}$, a sum of products of integral elements (by Proposition 20.4(i) and Lemma 20.17), which thus is integral over $\mathbb{Z}$. But $\frac{|G|}{n_i}$ is rational, and so is in $\mathbb{Z}$. □

This result can be improved slightly, at no extra charge.

COROLLARY 20.19. *n_i divides $[G:\ker\rho_i]$ for each i.*

Proof. ρ_i also yields an irreducible representation of $G/\ker\rho_i$. □

Using Remark 20.21 below, we improve this result further in Exercise 20. Our next application highlights a peculiar feature of the character table.

PROPOSITION 20.20. *If $\gcd(n_i, m_j) = 1$, then either $\chi_{ij} = 0$ or $|\chi_{ij}| = n_i$.*

Proof. Let $a = \frac{\chi_{ij}}{n_i}$. We want to prove that $|a|$ is 0 or 1. Write $1 = un_i + vm_j$ for suitable integers u, v. Then

$$a = (un_i + vm_j)\frac{\chi_{ij}}{n_i} = u\chi_{ij} + v\frac{m_j\chi_{ij}}{n_i} = u\chi_{ij} + v\alpha_{ij} \tag{20.7}$$

is integral over $\mathbb{Z}$, and $|a| \leq 1$ by Proposition 20.4(ii). If we could conclude that $a \in \mathbb{Q}$, then $a \in \mathbb{Z}$, and we would be done, but this need not be true. Yet we can make this approach succeed by using Galois theory.

Let ζ be a primitive $|G|$-th root of 1. Recalling from Example 4.48 of Volume 1 that $\mathbb{Q}(\zeta)$ is a Galois extension of $\mathbb{Q}$, let $H = \mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$. By Proposition 20.4(i), $\chi_{ij} = \sum_{u=1}^{n_i} \zeta_u \in \mathbb{Q}(\zeta)$, where ζ_u are $|G|$-th roots of 1, implying that $a \in \mathbb{Q}(\zeta)$. For any σ in H, clearly $\sigma(\zeta_u)$ is also a $|G|$-th root of 1 for each u, and thus $\sigma(\chi_{ij}) = \sum_{u=1}^{n_i} \sigma(\zeta_u)$ is also a sum of $|G|$-th roots of 1. Hence $|\sigma(a)| \leq \frac{n_i}{n_i} = 1$ for each σ in H. Let $b = \prod_{\sigma\in H} \sigma(a) \in \mathbb{Q}$ in view of Remark 4.103 of Volume 1. Each $\sigma(a)$ is integral over $\mathbb{Z}$, in analogy to Equation (20.7), so b is integral over $\mathbb{Z}$. It follows that $b \in \mathbb{Z}$. But $|b| = \prod |\sigma(a)| \leq 1$. Hence $|b|$ is 0 or 1.

If $|b| = 0$, then some $\sigma(a) = 0$, so $a = 0$, implying $\chi_{ij} = 0$.

If $|b| = 1$, then all the $|\sigma(a)| = 1$; in particular, $|a| = 1$, so $|\chi_{ij}| = n_i$. □

Application: Burnside's Theorem.

Proposition 20.20 can be used at times to locate zeroes in the character tables. However, we push onward to our main application, Burnside's Theorem, which involves simple groups.

Remark 20.21. For any $1 \leq i \leq t$, let $Z_i = \{g \in G : \rho_i(g) \in F\cdot I\} \lhd G$. Then $\rho_i(Z_i)$ is a finite multiplicative subgroup of F and thus is cyclic; cf. Proposition 4.1 of Volume 1.

THEOREM 20.22. *Suppose G is a finite nonabelian simple group. Then m_j cannot be a power (other than 1) of a prime number p.*

Proof. Otherwise write $m_j = p^u$ for $u \geq 1$. Note that $j > 1$. Also assume $i > 1$. Then $\ker\rho_i$ is a proper normal subgroup of G, and is thus trivial. Hence $\rho_i(G) \cong G$ is nonabelian, implying that Z_i (of Remark 20.21) is *not* G. It follows that $Z_i = \{e\}$. Thus, $|\chi_{ij}| < n_i$ by Remark 20.12. Hence, by Proposition 20.20, $\chi_{ij} = 0$ for any $i > 1$ such that p does not divide n_i.

Now, by Schur II,

$$0 = \sum_i \chi_{ij}\overline{\chi_{i1}} = \sum_i \chi_{ij}n_i = 1 + \sum_{i>1} n_i\chi_{ij} = 1 + \sum_{p|n_i} n_i\chi_{ij}.$$

Thus, writing $n_i = pn_i'$ we see that $\frac{1}{p} = -\sum_{p|n_i} n_i'\chi_{ij}$, which is integral over $\mathbb{Z}$. But $\frac{1}{p}$ is not integral over $\mathbb{Z}$, a contradiction. □

We are ready for a famous theorem of Burnside in the theory of groups.

THEOREM 20.23 (BURNSIDE'S THEOREM). *There does not exist a nonabelian simple group G of order p^uq^v (for p, q prime).*

Proof. Otherwise, take such a simple group G and a p-Sylow subgroup P of G. Being a p-group, P has a central element $z \neq e$. We show that $z \in Z(G)$, thereby arriving at the contradiction that $\langle z\rangle \lhd G$. Indeed, let $\mathcal{C}_j$ be the conjugacy class of z. Then $m_j = |\mathcal{C}_j| = \frac{|G|}{|C(z)|}$. But clearly $P \subseteq C(z)$, so by Lagrange's theorem $|P|$ divides $|C(z)|$. Therefore $m_j = \frac{|G|}{|C(z)|}$ divides $\frac{|G|}{|P|}$, a power of q, and hence must be 1 by Theorem 20.22. Hence $C(z) = G$, implying that $z \in Z(G)$, as desired. □

An alternate formulation for Burnside's Theorem involves solvable groups.

THEOREM 20.24. *Every group of order p^uq^v (for p, q prime) is solvable.*

Proof. Induction on $|G|$. We may assume that G has no nontrivial normal subgroup N, or else we apply induction to assert that N and G/N are each solvable (since each has smaller order), implying that G is solvable.

Now G, being simple, is Abelian by Theorem 20.23. □

Note that the smallest nonsolvable group, A_5, has order $60 = 2^2 \cdot 3 \cdot 5$.

Tensor products of representations

We can push the theory much further when introducing tensor products, which enable us both to multiply characters explicitly and also to induce representations from subgroups. Let us start with an innocuous-looking but effective result.

PROPOSITION 20.25. *For any groups G and H, there is an algebra isomorphism $F[G\times H]\to F[G]\otimes_F F[H]$ given by $(g,h)\mapsto g\otimes h$.*

Proof. The natural group injection $G\times H\to \mathrm{Unit}(F[G]\otimes F[H])$ given by $(g,h)\mapsto g\otimes h$ yields a homomorphism $\phi\colon F[G\times H]\to F[G]\otimes_F F[H]$ by Lemma 19.15. ϕ sends a base to a base, and therefore is 1:1 and onto. □

COROLLARY 20.26. *For any group G, there is a natural algebra homomorphism*

$$\Delta\colon F[G]\mapsto F[G]\otimes F[G]$$

satisfying $\Delta(g)=g\otimes g$.

Proof. The group homomorphism $G\to G\times G$ given by $g\mapsto(g,g)$ yields an algebra homomorphism

$$F[G]\mapsto F[G\times G]\cong F[G]\otimes_F F[G].$$ □

This homomorphism is called **comultiplication** in the group algebra (as compared to comultiplication in an algebraic group, defined in Remark 19B.13).

Products of characters.

Having seen in Remark 20.2(iii) that the sum of two characters is a character, we prove an analogous result for products of characters.

Definition 20.27. Suppose $\rho\colon G\to \mathrm{GL}(m,F)$ and $\tau\colon G\to \mathrm{GL}(n,F)$ are group representations whose corresponding G-modules are V and W, respectively. The **product** $\rho\otimes\tau$ is the representation (of degree mn) corresponding to the G-module $V\otimes_F W$, given the **diagonal action**

$$g(v\otimes w)=gv\otimes gw.$$

(This is not the module action for the tensor product as described in Equation (18.3) of Chapter 18, which only involves the first tensor component.)

Remark 20.28. Here is another way of viewing $\rho\otimes\tau$. The group representations ρ and τ induce algebra homomorphisms $\hat\rho\colon F[G]\to M_m(F)$ and $\hat\tau\colon F[G]\to M_n(F)$; taking the composite $(\hat\rho\otimes\hat\tau)\circ\Delta$ (of Corollary 20.26) yields an algebra homomorphism

$$F[G]\to M_m(F)\otimes M_n(F)$$

satisfying $g\mapsto\rho(g)\otimes\tau(g)$. Composing this with the natural isomorphism $M_m(F)\otimes M_n(F)\to M_{mn}(F)$ (of Corollary 18.36) gives us $\rho\otimes\tau$.

PROPOSITION 20.29. *$\chi_{\rho\otimes\tau}=\chi_\rho\chi_\tau$ for any f.d. representations ρ and τ.*

Proof. We write the representation of Remark 20.28 explicitly. Writing

$$\rho(g)=(a_{ij}),\qquad \tau(g)=(b_{k\ell}),$$

we see using Remark 18.37 that

$$\begin{aligned}\chi_{\rho\otimes\tau}(g)&=\mathrm{tr}\big((a_{ij})\otimes(b_{k\ell})\big)\\&=\sum_{i,k}a_{ii}b_{kk}\\&=\sum_i a_{ii}\sum_k b_{kk}=\mathrm{tr}\big((a_{ij})\big)\,\mathrm{tr}\big((b_{k\ell})\big)=\chi_\rho(g)\chi_\tau(g).\end{aligned}$$ □

COROLLARY 20.30. *The set of characters is closed under multiplication (as well as addition) in $\mathcal{R}=\mathbb{C}^{(t)}$ (defined preceding Theorem 20.5).*

Representations of direct products.

One easy application of tensor products is to the character table of the direct product $G\times H$ of groups G and H. The key is the Schur inner product of Theorem 20.5.

LEMMA 20.31. *For any representations ρ,ρ' of G and representations τ,τ' of H, we have*

$$\langle\chi_{\rho\otimes\tau},\ \chi_{\rho'\otimes\tau'}\rangle=\langle\chi_\rho,\chi_{\rho'}\rangle\langle\chi_\tau,\chi_{\tau'}\rangle.$$

In particular, $\rho\otimes\tau$ is an irreducible representation of $G\times H$ iff the representations ρ and τ are irreducible.

Proof. Since $\chi_{\rho\otimes\tau}=\chi_\rho\chi_\tau$, we have

$$\begin{aligned}\langle\chi_{\rho\otimes\tau},\ \chi_{\rho'\otimes\tau'}\rangle&=\frac{1}{|G||H|}\sum_{(g,h)\in G\times H}\chi_\rho(g)\chi_\tau(h)\overline{\chi_{\rho'}(g)\chi_{\tau'}(h)}\\&=\frac{1}{|G||H|}\sum_{(g,h)\in G\times H}\chi_\rho(g)\overline{\chi_{\rho'}(g)}\chi_{\tau'}(h)\overline{\chi_\tau(h)}\\&=\langle\chi_\rho,\ \chi_{\rho'}\rangle\langle\chi_\tau,\ \chi_{\tau'}\rangle.\end{aligned}$$

$\langle\chi_{\rho\otimes\tau},\ \chi_{\rho\otimes\tau}\rangle=1$ iff $\langle\chi_\rho,\chi_\rho\rangle=1=\langle\chi_\tau,\chi_\tau\rangle$, in view of Corollary 20.6(iii), which also yields the last assertion. □

Note that (g_1,h_1) and (g_2,h_2) are conjugate in $G\times H$, iff the g_i are conjugate in G and the h_i are conjugate in H; indeed

$$(g_2,h_2)=(a,b)(g_1,h_1)(a,b)^{-1}\quad\text{iff}\quad g_2=ag_1a^{-1}\text{ and }h_2=bh_1b^{-1}.$$

THEOREM 20.32. *If G and H are groups having respective character tables X_G and X_H (over $\mathbb{C}$), then the character table of $G\times H$ is the tensor product of the character tables of G and of H.*

Proof. Suppose $\rho_i: G \to \mathrm{GL}(n_i, F)$, $1 \le i \le t$, and $\tau_j: H \to \mathrm{GL}(m_j, F)$, $1 \le j \le u$, are all the irreducible representations (up to equivalence) of G and H respectively. By Lemma 20.31, all of the $\rho_i \otimes \tau_j$ are irreducible representations of $G \times H$ and analogously are mutually orthogonal.

Thus, we get tu inequivalent irreducible representations of $G \times H$, and since $G \times H$ has precisely tu conjugacy classes, these are all the equivalence classes of irreducible representations of $G \times H$. (Another way of verifying this assertion is via the degree formula

$$|G||H| = \sum_i n_i^2 \sum_j m_j^2 = \sum_{i,j} (n_i m_j)^2.)$$

The assertion is now immediate from Remark 18.37. □

Induced representations and their characters

Our next application of tensor products is a construction whereby we induce a representation of a group G from a given representation ρ of a subgroup H. We recall that a **transversal** $\{b_1, b_2, \dots\}$ of H in G is a subset of G such that G is the disjoint union $\bigcup_i b_i H$. Thus, any element g of G can be written uniquely as $b_i h$, where $h \in H$ depends on g.

Let us pave the way with some ad hoc examples.

Example 20.33. (i) Suppose $H \lhd G$. Any character $\chi: H \to F$ can be extended to a class function $\hat{\chi}: G \to F$ by putting $\hat{\chi}(g) = 0$ for all $g \in G \setminus H$. Hence $\hat{\chi}$ is a linear combination of characters of representations of G. (Shortly we shall see how to get $\hat{\chi}$ from a single representation.)

(ii) Suppose $[G : H] = k$, and take a transversal $\{b_1, b_2, \dots, b_k\}$ of H in G. Define the representation $\mathbf{1}^G: G \to \mathrm{GL}(k, F)$ by taking $\mathbf{1}^G(g)$ to be the $k \times k$ matrix whose i, j entry is 1 if $b_i^{-1} g b_j \in H$, and 0 otherwise. This is a permutation representation.

(iii) Generalizing (ii), taking any representation $\rho: H \to \mathrm{GL}(n, F)$, define the representation $\rho^G: G \to \mathrm{GL}(nk, F)$ by taking $\rho^G(g)$ to be the $nk \times nk$ matrix comprised of blocks of $n \times n$ matrices whose i, j block is $\rho(b_i^{-1} g b_j)$ if $b_i^{-1} g b_j \in H$ and 0 otherwise (for $1 \le i, j \le k$). This can be seen by direct calculation to be a representation.

(iv) In (iii), when $\deg \rho = 1$, we call ρ^G a **monomial representation**; note that here $\rho^G(g)$ is a permutation matrix times a diagonal matrix; i.e., each row has exactly one nonzero entry (and likewise for columns). The character of a monomial representation is called a **monomial character**.

To gain a deeper understanding of these constructions, let us start over again, using modules over group algebras. The group inclusion $H \subseteq G$ induces an algebra injection $F[H] \to F[G]$, so we view $F[H] \subseteq F[G]$.

Proposition 20.34. *If $\{b_1, b_2, \dots\}$ is a transversal of H in G, then*

$$F[G] = \bigoplus_i b_i F[H]$$

as right $F[H]$-modules.

Proof. Any element of G is in $\sum b_i F[H]$, proving that the transversal spans $F[G]$.

To prove independence, suppose $0 = \sum_i b_i a_i$ for $a_i \in F[H]$. Writing $a_i = \sum_{h \in H} \alpha_{i,h} h$ for $\alpha_{i,h} \in F$, we get

$$0 = \sum_{h \in H} \sum_i \alpha_{i,h} b_i h = \sum_{g \in G} \alpha_{i,h} g,$$

where $g = b_i h$ runs through the elements of G. Hence all $\alpha_{i,h} = 0$, so all $a_i = 0$. □

We are ready to define the induced representation.

Definition 20.35. Suppose $H < G$, and ρ is a representation of H. Let M_H denote the $F[H]$-module corresponding to ρ. The **induced representation** ρ^G is the representation of G corresponding to the $F[G]$-module

$$M_H^G = F[G] \otimes_{F[H]} M_H.$$

Note. We wrote M_H to stress that M is an $F[H]$-module. Also, this notation M_H^G has nothing to do with fixed subsets under a group action.

Remark 20.36. $M_H^G = F[G] \otimes_{F[H]} M_H$ is indeed an $F[G]$-module, since $F[G]$ is an $F[G], F[H]$-bimodule; the module multiplication is given by

$$g(a \otimes v) = ga \otimes v$$

for $a, g \in G$ and $v \in M_H$. Equivalent representations have equivalent induced representations, by Remark 18.6.

Remark 20.37. By Proposition 18.12, notation as in Proposition 20.34, any element of M_H^G can be written uniquely as $\sum b_i \otimes v_i$, where $v_i \in M_H$.

Note that $\dim_F M_H^G = [G : H] \dim_F M_H$, so $\deg \rho^G = [G : H] \deg \rho$.

Remark 20.38. Suppose $[G:H] = k < \infty$ and ρ is a representation of H of degree m. Let us compute the induced representation ρ^G and its character. We fix a transversal $\{b_1, b_2, \dots b_k\}$ of H in G.

Take the given $F[H]$-module M_H corresponding to ρ, i.e., $\rho(h)(v) = hv$ for all $h \in H,\ v \in M_H$. If $w_1, \dots, w_n$ is a base of M_H, then

$$\{b_j \otimes w_u : 1 \le j \le k,\ 1 \le u \le n\}$$

comprise a base of $M_H^G = F[G] \otimes M_H$ by Remark 20.37, and one needs only check the action of $\rho^G(g)$ on this base.

Clearly, gb_jH is some coset $b_{\pi_g(j)}H$, where $\pi_g(j) \in \{1, \dots, k\}$. The function $\pi_g: \{1, \dots, k\} \to \{1, \dots, k\}$ is 1:1 and thus a permutation. We write $\pi_g j$ for $\pi_g(j)$. Thus,

$$gb_j = b_{\pi_g j} h_{j,g} \tag{20.8}$$

for suitable $h_{j,g} \in H$, so for arbitrary $v \in M_H$ we have

$$\begin{aligned} g(b_j \otimes v) &= gb_j \otimes v \\ &= b_{\pi_g j} h_{j,g} \otimes v \\ &= b_{\pi_g j} \otimes h_{j,g} v. \end{aligned}$$

Let us interpret this as a matrix. The matrix for $\rho^G(g)$ is a $k \times k$ array of blocks each of size $m \times m$, such that the block that appears in the $(\pi_g j,\, j)$-position is the matrix for $\rho(h_{j,g})$ and all other blocks are 0.

There is a nonzero block at the (i, j) position exactly when $\pi_g(j) = i$; i.e.,

$$b_i H = gb_j H,$$

and thus $b_i^{-1} g b_j \in H$. In this case

$$gb_j = b_i (b_i^{-1} g b_j),$$

so comparing with (20.8) yields $h_{j,g} = b_i^{-1} g b_j$. Thus, when $b_i^{-1} g b_j \in H$, the block appearing in the (i, j) position is the matrix for $\rho(b_i^{-1} g b_j)$; otherwise this block is 0. This is precisely Example 20.33(iii).

Now let us compute the character $\chi = \chi_{\rho^G}$. Here we are concerned only with the diagonal blocks, i.e., $i = j$, so

$$\chi(g) = \sum \chi_\rho(b_i^{-1} g b_i),$$

summed over all i with $b_i^{-1} g b_i \in H$. More concisely, define $\hat\chi$ by

$$\hat\chi(g) = \begin{cases} \chi_\rho(g) & \text{if } g \in H; \\ 0 & \text{otherwise.} \end{cases} \tag{20.9}$$

Note that $\hat\chi(hgh^{-1}) = \hat\chi(g)$ for all g in G (since $hgh^{-1} \in H$ iff $g \in H$). Now we have the formula

$$\chi(g) = \sum_{i=1}^{k} \hat\chi(b_i^{-1} g b_i) = \frac{1}{|H|} \sum_{a \in G} \hat\chi(a^{-1} g a). \tag{20.10}$$

Although $\hat\chi$ is not necessarily a class function in general, it is a class function when $H \lhd G$.

Example 20.39. Suppose $H \lhd G$. Then $a^{-1} g a \in H$ iff $g \in H$, and we have

$$\chi(g) = \begin{cases} [G:H] \chi_\rho(g) & \text{if } g \in H; \\ 0 & \text{otherwise.} \end{cases}$$

In this case $\chi = [G:H]\hat\chi$.

Example 20.40. F is the $F[H]$-module corresponding to the unit representation $\mathbf{1}_H$, and thus $F[G] \otimes_{F[H]} F$ is the $F[G]$-module corresponding to the representation $(\mathbf{1}_H)^G$, which has been described in Example 20.33(ii) and is a permutation representation.

Let us see how trivial characters induce up to nontrivial characters when $G = S_n$. In each example $\{\chi_i : 1 \le i \le t\}$ denotes the irreducible characters of G, and we compute the character χ of the representation induced from the unit representation $\mathbf{1}_H$ for various $H < G$.

(i) $H = A_n$. Since $H \lhd G$ has index 2, we see from Example 20.39 that

$$\begin{cases} \chi(h) = 2 & \text{for} \quad h \in H; \\ \chi(g) = 0 & \text{for} \quad g \in G \setminus H. \end{cases}$$

Hence $\chi = \chi_1 + \chi_2$, where χ_1 is the trivial character and χ_2 is the character of the "sign representation." In other words, $\chi_2 = \chi - \chi_1$.

(ii) $n = 3$ and $H = \langle (12) \rangle$. Then $\chi(1) = [G : H] = 3$. Note that $g(12)g^{-1} \in H$ iff $g \in H$, so, by formula (20.9), $\chi((12)) = 1$. On the other hand, $g(123)g^{-1}$ is *never* in H since $\langle (123) \rangle \lhd G$. Thus, $\chi((123)) = 0$. But the vectors $(3\ 1\ 0) = (1\ 1\ 1) + (2\ 0\ -1)$, so we conclude that $\chi = \chi_1 + \chi_3$, cf. Example 20.13(iii), or $\chi_3 = \chi - \chi_1$.

We have computed each irreducible character of S_3 as a linear combination of characters induced from trivial characters of certain subgroups. This sort of induced representation plays a special role in the theory. For example, for the subgroup S^λ of S_n defined in Exercise 19.38, Sagan [**Sag**] shows how Young's theory described in Chapter 19 can be reformulated in terms of inducing the unit representation of S^λ to S_n. Our objective for the remainder of this chapter is to explain this phenomenon in general terms, culminating in Theorems 20.44 and 20.46 and the subsequent discussion.

Remark 20.41. Define the **coinduced module** $N = \mathrm{Hom}_{\mathbb{Z}[H]}(\mathbb{Z}[G], M_H)$, viewed as an $F[G]$-module as in Remark 18.43. When $[G\!:\!H] < \infty$, we claim that $N \cong M_H^G$. To verify this, notation as in Remark 20.38, we need to define a G-module map on the base $\{b_j \otimes w_u : 1 \le j \le k,\ 1 \le u \le n\}$ of M_H^G, taking values in N, and show that it is an isomorphism.

Note that $\{b_i^{-1} : 1 \le i \le k\}$ is a left transversal of H in G since, for any element $g \in G$, writing $g^{-1} = b_i h$, we have $g = h^{-1} b_i^{-1} \in H b_i^{-1}$. Hence, $\{b_i^{-1} : 1 \le i \le k\}$ comprises a base of $\mathbb{Z}[G]$ as a (left) $\mathbb{Z}[H]$-module, so we can define any $\mathbb{Z}[H]$-module map $\mathbb{Z}[G] \to M_H$ by declaring what the action is on the b_i^{-1}. Given $a \otimes w$, where $a = b_j h$, we define $\phi_{a,w}$ by

$$\phi_{a,w}(b_i^{-1}) = \delta_{ij} h w,$$

where δ_{ij} is the Kronecker delta. Thus $\phi_{b_j h, w} = \phi_{b_j, hw}$. By Equation (20.8), $b_{\pi_g i}^{-1} g = h_{i,g} b_i^{-1}$, so

$$(g\phi_{b_j,w})(b_{\pi_g i}^{-1}) = \phi_{b_j,w}(b_{\pi_g i}^{-1} g) = \phi_{b_j,w}(h_{i,g} b_i^{-1}) = h_{i,g}\phi_{b_j,w}(b_i^{-1}) = \delta_{ij} h_{i,g} w,$$

implying

$$g\phi_{b_j,w} = \phi_{b_{\pi_g j}, h_{j,g} w} = \phi_{b_{\pi_g j} h_{j,g}, w} = \phi_{g b_j, w},$$

so $b_j \otimes w_u \mapsto \phi_{b_j, w_u}$ defines the desired monic map of $\mathbb{Z}[G]$-modules. Comparing dimensions, we have an isomorphism.

Comparing induced representations.

Although a character induced from a trivial character can be reducible, the trivial character appears as a component in each instance of Example 20.40. This can be explained through the next result. Given a representation ρ of G and $H < G$, let ρ_H denote the restriction of ρ to H.

Theorem 20.42 (Frobenius Reciprocity Theorem). *Suppose $F \subseteq \mathbb{C}$ is a splitting field for a finite group G, σ is an irreducible representation of a subgroup H, and ρ is an irreducible representation of G. Then the multiplicity of ρ in σ^G is the same as the multiplicity of σ in ρ_H.*

Proof. By means of the Schur inner product on $\mathcal{R}$ of Theorem 20.5. Let $\chi = \chi_\rho$, a character of G, and $\psi = \psi_\sigma$, a character of H. We need to show that the multiplicity $\langle \psi^G, \chi \rangle_G$ of χ in ψ^G equals the multiplicity $\langle \psi, \chi_H \rangle_H$ of ψ in χ_H. Using Equation (20.9), we have

$$\begin{aligned}
\langle \chi^G, \chi \rangle_G &= \frac{1}{|G|}\frac{1}{|H|} \sum_{a,g \in G} \hat{\psi}(a^{-1} g a)\overline{\chi(g)} \\
&= \frac{1}{|G|}\frac{1}{|H|} \sum_{a,g \in G} \hat{\psi}(g)\overline{\chi(aga^{-1})} \\
&= \frac{1}{|G|}\frac{1}{|H|} \sum_{a \in G,\ g \in H} \psi(g)\overline{\chi(aga^{-1})} \\
&= \frac{1}{|H|} \sum_{g \in H} \psi(g)\overline{\chi(g)} \\
&= \langle \psi, \chi_H \rangle_H. \qquad \square
\end{aligned}$$

Perhaps this proof is too slick to "explain" why the two multiplicities are equal. For a more conceptual proof, see Exercises 25 and 26. Here is another easy application of module theory.

Theorem 20.43. *Suppose $H < K < G$ and ρ is a representation of H. Then*

(i) *$(\rho^K)^G$ and ρ^G are equivalent representations.*
(ii) *If $\rho = \rho_1 \oplus \rho_2$, then ρ^G is equivalent to $\rho_1^G \oplus \rho_2^G$.*
(iii) *$\rho^G \otimes \sigma$ and $(\rho \otimes \sigma_H)^G$ are equivalent representations, for any representation σ of G.*

Proof. Each assertion is equivalent to a familiar property of tensor products. Let M_H be the $F[H]$-module corresponding to ρ.

(i) $F[G] \otimes_{F[K]} (F[K] \otimes_{F[H]} M_H) \cong (F[G] \otimes_{F[K]} F[K]) \otimes_{F[H]} M_H \cong F[G] \otimes_{F[H]} M_H$ as $F[G]$-modules;.

(ii) Let M_j denote the $F[H]$-module corresponding to ρ_j for $j = 1, 2$. Then the decomposition $M_H \cong M_1 \oplus M_2$ lifts to a decomposition

$$M_H^G \cong F[G] \otimes (M_1 \oplus M_2) \cong (F[G] \otimes M_1) \oplus (F[G] \otimes M_2).$$

(iii) Let N_H denote the $F[H]$-module corresponding to σ. Although

$$F[G] \otimes_{F[H]} (M_H \otimes_F N_H) \cong (F[G] \otimes_{F[H]} M_H) \otimes_F N_H$$

as F-vector spaces, under the map

$$a \otimes (v \otimes w) \mapsto (a \otimes v) \otimes w,$$

this map does *not* preserve the G-module action of Definition 20.27. So let us try another tack. Define the maps $\psi_a : M_H \to F[G] \otimes_{F[H]} M_H$ given by

$v \mapsto a \otimes v$, and $\ell_a \colon N_H \to N_H$ given by $w \mapsto aw$. These are both linear transformations over F and thus produce

$$\psi_a \otimes \ell_a \colon M_H \otimes_F N_H \to (F[G] \otimes_{F[H]} M_H) \otimes_F N_H$$

given by $v \otimes w \mapsto (a \otimes v) \otimes aw$. Now we have a balanced map (over $F[H]$)

$$F[G] \times (M_H \otimes_F N_H) \to (F[G] \otimes_{F[H]} M_H) \otimes_F N_H$$

given by

$$(a, v \otimes w) \mapsto (\psi_a \otimes \ell_a)(v \otimes w) = (a \otimes v) \otimes aw.$$

This provides a map $F[G] \otimes_{F[H]} (M_H \otimes_F N_H) \to (F[G] \otimes_{F[H]} M_H) \otimes_F N_H$ given by

$$a \otimes (v \otimes w) \mapsto (a \otimes v) \otimes aw,$$

clearly an $F[G]$-module map and onto (since the inverse image of $(g \otimes v) \otimes w$ is $g \otimes (v \otimes g^{-1}w)$ for any g in G); comparing dimensions shows that it is an isomorphism. □

Artin's theorem on characters.

THEOREM 20.44 (ARTIN). *Every complex character (for a group G) is a linear combination (over $\mathbb{Q}$) of complex characters induced from cyclic subgroups of G.*

We prove Artin's theorem as a consequence of a more technical version. Given a subring C of $\mathbb{C}$, define $\mathcal{R}_C(G)$ to be the C-subalgebra of $\mathcal{R}$ (defined preceding Theorem 20.5) generated by the irreducible characters.

Remark 20.45. If $H < G$, then restriction of characters induces a canonical homomorphism $\mathrm{Res} \colon \mathcal{R}_{\mathbb{Q}}(G) \to \mathcal{R}_{\mathbb{Q}}(H)$, whereas induction of characters induces a canonical injection $\mathrm{Ind} \colon \mathcal{R}_{\mathbb{C}}(H) \to \mathcal{R}_{\mathbb{C}}(G)$, which restricts to an injection

$$\mathrm{Ind} \colon \mathcal{R}_{\mathbb{Q}}(H) \to \mathcal{R}_{\mathbb{Q}}(G).$$

THEOREM 20.46 (ARTIN'S THEOREM, TECHNICAL VERSION). *Suppose $\mathcal{S}$ is a set of subgroups of G such that G is the union of conjugates of subgroups from $\mathcal{S}$. Then $\mathcal{R}_{\mathbb{Q}}(G) = \sum_{H \in \mathcal{S}} \mathrm{Ind}(\mathcal{R}_{\mathbb{Q}}(H))$.*

Proof. For any set $\mathcal{S}$ of subgroups of G, Remark 20.45 yields homomorphisms

$$\mathrm{Res} \colon \mathcal{R}_{\mathbb{C}}(G) \to \bigoplus_{H \in \mathcal{S}} \mathcal{R}_{\mathbb{C}}(H) \quad \text{and} \quad \mathrm{Ind} \colon \bigoplus_{H \in \mathcal{S}} \mathcal{R}_{\mathbb{C}}(H) \to \mathcal{R}_{\mathbb{C}}(G)$$

(where we add the images); by Frobenius reciprocity these maps are adjoints with respect to the Schur bilinear form; cf. Definition 0.19 of Volume 1. By hypotheses, the natural map $\mathrm{Res} \colon \mathcal{R}_{\mathbb{C}}(G) \to \bigoplus_{H \in \mathcal{S}} \mathcal{R}_{\mathbb{C}}(H)$ is injective, so $\mathrm{Ind} \colon \bigoplus_{H \in \mathcal{S}} \mathcal{R}_{\mathbb{C}}(H) \to \mathcal{R}_{\mathbb{C}}(G)$ is onto, by Remark 0.20 of Volume 1. But $\mathcal{R}_{\mathbb{Q}}(G) \otimes_{\mathbb{Q}} \mathbb{C} \cong \mathcal{R}_{\mathbb{C}}(G)$; matching components in the tensor products (using Proposition 18.12), we see that $\mathrm{Ind} \colon \bigoplus_{H \in \mathcal{S}} \mathcal{R}_{\mathbb{Q}}(H) \to \mathcal{R}_{\mathbb{Q}}(G)$ is onto.

□

Theorem 20.44 follows, since the set of cyclic subgroups of G satisfies the hypothesis of Theorem 20.46. Furthermore, every irreducible complex character of a cyclic group has degree 1. A deeper theorem, by Brauer, that every character is a $\mathbb{Z}$-linear combination of characters induced from characters of degree 1, involves a broader class of subgroups; cf. [**Jac5**].

Chapter 21

Lie Algebras and Other Nonassociative Algebras

The theory of Lie algebras is the outcome of a successful effort to develop an algebraic structure theory from a previously existing mathematical notion, namely Lie groups, which we encountered in Remark 19A.22. Ironically, nowadays the reference to Lie groups is often replaced by algebraic groups (Appendix 19B) whose theory, like that of abstract Lie algebras, also was developed in the twentieth century; cf. Appendix 21A.

Our aim in this chapter is to develop from scratch the basic algebraic theory for f.d. (i.e., finite-dimensional) Lie algebras over a field F that often is algebraically closed of characteristic 0. The prerequisites from Volume 1 are minimal — some matrix theory, including the Jordan decomposition (Chapter 2) and bilinear forms (Chapter 0), and rudimentary facts about derivations (Appendix 6B) are enough to develop the basic structure theory of f.d. Lie algebras, and Cartan's classification of the f.d. semisimple Lie algebras requires the use of quadratic forms (Appendix 0A). The classification of Cartan matrices in terms of Dynkin diagrams is independent of the rest of the theory and also has application to Appendix 25C, so we treat it separately in Chapter 22.

Lie algebras also are studied through their universal (associative) enveloping algebras, described in Appendix 21C, along with the quantum analogs.

Lie algebras also serve as our first introduction to nonassociative algebras, and in Appendix 21B we deal with other nonassociative algebras such as alternative algebras and Jordan algebras. In Appendix 23B, we see how identities of Lie algebras play a key role in group theory, in particular with Zelmanov's solution of the restricted Burnside problem.

We start by placing Lie algebras in the more general framework of nonassociative algebras.

Definition 21.0. A **nonassociative algebra** over a commutative ring C is a C-module A together with a C-bilinear multiplication $A \times A \to A$ that is distributive over addition and satisfies the axiom

$$c(a_1a_2) = (ca_1)a_2 = a_1(ca_2), \qquad \forall c \in C,\ a_i \in A$$

(which is the same as in Definition 5.1 of Volume 1). (Thus a nonassociative algebra induces a map $A \otimes_C A \to A$; cf. Remark 18.19.)

"Nonassociative" more precisely means "not necessarily associative"; note that the existence of a unit element 1 is not required. Thus, we see that an associative algebra is merely a nonassociative algebra with 1 whose multiplication is associative, if the reader will forgive the abuse of language. For the remainder of this chapter, *algebra* denotes "nonassociative algebra." For example, any C-module A can be made into an algebra via the **trivial** multiplication $ab = 0$ for all $a, b \in A$; although structurally uninteresting, the trivial algebra plays an important role at times.

Let us mimic the associative structure theory. A **homomorphism** $f\colon A \to B$ of algebras is a C-module map satisfying $f(a_1a_2) = f(a_1)f(a_2)$. An **ideal** I of A is an additive subgroup of A such that $Ia \subseteq I$ and $aI \subseteq I$, $\forall a \in A$. The ideals are precisely the kernels of homomorphisms; cf. Exercise 21B.1. An algebra without proper nonzero ideals is called **simple**.

The classical study of nonassociative algebras commences with various subsets of matrices $M_n(F)$ closed under suitable nonassociative multiplications. Two of the most prominent such subsets involve the matrix transpose (t). Let K be the set of skew-symmetric matrices, and S be the set of symmetric matrices. Although neither K nor S is closed under the usual matrix multiplication, K is closed under a new multiplication given by $(a, b) \mapsto ab - ba$, and S is closed under the multiplication $(a, b) \mapsto ab + ba$. This gives rise to two kinds of nonassociative algebras, **Lie algebras** and **Jordan algebras**, respectively. We already came up with the Lie product $[a, b] = ab - ba$ on $M_n(F)$ in Remark 19A.22, but now we start over again with a self-contained exposition that is more elementary.

Lie algebras

Definition 21.1. A **Lie algebra** over a commutative (associative) ring C is a nonassociative C-algebra L whose multiplication $(a, b) \mapsto [ab]$, called **Lie multiplication**, satisfies the following two conditions for all $a, b, c \in L$:

(i) (**Anticommutativity**) $[aa] = 0$;

(ii) (**Jacobi identity**) $[[ab]c] + [[bc]a] + [[ca]b] = 0$.

Occasionally we write $[a, b]$ instead of $[ab]$ to avoid ambiguity.

A **Lie subalgebra** of a Lie algebra L is a C-submodule of L that is closed under the given Lie multiplication (and is thus a Lie algebra in its own right).

In most of the literature, Lie algebras are denoted by $\mathfrak{g}$, to emphasize their connection with Lie groups. We prefer L, for "Lie."

Remark 21.2. Anticommutativity shows that a nonzero Lie algebra L **cannot** possess a unit element 1, since otherwise $1 = [11] = 0$. Furthermore, for all a, b in L, anticommutativity implies that

$$0 = [a + b, a + b] - [aa] - [bb] = [ab] + [ba],$$

so $[ab] = -[ba]$.

Hence we can rewrite the Jacobi identity as

$$[[ab]c] = [a[bc]] - [b[ac]]. \tag{21.1}$$

Example 21.3. For any associative algebra R, we define the **Lie commutator**

$$[a, b] = ab - ba,$$

with respect to which R becomes a Lie algebra, denoted as R^-.

One of the most useful (and straightforward) properties of Lie commutators in an associative algebra R is the following equality:

$$[a, bc] = [a, b]c + b[a, c], \qquad \forall a, b, c \in R. \tag{21.2}$$

A special case: If $[a, c] = 0$, then $[a, bc] = [a, b]c$. Furthermore, (21.2) shows that R^- is indeed a Lie algebra. The Jacobi identity (21.1) follows from the computation:

$$\begin{aligned} [a, [b, c]] &= [a, bc] - [a, cb] \\ &= [a, b]c + b[a, c] - [a, c]b - c[a, b] \\ &= [[a, b], c] + [b, [a, c]]. \end{aligned}$$

Remark 21.4. Given an element a in a Lie algebra L, we define the **adjoint map** $\mathrm{ad}_a \in \mathrm{End}_C L$ by $\mathrm{ad}_a(y) = [ay]$ for y in L. Then (21.1) becomes

$$\mathrm{ad}_{[ab]} = \mathrm{ad}_a \,\mathrm{ad}_b - \mathrm{ad}_b \,\mathrm{ad}_a \,. \tag{21.3}$$

In particular if $[ab] = 0$, then $\mathrm{ad}_a \,\mathrm{ad}_b = \mathrm{ad}_b \,\mathrm{ad}_a$.

The **adjoint algebra** $\mathrm{ad}\, L$ is defined as $\{\mathrm{ad}_a : a \in L\}$. Equation (21.3) shows that $\mathrm{ad}\, L$ is itself a Lie subalgebra of $(\mathrm{End}_C L)^-$.

(There is some ambiguity here. If a is in a Lie subalgebra H of L, ad_a could denote the adjoint map either in $\mathrm{End}\, H$ or $\mathrm{End}\, L$. To avoid confusion in such situations, we write $\mathrm{ad}_L H$ when we are considering the adjoints of elements of H acting on L. Thus $\mathrm{ad}\, L$ means $\mathrm{ad}_L L$.)

We are most interested in the case when the base ring is a field F and when L is f.d. (finite-dimensional) over F. Here are some examples of Lie algebras over a field.

Example 21.5. (i) A Lie algebra L is called **Abelian** when the Lie product is trivial, i.e., $[ab] = 0$ for all $a, b \in L$. Obviously, any 1-dimensional Lie algebra over a field is Abelian, since $[aa] = 0$. For any associative algebra R, the Lie algebra R^- is Abelian iff R is commutative.

(ii) There is a nonabelian two-dimensional Lie algebra with base $\{a, b\}$, whose Lie product is given by $[ab] = a$. This is the only two-dimensional Lie algebra up to isomorphism; cf. Exercise 1.

(iii) The most famous case of Example 21.3 is $M_n(F)^-$, which is denoted as $gl(n, F)$; accordingly, the Lie subalgebras of $gl(n, F)$ are called **linear**. In Exercise 21A.6(i), $gl(n, F)$ is displayed as the Lie algebra of the algebraic group $\mathrm{GL}(n, F)$.

(iv) If $(R, *)$ is any associative algebra with involution, then the set of skew-symmetric elements forms a Lie subalgebra of R^-.

(v) The set of upper triangular matrices is a Lie subalgebra of $gl(n, F)$.

(vi) The set of strictly upper triangular matrices (i.e., zero on the diagonal) is a Lie subalgebra of the Lie algebra of (v).

However, we feature the following examples.

Example 21.6. The **classical** Lie algebras, called A_n, B_n, C_n, D_n, are of special import in the structure theory. We introduce them from our current, elementary, point of view; they are realized in Exercise 21A.6 as Lie algebras of algebraic groups.

A_n, also called $sl(n + 1, F)$, is the Lie subalgebra of $gl(n + 1, F)$ comprised of matrices of trace 0. (This is closed under Lie multiplication since

$\operatorname{tr}[a,b] = \operatorname{tr}(ab) - \operatorname{tr}(ba) = 0$.) As a vector space, $sl(n+1,F)$ has a base

$$\{e_{ij} : i \neq j\} \cup \{e_{ii} - e_{i+1,i+1} : 1 \leq i \leq n\};$$

hence $\dim sl(n+1,F) = (n+1)^2 - 1 = n^2 + 2n$.

B_n, C_n, and D_n are each constructed as instances of Example 21.5 (iv), where $(*)$ is an involution on $M_k(F)$. Taking $(*)$ to be the transpose on $M_k(F)$, we get the Lie algebra consisting of all skew-symmetric matrices, which is a Lie subalgebra of $sl(k,F)$. When $k = 2n+1$ is odd, this Lie algebra is called B_n; when $k = 2n$ is even, it is called D_n. Hence

$$\dim B_n = \frac{(2n+1)2n}{2} = 2n^2 + n, \quad \dim D_n = \frac{2n(2n-1)}{2} = 2n^2 - n.$$

Taking $(*)$ to be the canonical symplectic involution on $M_{2n}(F)$ (cf. Example 14.33(iii)), we get the Lie subalgebra C_n, or $sp(n,F)$, of $sl(n,F)$ consisting of matrices of the form $\begin{pmatrix} S & T \\ U & -S^t \end{pmatrix}$, where $T^t = T$ and $U^t = U$. Thus

$$\dim C_n = n^2 + 2\frac{n}{2}(n+1) = 2n^2 + n = \dim B_n.$$

A more sophisticated way of describing B_n, C_n, and D_n is given in Exercises 8 and 9.

Let us see how this works for $n = 1$. A base for $A_1 = sl(2,F)$ is

$$e = e_{12}, \quad f = e_{21}, \quad h = e_{11} - e_{22};$$

they satisfy the relations

$$[ef] = h; \quad [eh] = -2e; \quad [fh] = 2f. \tag{21.4}$$

As we shall see, the Lie algebra $sl(2,F)$ plays a special role in the theory, for f.d. semisimple Lie algebras over an algebraically closed field of characteristic 0 can be built by combining copies of $sl(2,F)$ (cf. Remark 21.81 and Theorem 21.108).

The Lie algebra D_1 is 1-dimensional with base $e_{12} - e_{21}$ and thus is Abelian; B_1 and C_1 are both 3-dimensional and isomorphic to A_1.

Let us now consider Lie algebras in terms of the general algebraic structure theory sketched after Definition 21.0. **Lie homomorphisms** of Lie algebras are C-module maps $f\colon L_1 \to L_2$ satisfying $f[ab] = [f(a)f(b)]$ for all $a, b \in L_1$. A **Lie ideal** of L is a C-submodule I such that $[aI] \subseteq I$ for all $a \in L$. (Anticommutativity yields $[Ia] \subseteq I$.) We write $I \lhd L$, relying on the context to resolve the ambiguity of whether we are considering ideals of Lie

algebras or associative algebras. One important example of a Lie ideal is $\{a \in L : [aL] = 0\}$, called the **center** $Z(L)$ of L. The sum of Lie ideals and the intersection of Lie ideals are Lie ideals, by an easy verification.

The kernel of any Lie homomorphism is a Lie ideal. Conversely, if $I \lhd L$, then L/I is a Lie algebra under the Lie multiplication $[a+I, b+I] = [ab]+I$, and the natural map $L \to L/I$ is a Lie homomorphism with kernel I. The analogs of the Noether structure theorems (cf. Theorems 0.1 and 0.2 of Volume 1) then hold for Lie algebras; we leave the verifications to the reader.

Example 21.7. Another way of constructing Lie algebras plays a special role in the theory. Recall from Volume 1 that a derivation of a C-algebra R (not necessarily associative) is a C-module map $\delta\colon R \to R$ satisfying

$$\delta(ab) = \delta(a)b + a\delta(b), \qquad \forall a, b \in R. \tag{21.5}$$

Many properties we proved earlier still hold when R is nonassociative. In particular, the set of C-derivations of R is a Lie subalgebra of $(\operatorname{End}_C R)^-$, denoted $\operatorname{Deriv}(R)$. (The easy verification was already given in Proposition 10A.3 of Volume 1.)

In the notation of Lie multiplication, (21.5) is rewritten as

$$\delta([ab]) = [\delta(a)b] + [a\delta(b)], \qquad \forall a, b \in L. \tag{21.6}$$

Remark 21.7′.

(i) For any $c \in L$, ad_c defines a derivation (seen at once by (21.1)); such a derivation is called **inner**.

(ii) The same argument shows that ad_c defines a derivation on any Lie ideal I of L, although this derivation need not be inner with respect to I.

Remark 21.8. $\operatorname{ad} L \lhd \operatorname{Deriv}(L)$. Indeed,

$$[\delta, \operatorname{ad}_a](b) = \delta[ab] - [a\delta(b)] = [\delta(a)b]$$

for any $\delta \in \operatorname{Deriv} L$ and $a, b \in L$, proving that

$$[\delta, \operatorname{ad}_a] = \operatorname{ad}_{\delta(a)}. \tag{21.7}$$

Example 21.9. For any Lie algebra L, there is a Lie homomorphism $\varphi\colon L \to \operatorname{ad} L$ given by $a \mapsto \operatorname{ad}_a$, in view of Equation (21.3). Clearly $\ker\varphi = Z(L)$, implying that

$$\operatorname{ad} L \cong L/Z(L).$$

As in the associative theory, the ideal-theoretic approach leads us to the following basic notion.

Definition 21.10. A Lie algebra L is **simple** if L has no Lie ideals other than 0 and L itself.

Although the basic structure theory can be carried out over an arbitrary base ring C, from now on we assume that L is a Lie algebra over a field F. One of our main goals will be to characterize the simple f.d. Lie algebras over an algebraically closed field F of characteristic 0. Note that the analogous question for associative algebras is rather easy; the Wedderburn-Artin Theorem implies that any f.d. simple associative algebra over such F has the form $M_n(F)$. The situation is much subtler for Lie algebras:

Example 21.10′. In characteristic $\neq 2$, $A_n, B_n, C_n (n \geq 3), D_n (n \geq 3)$ are all simple Lie algebras; we verify A_n now and leave the others for Exercise 10.

Any Lie ideal $I \neq 0$ of $A_n = sl(n+1, F)$ contains some nonzero matrix $a = \sum_{ij} \alpha_{ij} e_{ij}$; changing base if necessary, one may assume that $\alpha_{12} \neq 0$. Then

$$[e_{21} a] = \sum_j \alpha_{1j} e_{2j} - \sum_i \alpha_{i2} e_{i1},$$

implying that

$$e_{21} = \frac{1}{2\alpha_{12}} [[e_{21} a] e_{21}] \in I.$$

For $n = 1$, we get $e_{11} - e_{12} = [e_{12} e_{21}] \in I$, so $e_{12} = \frac{1}{2}[e_{11} - e_{22}, e_{12}] \in I$. Thus, we may assume $n \geq 2$; $e_{i1} = [e_{i2} e_{21}] \in I$ for all $i \notin 1$, and then $e_{ij} = [e_{i1} e_{1j}] \in I$ for all $i \neq \{1, j\}$. But then $e_{1j} = [e_{1i} e_{ij}] \in I$. Hence, $e_{ij} \in I$ for all $i \neq j$, and finally,

$$e_{ii} - e_{jj} = [e_{ij} e_{ji}] \in I.$$

(A similar proof works for $n \geq 2$ in all characteristics. For $n = 1$, we need to assume characteristic $\neq 2$; cf. Exercise 12. D_2 is *never* simple; cf. Exercise 13.)

Herstein proved a series of striking theorems that show the simplicity of such Lie algebras in a very general context; cf. Exercises 14–28 in the noninvolutory case and Herstein [**Hers3**] for more details in the involutory case. One can get a shorter and more conceptual proof of Herstein's theorems by means of Jacobson's Density Theorem; another more Lie-theoretic approach was taken by Benkart [**Ben**] using "inner ideals."

Remark 21.11. Given a Lie subalgebra A of L, define its **normalizer** $N_L(A)$ to be $\{b \in L : [bA] \subseteq A\}$. $N_L(A)$ is a Lie subalgebra of L by the Jacobi identity, and $A \lhd N_L(A)$ by definition.

Lie representations

Definition 21.12. A **Lie representation** $\rho\colon L \to \operatorname{End} V$ is a Lie homomorphism $\rho\colon L \to (\operatorname{End} V)^-$ for a vector space V over a field F. The Lie representation ρ is **finite-dimensional** (f.d.) of **degree** n if $\dim_F V = n$; then we have $\rho\colon L \to gl(n, F)$. A representation ρ is called **faithful** if $\ker \rho = 0$.

It would be difficult to overestimate the importance of representations in the theory of Lie algebras.

Example 21.12′. Given a Lie algebra L, define the **adjoint representation** $L \to (\operatorname{End} L)^-$ by $a \mapsto \operatorname{ad}_a$; this is a Lie representation by (21.3).

The adjoint representation plays an analogous role to the regular representation of groups and of algebras. However, there is a mild change from group representations and the associative theory.

Remark 21.13. Whereas the regular representation of an associative algebra is faithful, the kernel of the adjoint representation of a Lie algebra L is $Z(L)$. Thus $L/Z(L)$ is naturally isomorphic to a Lie subalgebra of $(\operatorname{End} L)^-$.

A deep theorem of Ado (in characteristic 0) and Iwasawa (in characteristic p) says that every f.d. Lie algebra does have a faithful (f.d.) representation, but the proof requires additional techniques; a proof of Ado's theorem is sketched in Exercises 21C.9–21C.13.

Lie modules.

Just as in group representation theory, we study representations by translating to modules, but we need a Lie-theoretic version. Since we take our Lie algebra L over a field F, we may bring in linear algebra.

Definition 21.14. A **Lie module** over a Lie algebra L is an F-vector space V endowed with scalar multiplication $L \times V \to V$ satisfying

$$[ab]v = a(bv) - b(av)$$

for all $a, b \in L$, $v \in V$. In this case, we also say L **acts on** V. We say that the action is **trivial** if $av = 0$ for all a in L and v in V. For convenience, we write $a^2 v$ for $a(av)$, and inductively $a^m v$ for $a(a^{m-1} v)$.

A Lie **submodule** of V is an F-subspace that is closed under the action of L. The Lie module V is called **simple** if it has no proper nonzero Lie submodules.

Example 21.15. One obvious example of a Lie module is L itself, with the adjoint action given by the usual Lie multiplication. (This follows from (21.3).) The Lie L-submodules of L are precisely the Lie ideals of L.

There is a 1:1 correspondence between Lie representations of degree n and Lie modules that are of dimension n as vector spaces. Namely, given a Lie representation $\rho\colon L \to gl(n,F)$, L acts on $F^{(n)}$ via $av = \rho(a)(v)$ for $a \in L$ and $v \in F^{(n)}$. Conversely, given a Lie module structure on $V = F^{(n)}$, define $\rho\colon L \to \operatorname{End}_F V$ by taking $\rho(a)$ to be the map $v \mapsto av$; then

$$\rho([ab])v = [ab]v = a(bv) - b(av) = [\rho(a), \rho(b)]v,$$

implying that ρ is a representation. This enables us to study representations in terms of the structure theory of Lie modules.

Digression. The same line of reasoning also gives us a way of verifying that a candidate L for a Lie algebra is in fact a Lie algebra. Suppose that we have a 1:1 map $L \to (\operatorname{End}_F V)^-$. Then the Jacobi identity can be transferred to L from the known Lie structure of $(\operatorname{End}_F V)^-$.

Here is a useful way of producing Lie submodules and Lie ideals.

Remark 21.16. (i) If V is a Lie module over L and $A \lhd L$, then AV is a Lie submodule of V. (Indeed, if $b \in L$, $a \in A$, and $v \in V$, then

$$b(av) = [ba]v + a(bv) \in AV + AV \subseteq AV.)$$

(ii) Let us interpret this for $V = L$, using Example 21.15. Given subsets S_1, S_2 of L, define $[S_1S_2]$ to be the F-subspace of L spanned by all $\{[ab] : a \in S_1,\ b \in S_2\}$. Then (i) says the Lie product $[I_1I_2]$ of Lie ideals I_1, I_2 is a Lie ideal.

Surprisingly, trivial modules play a role, because of the following observation.

Remark 21.17. $[LL]$ acts trivially on any Lie module V (over L) that is one-dimensional over F. In particular, if $L = [LL]$, then every degree 1 representation of L is trivial. (Indeed, suppose $a, b \in L$, and write $V = Fv$. Then $av = \alpha v$ and $bv = \beta v$ for suitable $\alpha, \beta \in F$, so

$$[a,b]v = a(\beta v) - b(\alpha v) = (\alpha\beta - \beta\alpha)v = 0.)$$

Remark 21.18. (i) Suppose V is a Lie module over L. Define

$$\operatorname{Ann}_L V = \{a \in L : aV = 0\},$$

a Lie ideal of L (for if $aV = 0$, then $[ab]v = a(bv) - b(av) = 0 - b0 = 0$, $\forall v \in V$).

It follows that V is also a Lie module over $L/\operatorname{Ann}_L V$, where we define $(a + \operatorname{Ann}_L V)v = av$, and V has the same Lie submodules over both Lie algebras.

(ii) If V is a Lie module over L, then $\operatorname{End}_F V$ is also a Lie module over L, under the multiplication

$$[af](v) = af(v) - f(av). \tag{21.8}$$

(Indeed, for $a, b \in L$, $f \in \operatorname{End}_F V$, and $v \in V$, we have

$$\begin{aligned}[a[bf]](v) &= a([bf]v) - [bf](av)\\ &= a((bf(v) - f(bv)) - (bf(av) - f(b(av)))\\ &= a(bf(v)) - af(bv) - bf(av) + f(b(av));\end{aligned}$$

and likewise

$$[b[af]](v) = b(af(v)) - bf(av) - af(bv) + f(a(bv)),$$

so

$$\begin{aligned}[a[bf]](v) - [b[af]](v) &= a(bf(v)) + f(b(av)) - b(af(v)) - f(a(bv))\\ &= [ab]f(v) - f([ab]v) = [[ab]f](v),\end{aligned}$$

as desired.)

Remark 21.19. Given a Lie representation $\rho\colon L \to \operatorname{End}_F V$, let $\bar{L}$ denote the associative subalgebra (with 1) of $\operatorname{End}_F V$ generated by $\rho(L)$. Then V is naturally an $\bar{L}$-module. The Lie L-submodules of V are precisely the Lie $\rho(L)$-submodules (cf. Remark 21.18(i)), which are the $\bar{L}$-submodules of V. This easy observation enables us to translate much of the module theory over associative rings to Lie module theory and thus to Lie representation theory, although direct proofs may be more satisfying aesthetically. (For example, the Jordan-Hölder theory of composition series of modules works analogously for Lie modules; cf. Exercise 31.)

Viewing the above setup more abstractly, suppose R is an associative algebra, and L is a Lie subalgebra of R^-. Let $\bar{L}$ denote the associative subalgebra (including 1) of R generated by L. Many important connections link the Lie structure of L to the associative structure of $\bar{L}$. The usual notation for multiplication denotes the associative multiplication in R.

PROPOSITION 21.20. *Suppose I is a Lie ideal of $L \subseteq R^-$. Then $\bar{L}I = I\bar{L} \lhd \bar{L}$. In particular, if $I^n = 0$, then $(\bar{L}I)^n = 0$.*

Proof. Everything follows from the claim $\bar{L}I = I\bar{L}$, and by symmetry we need only prove that $I\bar{L} \subseteq \bar{L}I$. Thus, for $a \in I$ and $b = b_1 \cdots b_t$ for $b_i \in L$, we need to show that $ab \in \bar{L}I$. Note that (21.2) implies inductively that

$$[a, b_1 \cdots b_t] = \sum b_1 \cdots b_{i-1}[ab_i]b_{i+1} \cdots b_t.$$

Hence

$$ab = ba + [ab] = ba + \sum_i b_1 \cdots b_{i-1}[ab_i]b_{i+1} \cdots b_t$$

which belongs to $\bar{L}I$, seen by induction on t (noting that $[ab_i] \in I$). □

The Jordan decomposition.

We recall the Jordan decomposition $a = \mathbf{s} + \mathbf{n}$ of a transformation a into its semisimple and nilpotent parts, which commute with each other (cf. Theorem 2.72ff of Volume 1); the Jordan decomposition is unique by Theorem 2.75 of Volume 1.

PROPOSITION 21.21. *Suppose L is a Lie subalgebra of $gl(n, F)$, with F algebraically closed.*

(i) *If $a \in L$ is semisimple as a linear transformation, then ad_a also is semisimple; moreover, if $\{\alpha_i : 1 \leq i \leq n\}$ are the eigenvalues of a, then $\{\alpha_i - \alpha_j : 1 \leq i, j \leq n\}$ are the eigenvalues of ad_a.*

(ii) *If $a \in L$ is nilpotent as a transformation in $gl(n, F)$, then ad_a is nilpotent.*

(iii) *If $a = \mathbf{s} + \mathbf{n}$ is the Jordan decomposition of a, then*

$$\mathrm{ad}_a = \mathrm{ad}_{\mathbf{s}} + \mathrm{ad}_{\mathbf{n}}$$

is the Jordan decomposition of ad_a.

Proof. (i) We can write $a = \sum \alpha_i e_{ii}$ with respect to a suitable base $e_1, \ldots, e_n$ of $F^{(n)}$. But then

$$\mathrm{ad}_a(e_{ij}) = (\alpha_i - \alpha_j)e_{ij},$$

implying that ad_a is diagonal with respect to the base $\{e_{ij} : 1 \leq i, j \leq n\}$ of $\mathrm{End}\, F^{(n)}$.

(ii) $\mathrm{ad}_a = \ell_a - r_a$, where these denote respectively the left and right multiplication maps by a; ℓ_a and r_a commute, so

$$\mathrm{ad}_a^{2n-1} = (\ell_a - r_a)^{2n-1} = \sum_{j=0}^{2n-1} \binom{2n-1}{j} \ell_a^j r_a^{2n-1-j} = 0.$$

(iii) $\mathrm{ad}_a = \mathrm{ad}_{\mathbf{s}+\mathbf{n}} = \mathrm{ad}_{\mathbf{s}} + \mathrm{ad}_{\mathbf{n}}$, where $\mathrm{ad}_{\mathbf{s}}$ is semisimple by (i) and $\mathrm{ad}_{\mathbf{n}}$ is nilpotent by (ii). Furthermore $\mathrm{ad}_{\mathbf{s}}$ and $\mathrm{ad}_{\mathbf{n}}$ commute by Remark 21.4, so we have the Jordan decomposition. □

Restricted Lie algebras in characteristic > 0.

Soon we are to focus on characteristic 0, partly because the situation in characteristic $p > 0$ becomes far more intricate; cf. Block [**Bl**]. However, some useful extra structure becomes available in characteristic p.

Since Lie algebras are intimately related to derivations, it is worth noting that if $\delta : R \to R$ is a derivation of an arbitrary algebra of characteristic p, then δ^p is also a derivation; cf. Corollary 6B.6. Thus $\mathrm{Deriv}\, R$ has another operation, $\mathrm{ad}_a \mapsto \mathrm{ad}_a^p$. In particular, if L is a Lie algebra of characteristic p, then ad_a^p is a derivation for any $a \in L$.

Often, ad_a^p is the adjoint of an element of L, denoted $a^{[p]}$, which is uniquely defined when $Z(L) = 0$. The map $a \mapsto a^{[p]}$ defines an important unary operation.

Definition 21.21′. A linear Lie subalgebra $L \subseteq gl(n, F)$ of characteristic p is called **restricted** if L has a unary operation $a \mapsto a^{[p]}$ such that the natural injection $\rho : L \to M_n(F)^-$ satisfies $\rho(a^{[p]}) = \rho(a)^p$.

The definition for arbitrary Lie algebras is given in Exercise 30. In some ways, restricted Lie algebras are more amenable to study than Lie algebras of characteristic 0; cf. Exercise 21C.8.

Nilpotent and solvable Lie algebras

So far, we have modelled the Lie theory after the associative theory. In many ways, the Lie theory is closer to group theory than to the theory of associative algebras. For example, we have:

Definition 21.22. Writing L^1 and $L^{(0)}$ each to denote L, we define inductively $L^{k+1} = [LL^k]$ and $L^{(k)} = [L^{(k-1)}, L^{(k-1)}]$.

The Lie algebra L is called **nilpotent** (of index k) if $L^k = 0$ for some k; L is called **solvable** if $L^{(k)} = 0$ for some k.

$L^2 = L^{(1)} = [L, L]$, which we also denote as L', is one of the main tools used in studying L.

Example 21.23. (i) $L' \lhd L$ by Remark 21.16(ii), and L/L' is Abelian. In particular, $L' = 0$ iff L is Abelian. More generally, for any $I \lhd L$, $(L/I)^k$ (resp. $(L/I)^{(k)}$) is the image of L^k (resp. $L^{(k)}$) in L/I. It follows that any

homomorphic image of a nilpotent (resp. solvable) Lie algebra is nilpotent (resp. solvable).

(ii) $L^{(k)}$ and L^k are Lie ideals for all k, by Remark 21.16(ii) and induction on k.

(iii) Clearly $L^{(k)} \subseteq L^{k+1}$, so all nilpotent Lie algebras are solvable.

Remark 21.24. $(L \otimes_F K)' = L' \otimes_F K$ for any field extension K of F. Consequently, nilpotence and solvability are preserved under base field extension. This often enables us to reduce to the case where F is algebraically closed.

Remark 21.25. Any F-subspace A of L containing L' is a Lie ideal of L, since $[LA] \subseteq L' \subseteq A$.

Example 21.26. (i) In the important case $L = gl(n, F)$, we have $L' = sl(n, F)$, which has codimension 1 in $gl(n, F)$. Indeed, any element of L' has trace 0, so $L' \subseteq sl(n, F)$; on the other hand, by Example 21.6, $sl(n, F)$ has a base comprised of Lie commutators $e_{ij} = [e_{ii}, e_{ij}]$ for $i \neq j$ and $e_{ii} - e_{i+1,i+1} = [e_{i,i+1}, e_{i+1,i}]$.

Furthermore, $\mathrm{tr}(\alpha I) = n\alpha$. If $\mathrm{char}(F) \nmid n$, this says $L' \cap F \cdot 1 = 0$, so $L = L' \oplus F$.

(ii) The Lie algebra of Example 21.5(vi) is nilpotent, as seen by direct computation and induction.

(iii) Example 21.5(ii) is solvable but not nilpotent. Likewise, the Lie algebra L of Example 21.5(v) is solvable (but not nilpotent) since L' is the nilpotent Lie algebra of Example 21.5(vi).

Engel's and Lie's Theorems.

Perhaps surprisingly, the examples of f.d. nilpotent and solvable Lie algebras given in Example 21.26 are the only ones in some sense, as we see now via important structural theorems of Engel and Lie. We start with nilpotent Lie algebras, for which we have already done the necessary work.

THEOREM 21.27. *Suppose L is a Lie subalgebra of R^- for a f.d. associative algebra R, and ad_a is nilpotent for every $a \in L$. Then $\mathrm{ad}\, L$ is nilpotent under the multiplication of R, and L is a nilpotent Lie algebra.*

Proof. Viewed in $\mathrm{End}_F R$, $\mathrm{ad}\, L$ satisfies the condition of Theorem 15.23 (taking $\nu = -1$), so $\mathrm{ad}\, L$ is nilpotent with respect to the multiplication of R, implying that the Lie algebra L is nilpotent. □

We can view Theorem 21.27 more explicitly. We say that a subset S of L **acts nilpotently** on an L-module V if $S^k V = 0$ for some k. (Then there exists such $k \leq \dim V$, by Lemma 15.22.)

LEMMA 21.28. *Suppose the Lie algebra L acts nilpotently on an L-module V. Then there exists $v \neq 0$ in V such that $Lv = 0$.*

Proof. For k minimal with $L^k V = 0$, take any $v \neq 0$ in $L^{k-1}V$. □

THEOREM 21.29 (ENGEL'S THEOREM). *If $L \subseteq gl(n, F)$ is a Lie algebra of nilpotent transformations, then, under a suitable choice of base, L becomes a Lie subalgebra of the algebra of strictly upper triangular matrices.*

Proof. Let $V = F^{(n)}$. By Theorem 21.27, L acts nilpotently on V, so there is some $0 \neq v \in V$ with $Lv = 0$. Then Fv is a Lie submodule of V, so L acts nilpotently on $V_1 = V/Fv$. By induction on n, we can find a base $\bar{b}_1, \ldots, \bar{b}_{n-1}$ of V_1 such that $L\bar{b}_i \subseteq \sum_{j>i} F\bar{b}_j$ for $1 \leq i \leq n-1$. Writing $\bar{b}_i = b_i + Fv$, we let $b_n = v$ and see that $Lb_i \subseteq \sum_{j>i} Fb_j$ for $1 \leq i \leq n$, as desired.

□

Warning. Engel's Theorem does NOT say that any f.d. nilpotent Lie subalgebra L of $gl(n, F)$ can be put into strictly upper triangular form, since the action of L on $F^{(n)}$ need not be nil. For example, F itself is Abelian (as a Lie algebra), but acts as scalar multiplication. However, the theorem does say that $\mathrm{ad}\, L$ can be put into strictly upper triangular form.

Turning to solvable algebras, we have the following important property drawn from its analog in group theory.

PROPOSITION 21.30. *Suppose I is a Lie ideal of L. The Lie algebra L is solvable iff the Lie algebras I and L/I are both solvable.*

Proof. ($\Rightarrow$) Clear, in view of Example 21.23.

($\Leftarrow$) By hypothesis $(L/I)^{(k)} = 0$ for some k, i.e., $L^{(k)} \subseteq I$. But $I^{(m)} = 0$ for some m, so $L^{(k+m)} = 0$. □

We say that L **acts solvably** on an L-module V if $L^{(k)}V = 0$ for some k. (By Lemma 15.22, one could take $k \leq \dim V$.)

LEMMA 21.31. *Suppose V is a f.d. vector space over an algebraically closed field F and is a simple Lie module over a Lie algebra L. If L acts solvably on V, then $\dim_F V = 1$.*

Proof. For each k, $L^{(k)}V$ is a Lie submodule of V by Remark 21.16, and thus is V or 0. By hypothesis we must have $L^{(k+1)}V < L^{(k)}V$ for some k, so replacing L by $L^{(k)}$ we may assume that $L'V = 0$. Then V has the same Lie submodule structure over L/L', and we may assume that L is Abelian. Take an eigenvector v of some element $a \neq 0$ in L. Writing $av = \alpha v$, let $W = \{w \in V : aw = \alpha w\}$, a nonzero subspace of V. Then $LW \subseteq W$ is a nonzero L-submodule of V, implying that $W = V$. Hence L acts as scalar multiplication on V, so clearly $\dim V = 1$. □

THEOREM 21.32 (LIE'S THEOREM). *Suppose L is a Lie subalgebra of $gl(n, F)$ acting solvably on $F^{(n)}$, with F an algebraically closed field. Then $F^{(n)}$ (as a Lie module over L) has a composition series each of whose factors has dimension* 1*; hence the elements of L act in simultaneous upper triangular form with respect to a suitable base of $F^{(n)}$.*

Proof. Take a nonzero simple Lie submodule V of $F^{(n)}$ over L. By Lemma 21.31, $V = Fv$ for $0 \neq v \in V$; taking a base of $F^{(n)}$ including v and passing to $F^{(n)}/V$, we apply induction on n. □

An interesting instance of this hypothesis, needed later, is when V itself is a f.d. Lie algebra and L is a f.d. solvable Lie subalgebra acting on V via the adjoint action.

This reasoning ties the solvability of L to the nilpotence of L'.

COROLLARY 21.33. *Suppose L is a Lie algebra acting on $V = F^{(n)}$, for F algebraically closed. Then L acts solvably on V iff L' acts nilpotently on V.*

Proof. ($\Leftarrow$) Obviously, $L^{(n+1)}V \subseteq (L')^n V = 0$.

($\Rightarrow$) By Lemma 21.31, take $v \in V$ with $Lv = 0$. L acts solvably on V/Fv, implying that $(L')^{n-1}(V/Fv) = 0$ by induction. But then $(L')^{n-1}V \subseteq Fv$, so $(L')^n V = 0$. □

An alternative proof: Write L in upper triangular form, using Lie's Theorem; $(\operatorname{ad} L)'$ is strictly upper triangular (by inspection, since the diagonal is Abelian), implying that $(\operatorname{ad} L)'$ is nilpotent.

Using the adjoint representation, we conclude:

COROLLARY 21.34. *The following conditions are equivalent for a Lie algebra L:*

(i) *L is solvable;*
(ii) *$\operatorname{ad} L$ is solvable;*
(iii) *$(\operatorname{ad} L)'$ is nilpotent.*

Proof. (i) ⇒ (ii) Clear by Example 21.9.

(ii) ⇒ (iii) Tensoring by the algebraic closure of the base field F, in view of Remark 21.24, we may assume that F is algebraically closed. But then $\operatorname{ad} L$ may be viewed as a Lie algebra of upper triangular matrices, so we conclude with Corollary 21.33.

(iii) ⇒ (i) $\operatorname{ad} L \cong L/Z(L)$ is solvable, so L is solvable by Proposition 21.30. □

We extend these results in Exercises 39 and 50 by exploiting the structure of associative algebras.

Remark 21.35. Lie's Theorem can be reformulated for any Lie representation $\rho: L \to gl(n, F)$, where F is algebraically closed and L is solvable. Namely, $\rho(L)$ is also solvable, so can be put in simultaneous upper triangular form with respect to a suitable choice of base of $F^{(n)}$.

Cartan's first criterion for solvable algebras.

Our next goal is a criterion for solvability, which requires characteristic 0. We need another fact about matrices.

LEMMA 21.36. *Suppose $A \subseteq B$ are subspaces of $gl(n, F)$ with $\operatorname{char}(F) = 0$. Let $W = \{w \in gl(n, F) : \operatorname{ad}_w(B) \subseteq A\}$. If $a \in W$ satisfies $\operatorname{tr}(aw) = 0$, $\forall w \in W$, then the matrix a is nilpotent.*

Proof. Let $\alpha_1, \ldots, \alpha_n$ be the eigenvalues of a. After a change of base, we may assume that the semisimple part $\mathbf{s}$ of a is diagonal, with entries $\alpha_1, \ldots, \alpha_n$. By hypothesis, $\mathbb{Q} \subseteq F$. We claim that any linear functional $f: \sum_{i=1}^m \mathbb{Q}\alpha_i \to \mathbb{Q}$ must be 0; this means all $\alpha_i = 0$, as desired. Let us define w to be the diagonal matrix with entries $f(\alpha_1), \ldots, f(\alpha_n)$. By Lagrange interpolation (Remark 0.6 of Volume 1), there is a polynomial $p \in F[\lambda]$ having degree $\leq n^2$ such that

$$p(\alpha_i - \alpha_j) = f(\alpha_i - \alpha_j), \quad \forall i, j.$$

Taking $i = j$ we see that $p(0) = 0$, so p has constant term 0.

In view of Proposition 21.21(i), the eigenvalues of $\operatorname{ad}_{\mathbf{s}}$ are $\alpha_i - \alpha_j$, and the eigenvalues of ad_w are $f(\alpha_i) - f(\alpha_j) = p(\alpha_i - \alpha_j)$, proving that $\operatorname{ad}_w = p(\operatorname{ad}_{\mathbf{s}})$.

Since $\operatorname{ad}_{\mathbf{s}}$ (and thus ad_w) is a polynomial in ad_a without constant term, and since $\operatorname{ad}_a(B) \subseteq A$, we see that $\operatorname{ad}_w(B) \subseteq A$, so $w \in W$. Hence

$$0 = \operatorname{tr}(aw) = \sum_{i=1}^n \alpha_i f(\alpha_i),$$

so $0 = f(\sum \alpha_i f(\alpha_i)) = \sum f(\alpha_i)^2$, implying that each $f(\alpha_i) = 0$, as desired. □

To proceed further, we refine our study of representations by recalling the trace bilinear form on $M_n(F)$:

$$\langle v, w\rangle = \operatorname{tr}(vw).$$

This is transferred via the Lie representation $\rho\colon L \to gl(n,F)$ to the Lie algebra L, viz.

$$\langle a, b\rangle = \operatorname{tr}(\rho(a)\rho(b)), \qquad \forall a, b \in L.$$

Remark 21.37. For any $n \times n$ matrices u, v, w, one has

$$\operatorname{tr}([uv]w) = \operatorname{tr}(uvw) - \operatorname{tr}(vuw) = \operatorname{tr}(uvw) - \operatorname{tr}(uwv) = \operatorname{tr}(u[vw]). \tag{21.9}$$

Hence the trace bilinear form of a f.d. Lie representation ρ is **associative** in the sense that

$$\langle [ab], c\rangle = \langle a, [bc]\rangle, \qquad \forall a, b, c \in L. \tag{21.10}$$

THEOREM 21.38. *Suppose $L \subseteq gl(n,F)$ in characteristic 0 such that $\operatorname{tr}(aL') = 0$ for all $a \in L$. Then L' is a nilpotent Lie algebra.*

Proof. Let $W = \{w \in gl(n,F) : \operatorname{ad}_w(L) \subseteq L'\}$. Take an arbitrary element $a = \sum_i [a_i b_i] \in L'$. Equation (21.9) implies that, for each i,

$$\operatorname{tr}([a_i b_i]w) = \operatorname{tr}(a_i[b_i w]) \in \operatorname{tr}(a_i L') = 0,$$

by hypothesis. Hence $\operatorname{tr}(aW) = 0$, so each $a \in L'$ is nilpotent by Lemma 21.36, and we are done by Theorem 21.29. □

Now we utilize the adjoint representation.

Definition 21.39. The **Killing form** of a f.d. Lie algebra L is the trace bilinear form of the adjoint representation, i.e.,

$$\langle a, b\rangle = \operatorname{tr}(\operatorname{ad}_a \operatorname{ad}_b).$$

Remark 21.40. Equation (21.10) can be rewritten as

$$-\langle \operatorname{ad}_b a, c\rangle = \langle a, \operatorname{ad}_b c\rangle. \tag{21.11}$$

THEOREM 21.41 (CARTAN'S FIRST CRITERION). *Suppose L is a f.d. Lie algebra of characteristic 0. Then L is solvable iff its Killing form vanishes identically on L'.*

Proof. ($\Rightarrow$) By Corollary 21.34, $\operatorname{ad} L'$ is nilpotent and thus could be put in strictly triangular form; hence, $\operatorname{tr}(\operatorname{ad} L' \operatorname{ad} L') = 0$.

($\Leftarrow$) By Theorem 21.38, $(\operatorname{ad} L)'$ is nilpotent, implying that L is solvable by Corollary 21.34. □

The radical of a Lie algebra.

Let us delve deeper into the structure of f.d. Lie algebras.

PROPOSITION 21.42. *The sum $I_1 + I_2$ of two solvable Lie ideals is solvable.*

Proof. $(I_1 + I_2)/I_1 \cong I_2/(I_1 \cap I_2)$ is solvable, and $I_1 \cap I_2$ is solvable by Proposition 21.30, so $I_1 + I_2$ is solvable, again by Proposition 21.30. □

Definition 21.43. The **radical** $\operatorname{rad}(L)$ of a f.d. Lie algebra L is the (unique) largest solvable ideal of L.

Remark 21.43′. $\operatorname{rad}(L/\operatorname{rad}(L)) = 0$. (Indeed, if $I/\operatorname{rad}(L)$ is a solvable ideal of $L/\operatorname{rad}(L)$, then, by Proposition 21.30, I is a solvable ideal of L, implying that $I \subseteq \operatorname{rad}(L)$, and $I/\operatorname{rad}(L) = 0$.) If $\operatorname{rad}(L)^{(k)} = 0$, then $\operatorname{rad}(L)^{(k-1)}$ is an Abelian ideal of L.

(See Exercise 32 for the analog for nilpotent ideals. However, the corresponding assertion of Remark 21.43′ fails; cf. Exercise 33. We return to this issue in Appendix 23B.)

Semisimple Lie algebras

A Lie algebra L is **semisimple** if $\operatorname{rad}(L) = 0$. Our main interest is in semisimple Lie algebras, which play the parallel role to semisimple rings in the associative theory.

Remark 21.44. It follows from Remark 21.43′ that a f.d. Lie algebra L is semisimple iff it has no nonzero Abelian ideals. Note that any 1-dimensional Abelian Lie algebra is simple, but not semisimple.

In particular, if L is f.d. semisimple, then $Z(L) = 0$, so the adjoint representation of L is faithful.

Cartan's first criterion holds the key to semisimplicity. First we place the radical of the trace bilinear form in the context of Lie ideals.

LEMMA 21.45. *If $I \lhd L$, then $I^\perp \lhd L$ (taken with respect to the trace bilinear form of any representation).*

Proof. If $a \in I$, $b \in L$, and $c \in I^\perp$, then

$$\langle a, [bc]\rangle = \langle [ab], c\rangle = 0,$$

proving that $[LI^\perp] \subseteq I^\perp$. □

LEMMA 21.46. *Suppose L is a f.d. Lie algebra (of arbitrary characteristic). Then any Abelian Lie ideal I is contained in the radical of the Killing form of L. In particular, if the Killing form is nondegenerate, then the adjoint representation is faithful.*

Proof. Take $a \in I$ and $b \in L$ arbitrary. Then $\operatorname{ad}_a \operatorname{ad}_b \colon L \to I$, and

$$(\operatorname{ad}_a \operatorname{ad}_b)^2 \colon L \to [II] = 0.$$

Hence $0 = \operatorname{tr}(\operatorname{ad}_a \operatorname{ad}_b) = \langle a, b\rangle$, $\forall b \in L$. The last assertion follows, since $Z(L) = 0$. □

(See Exercise 41 for a generalization of this lemma.)

THEOREM 21.47 (CARTAN'S SECOND CRITERION). *A f.d. Lie algebra L of characteristic 0 is semisimple iff its Killing form is nondegenerate.*

Proof. ($\Rightarrow$) The radical I of the Killing form is a Lie ideal by Lemma 21.45; hence I is solvable by Cartan's first criterion, and thus is 0.

($\Leftarrow$) L has no nonzero Abelian ideals, by Lemma 21.46, and so is semisimple by Remark 21.44. □

One important application is a structural result. The direct sum $A \oplus B$ of two Lie algebras is a Lie algebra, with componentwise Lie multiplication $[(a_1, b_1), (a_2, b_2)] = ([a_1a_2], [b_1b_2])$; then the Lie structure of $L = A \oplus B$ is determined by the Lie structures on A and B. Note also that $A \oplus 0$ is a Lie ideal of L, which we identify with A, and likewise for $0 \oplus B$. From this point of view, we would like to know when we can decompose a given Lie algebra as a direct sum $A \oplus B$.

Remark 21.48. If $L = A \oplus B$ is a direct sum of Lie algebras, then any Lie ideal of A is a Lie ideal of L.

PROPOSITION 21.49. *Suppose L is any f.d. semisimple Lie algebra of characteristic 0. If $I \lhd L$, then $L = I \oplus I^\perp$ as Lie algebras, and in particular, I and $I^\perp \cong L/I$ are both semisimple.*

Proof. $I^\perp \lhd L$ by Lemma 21.45. Thus

$$\langle (I \cap I^\perp)', L\rangle \subseteq \langle I \cap I^\perp, [I \cap I^\perp, L]\rangle \subseteq \langle I \cap I^\perp, I \cap I^\perp\rangle = 0,$$

implying that $(I \cap I^\perp)' = 0$ by nondegeneracy of the Killing form. Hence $I \cap I^\perp$ is an Abelian ideal, which must be 0, implying that $L = I \oplus I^\perp$ by Remark 0.13 of Volume 1 (as vector spaces and thus as Lie algebras). Any solvable Lie ideal of I is a solvable Lie ideal of L and thus 0. Hence I (and likewise $I^\perp$) are semisimple, and clearly $I^\perp \cong L/I$. □

COROLLARY 21.50. *If L is a f.d. semisimple Lie algebra of characteristic 0 and $f \colon L \to L_1$ is a Lie homomorphism, then $L \cong \ker f \oplus f(L)$.*

Proof. $f(L) \cong L/\ker f \cong \ker f^\perp$. □

THEOREM 21.51. *Any f.d. semisimple Lie algebra L of characteristic 0 is a direct sum $\bigoplus S_i$ of simple nonabelian Lie subalgebras S_i, with each $S_i \lhd L$. Furthermore, each $S_i' = S_i$, and any Lie ideal of L is a direct sum of some of the S_i.*

Proof. If L is not itself simple, write $L = I \oplus I^\perp$ for some proper Lie ideal I and apply induction (on dimension) to both I and $I^\perp$ to obtain the first assertion. For each i, $S_i' \neq 0$ since otherwise S_i would be an Abelian Lie ideal of L, contrary to L semisimple; hence $S_i' = S_i$.

For any Lie ideal $A \lhd L$, write A_i for the projection of A to S_i. If $A_i \neq 0$, then $A_i = S_i$ and so $A_i = [AS_i] \subseteq A$. It follows that $A = \oplus A_i$. □

COROLLARY 21.52. *$L = L'$ for any f.d. semisimple Lie algebra L of characteristic 0.*

Proof. Write $L = \bigoplus S_i$ as in the theorem. Then $S_i' = S_i$, implying that $L' = \bigoplus S_i' = L$. □

COROLLARY 21.53. *The trace bilinear form of any representation ρ of a f.d. semisimple Lie algebra is nondegenerate.*

Proof. Passing to $L/\ker\rho$, which remains semisimple by Proposition 21.49, we may assume that ρ is faithful. The radical I of the trace bilinear form is a Lie ideal, by Lemma 21.45, and I' is nilpotent by Theorem 21.38; hence I is solvable and thus is 0. □

Here is another application of the Killing form.

THEOREM 21.54 (ZASSENHAUS). *Every derivation of a f.d. semisimple Lie algebra L of characteristic 0 is inner.*

Proof. Recall from Remark 21.8 that $\operatorname{ad} L \lhd \operatorname{Deriv} L$. Extending a base of $\operatorname{ad} L$ to $\operatorname{Deriv} L$, we see that the restriction of the Killing form of $\operatorname{Deriv} L$ to $\operatorname{ad} L$ is the Killing form of $\operatorname{ad} L$, which is nondegenerate by Theorem 21.47, since $\operatorname{ad} L \cong L$ by Lemma 21.46. But for any $\delta \in (\operatorname{ad} L)^\perp$, (21.7) implies $\operatorname{ad}_{\delta(a)} = [\delta, \operatorname{ad}_a] \in (\operatorname{ad} L)^\perp \cap \operatorname{ad} L = 0$, implying $\delta(a) = 0$. This proves that $(\operatorname{ad} L)^\perp = 0$, so

$$\operatorname{Deriv} L = \operatorname{ad} L \oplus (\operatorname{ad} L)^\perp = \operatorname{ad} L \oplus 0 = \operatorname{ad} L,$$

as desired. □

The Casimir element and Weyl's Theorem.

Since our best tools (Cartan's criteria, Zassenhaus' theorem) only work in characteristic 0, from now on we carry the assumption that $\operatorname{char}(F) = 0$.

Our next goal is to prove the important structure theorem of Weyl, that (in characteristic 0) every f.d. Lie representation of a f.d. semisimple Lie algebra is completely reducible. Although similar in language to the Wedderburn-Artin Theorem, this result is considerably more sublime, since the theorem fails for infinite-dimensional representations, and also fails in characteristic $\neq 0$, as shown in Jacobson [**Jac1**].

We start with a given Lie representation $\rho\colon L \to gl(m, F)$, which we assume is faithful since we may replace L by $L/\ker\rho$, in view of Proposition 21.49. Let $\dim L = n$. By Corollary 21.53, the trace bilinear form of ρ is nondegenerate on L, so any base $e_1, \dots, e_n$ of L has a dual base $e_1^*, \dots, e_n^* \in L$, i.e., $\langle e_i, e_j^* \rangle = \delta_{ij}$. These hypotheses and notation remain in effect throughout this discussion.

LEMMA 21.55. *Writing $[ae_i] = \sum \alpha_{ij} e_j$, then $[ae_i^*] = \sum -\alpha_{ji} e_j^*$. In particular, $\sum_i [ae_i] e_i^* = -\sum_i e_i [ae_i^*]$.*

Proof. $\alpha_{ij} = \langle [ae_i], e_j^* \rangle = -\langle [e_i a], e_j^* \rangle = -\langle e_i, [ae_j^*] \rangle$ by (21.10), so $-\alpha_{ij} = \langle [ae_j^*], e_i \rangle$. The first assertion follows at once by switching i and j, and the second assertion holds since

$$\sum_i [ae_i] e_i^* = \sum_{i,j} \alpha_{ij} e_j e_i^* = \sum_j e_j \sum_i \alpha_{ij} e_i^* = -\sum_j e_j [ae_j^*]. \qquad \square$$

Definition 21.56. The **Casimir element** c_ρ of ρ is $\sum_{i=1}^n \rho(e_i)\rho(e_i^*)$, computed in $M_m(F)$ (viewing $\rho(L) \subseteq gl(m, F) \subset M_m(F)$).

The Casimir element plays a critical role in the structure theory because of the following observation.

LEMMA 21.57. $\operatorname{tr}(c_\rho) = n$ *and* $[\rho(L), c_\rho] = 0$. *Viewing $V = F^{(m)}$ as an L-module by means of ρ, we have an L-module map $\ell_{c_\rho}\colon V \to V$ given by $v \mapsto c_\rho v$.*

Proof. $\operatorname{tr}(c_\rho) = \sum_{i=1}^n \operatorname{tr}(\rho(e_i)\rho(e_i^*)) = \sum_{i=1}^n \langle e_i, e_i^* \rangle = n$. For the second assertion, note that for any $a \in L$,

$$\begin{aligned}[\rho(a), c_\rho] &= \Big[\rho(a), \sum \rho(e_i)\rho(e_i^*)\Big] \\ &= \sum [\rho(a), \rho(e_i)]\rho(e_i^*) + \sum \rho(e_i)[\rho(a), \rho(e_i^*)] = 0,\end{aligned}$$

by Lemma 21.55.

Now $\ell_{c_\rho}(av) = c_\rho \rho(a)(v) = \rho(a) c_\rho(v) = a\ell_{c_\rho}(v)$. □

THEOREM 21.58 (WEYL'S THEOREM). *Any f.d. representation of a f.d. semisimple Lie algebra L (of characteristic 0) is completely reducible.*

Proof. We view $V = F^{(m)}$ as a Lie module via ρ, and want to prove that any Lie submodule W has a complement (as a Lie submodule). As noted above, we may assume that V is faithful over L. Let $\bar{V} = V/W$.

Special Case. Assume that $\dim \bar{V} = 1$. It suffices to find a 1-dimensional Lie submodule $W_1 \subset V$ not contained in W, for then $W \cap W_1 = 0$, implying that $V = W \oplus W_1$.

First, assume that W is a simple Lie module. Let $c = c_\rho$ be the Casimir element of ρ, and define $\ell_c\colon V \to V$ by $v \mapsto cv$. $L' = L$, so $L\bar{V} = 0$ by Remark 21.17; i.e., $LV \subseteq W$. Hence, $e_i e_i^* V \subseteq e_i V \subseteq W$ for each i (notation as in Definition 21.56), implying that $cV \subseteq W$. Thus $\ker \ell_c \neq 0$. By Lemma 21.57, cV is a Lie submodule of W, which must be W or 0. Likewise $cW \leq W$. If $cW = 0$, then $c^2 V = 0$ implying $\operatorname{tr}(c) = 0$, contrary to Lemma 21.57. Hence $cW = W$, implying that $cV = W$, so $\dim \ker \ell_c = 1$ and $W \cap \ker \ell_c = 0$, as desired.

Now we prove the special case by induction on $m = \dim V$. By the previous paragraph we are done unless W has a proper Lie submodule $Y \neq 0$. Then $\bar{V} \cong (V/Y)/(W/Y)$, so by induction, W/Y has a complement W_1/Y in V/Y, where clearly $\dim(W_1/Y) = 1$. But $\dim W_1 < m$, so, again by induction, Y has a 1-dimensional complement Y_1 in W_1. Now

$$W \cap Y_1 \subseteq W \cap W_1 \subseteq Y,$$

implying that $W \cap Y_1 \subseteq Y \cap Y_1 = 0$. Hence Y_1 is a complement to W. We have proved the special case.

General Case. Let $\mathcal{V} = \{f \in \mathrm{Hom}_F(V, W) : f|_W \text{ is scalar multiplication}\}$, and let $\mathcal{W} = \{f \in \mathcal{V} : f|_W = 0\}$. $\mathrm{End}_F V$ is a Lie module over L, by Remark 21.18(ii), and $\mathcal{V}, \mathcal{W}$ are Lie submodules. (Indeed if $f \in V$, then there is $\alpha \in F$ with $f(w) = \alpha w$ for all $w \in W$, and

$$[a, f](w) = af(w) - f(aw) = \alpha aw - \alpha aw = 0.) \tag{21.12}$$

But $\mathcal{W}$ is clearly of codimension 1 in $\mathcal{V}$, so by the Special Case has a 1-dimensional Lie module complement $\mathcal{Y}$ in $\mathcal{V}$. Take $0 \neq f \in \mathcal{Y}$. In particular, $f|_W$ is a nonzero scalar α, so replacing f by $\alpha^{-1}f$ we may assume that $\alpha = 1$; i.e., $f: V \to W$ is a projection of F-vector spaces. Moreover, we claim that $\ker f$ is a Lie submodule of V. Indeed, for any $a \in L$, $[a, f] \in Ff$ since $\mathcal{Y}$ is 1-dimensional; hence for any $v \in \ker f$ we have

$$f(av) = af(v) - [a, f](v) = 0 - 0 = 0.$$

Consequently, $V = W \oplus \ker f$ as L-modules. □

Another result, reminiscent of Wedderburn's Principal Theorem from the associative theory, is Levi's Theorem (Exercise 48), that if $S = L/\operatorname{rad} L$, then L has a Lie subalgebra isomorphic to S, with $L = S \oplus \operatorname{rad} L$ as vector spaces (but not necessarily as Lie algebras). This has some interesting structural consequences given in Exercises 49 and 50, and its proof could serve as an introduction to cohomological methods in Lie algebras, to be described in Chapter 25.

The structure of f.d. semisimple Lie algebras

For most of the remainder of this chapter, we push on toward the Killing-Cartan classification of f.d. semisimple Lie algebras in characteristic 0. The theory is an amazing edifice, built entirely from elementary (albeit intricate) arguments from linear algebra. The idea is to find a suitable nilpotent subalgebra $\mathfrak{h}$ (called a Cartan subalgebra), from which we extract a simultaneous

eigenspace decomposition of L (as $\mathfrak{h}$-module), leading to an inner product whose remarkable properties can be encoded into the celebrated **Dynkin diagrams**.

Root space decomposition with respect to a nilpotent Lie subalgebra.

Let us start by examining simultaneous eigenspaces. Let $n = \dim_F L$. We use the **(generalized) eigenspace decomposition $L = \bigoplus_\alpha L_\alpha$ with respect to** ad_a (acting on L), cf. Remark 2.70 of Volume 1, for which we need F to contain all the generalized eigenvalues of ad_a. For this reason, we continue to assume that F is algebraically closed. (Lacking this assumption on F, one could still make initial progress with the Fitting decomposition of Remark 2.67 of Volume 2, by means of Proposition 16.46.)

Recall from linear algebra that $(\mathrm{ad}_a - \alpha I)^n L_\alpha = 0$. Thus ad_a acts nilpotently on L_0. We call L_0 the **null component** of ad_a, denoted $\mathrm{Null}(a)$. The following easy observation is crucial.

Remark 21.59. $a \in \mathrm{Null}(a)$, since $[aa] = 0$. More generally, if N is a nilpotent Lie algebra containing a, then $N \subseteq \mathrm{Null}(a)$.

Since ad_a is a derivation, we recall an easy result from Corollary 6B.8 of Volume 1: For any generalized eigenvalues α and β,

$$[L_\alpha, L_\beta] \subseteq L_{\alpha+\beta}, \tag{21.13}$$

which is taken formally to be 0 unless $\alpha + \beta$ is an eigenvalue of ad_a. In particular, $\mathrm{Null}(a) = L_0$ is a nonzero Lie subalgebra of L, and each L_α is a Lie module over $\mathrm{Null}(a)$. But all depends on the choice of a. Our objective is to refine the eigenspace decomposition so that the same decomposition works simultaneously (in the sense of Remark 2.71) for each a in a given nilpotent subalgebra.

Definition 21.60. Suppose N is a nilpotent Lie subalgebra of L. A **root space decomposition** is comprised of linear maps $\mathbf{a}: N \to F$, called **roots**, together with a vector space decomposition

$$L = \oplus L_{\mathbf{a}} \tag{21.14}$$

such that, for any $a \in N$, each $L_{\mathbf{a}} \neq 0$ is a generalized eigenspace of $\mathrm{ad}\, a$ with generalized eigenvalue $\mathbf{a}(a)$; in other words $(\mathrm{ad}\, a - \mathbf{a}(a)I)^n L_{\mathbf{a}} = 0$, $\forall a \in N$. $L_{\mathbf{a}}$ is called the **root space** of the root $\mathbf{a}$.

(In the definition one might have $\mathbf{a}(a) = \mathbf{b}(a)$ for some $a \in N$, for roots $\mathbf{a} \neq \mathbf{b}$.) As we shall see soon, the roots play the starring role in the theory. Clearly only a finite number of roots can exist, since $\dim L < \infty$, and we achieve the root space decomposition by iterating the previous argument.

PROPOSITION 21.61 (ROOT SPACE DECOMPOSITION). *Suppose N is a nilpotent Lie subalgebra of a f.d. Lie algebra L. Then there exists a root space decomposition $L = \bigoplus_{\mathbf{a}} L_{\mathbf{a}}$, which is unique with respect to the choice of N. Furthermore, if $\mathbf{a}, \mathbf{b}$, and $\mathbf{a}+\mathbf{b}$ are roots, then $[L_{\mathbf{a}} L_{\mathbf{b}}] \subseteq L_{\mathbf{a}+\mathbf{b}}$; in particular, each $L_{\mathbf{a}}$ is a Lie module over $L_{\mathbf{0}}$.*

Proof. Take the generalized eigenspace decomposition $L = \oplus L_\alpha$ with respect to some $a_1 \in N$. Now $N \subseteq \text{Null}(a_1)$ by Remark 21.59, so taking any $a_2 \in N$, $\text{ad}_{a_2} L_\alpha \subseteq L_\alpha$ for each α. Hence the L_α decompose further into eigenspaces with respect to ad_{a_2}, thereby yielding a simultaneous eigenspace decomposition for ad_{a_1} and ad_{a_2}. Continuing for $a_3, \dots, a_u$ in N, this process must eventually stabilize since the number of subspaces cannot be greater than $\dim L$. When the process stabilizes, we have the desired decomposition; we would get the same result starting with any eigenspace decomposition of N, thereby proving uniqueness. We check that $[L_{\mathbf{a}} L_{\mathbf{b}}] \subseteq L_{\mathbf{a}+\mathbf{b}}$ at each element $a_i \in N$. □

One can interpret Proposition 21.61 as saying that the Lie algebra L is **graded** by the roots of N. This key facet lies at the heart of the theory, and often is taken as an axiom when generalizing to infinite-dimensional Lie algebras.

COROLLARY 21.62.

(i) *For each $L_{\mathbf{a}}$, one can choose a suitable base such that, for each $a \in N$, ad_a acts as an upper triangular matrix on $L_{\mathbf{a}}$ whose diagonal part is the scalar matrix $\mathbf{a}(a)I$. In particular, if $\mathbf{a}(a) \neq 0$, then the restriction of ad_a to $L_{\mathbf{a}}$ is nonsingular, and thus is onto.*

(ii) $\mathbf{a}(a) = 0$ *for all roots* $\mathbf{a}$ *and all* $a \in N'$.

Proof. (i) The assertion follows at once from Remark 21.35, viewing N as a solvable Lie algebra acting via the adjoint action on L.

(ii) ad_a acts nilpotently on N', by Corollary 21.33. □

Since the generalized eigenvalue for each $a \in N$ on $L_{\mathbf{0}}$ is 0, $L_{\mathbf{0}}$ is called the **null component** of the nilpotent subalgebra N, and is denoted $\text{Null}(N)$; note that $N \subseteq \text{Null}(N)$. The root space decomposition is to be refined until we reach Summary 21.80. First we bring in the Killing form $\langle\ ,\ \rangle$.

Remark 21.63. Let $n_{\mathbf{a}} = \dim L_{\mathbf{a}}$.

(i) For $a \in N$, the trace of ad_a restricted to L_a is $n_{\mathbf{a}}\mathbf{a}(a)$, by Corollary 21.62(i).

(ii) $\langle a, b\rangle = \sum_{\mathbf{a}} n_{\mathbf{a}}\mathbf{a}(a)\mathbf{a}(b)$ for all $a, b \in N$. This is clear since, according to Corollary 21.62, the entries along the diagonal of $\text{ad}_a \text{ad}_b$ are the products

of the respective diagonal entries of ad_a and ad_b, namely $\mathbf{a}(a)\mathbf{a}(b)$, taken $n_{\mathbf{a}}$ times for each root $\mathbf{a}$.

PROPOSITION 21.64. *$L_{\mathbf{b}} \perp L_{\mathbf{a}}$ (with respect to the Killing form) for any roots $\mathbf{a} \neq -\mathbf{b}$.*

Proof. We aim to show that $\langle b, a\rangle = 0$ for all b in $L_{\mathbf{b}}$ and a in $L_{\mathbf{a}}$.

Special Case. First assume that $\mathbf{b} = \mathbf{0}$. Then $\mathbf{a} \neq \mathbf{0}$, so there is $c \in N$ such that $\mathbf{a}(c) \neq 0$, and thus $\text{ad}_c : L_{\mathbf{a}} \to L_{\mathbf{a}}$ is onto. Hence, for any $a \in L_{\mathbf{a}}$, we have $a = (\text{ad}_c)^n \tilde{a}$ for suitable $\tilde{a} \in L_{\mathbf{a}}$, so, in view of (21.11),

$$\langle b, a\rangle = \langle b, (\text{ad}_c)^n \tilde{a}\rangle = \pm\langle (\text{ad}_c)^n b, \tilde{a}\rangle = \langle 0, \tilde{a}\rangle = 0.$$

General Case. In view of the special case, we may assume that neither $\mathbf{a}$ nor $\mathbf{b}$ are 0. Take $c \in N$ such that $\mathbf{a}(c) \neq 0$. For any $a \in L_{\mathbf{a}}$, we have $a = \text{ad}_c \tilde{a}$ for suitable $\tilde{a} \in L_{\mathbf{a}}$, so this time

$$\langle b, a\rangle = \langle b, [\tilde{a}c]\rangle = \langle [b\tilde{a}], c\rangle = 0$$

by the special case (noting that $c \in L_{\mathbf{0}}$, $[b\tilde{a}] \in L_{\mathbf{a}+\mathbf{b}}$, and $\mathbf{a}+\mathbf{b} \neq 0$). □

This result reduces much of the study of the Killing form to $\text{Null}(N)$.

COROLLARY 21.65. *If L is f.d. semisimple with nilpotent Lie subalgebra N, then the restriction of the Killing form to the null component of N is nondegenerate.*

Proof. The Killing form is nondegenerate on L, by Cartan's criterion. But $L_{\mathbf{0}} = \text{Null}(N)$ is orthogonal to all $L_{\mathbf{a}}$, $\mathbf{a} \neq \mathbf{0}$, and thus $\text{rad}\, L_{\mathbf{0}}$ is orthogonal to $L_{\mathbf{0}} \oplus (\bigoplus_{\mathbf{a}\neq\mathbf{0}} L_{\mathbf{a}}) = L$, implying that $\text{rad}\, L_{\mathbf{0}} = 0$. □

Cartan subalgebras

Corollary 21.65 leads us to study the structure of the Killing form restricted to the null component $\text{Null}(N)$ of a nilpotent Lie subalgebra N. On the other hand, Remark 21.63 shows us how to compute the Killing form on the restriction to N. Thus, the most amenable situation would be achieved when $\text{Null}(N) = N$. Let us see how to obtain such N. Recall from Remark 21.11 the definition of the normalizer $N_L(A)$ of a Lie subalgebra A.

Definition 21.66. A **Cartan subalgebra** of a Lie algebra L is a nilpotent Lie subalgebra $\mathfrak{h}$ that is its own normalizer in L; i.e., if $[a, \mathfrak{h}] \subseteq \mathfrak{h}$, then $a \in \mathfrak{h}$.

For example, the set of diagonal matrices is a Cartan subalgebra of $gl(n, F)$. On the other hand, the set of strictly upper triangular matrices is a nilpotent subalgebra that is not contained in any Cartan subalgebra.

PROPOSITION 21.67. *A nilpotent Lie subalgebra $\mathfrak{h}$ of L is Cartan iff $\mathfrak{h} = \mathrm{Null}(\mathfrak{h})$.*

Proof. Let $N = N_L(\mathfrak{h}) \supseteq \mathfrak{h}$, and $N_0 = \mathrm{Null}(\mathfrak{h}) \supseteq N$ (since $[N\mathfrak{h}] \subseteq \mathfrak{h}$).

($\Leftarrow$) $\mathfrak{h} = N_0 \supseteq N \supseteq \mathfrak{h}$, so equality holds.

($\Rightarrow$) We are given $\mathfrak{h} = N \subseteq N_0$; we are done unless $N \subset N_0$. Clearly $\mathfrak{h}$ acts nilpotently (via the adjoint) on N_0 and thus on $V = N_0/N$, so Lemma 21.28 implies that $\mathfrak{h}v = 0$ for some nonzero v in V. Writing $v = a+N$ for $a \in N_0 \setminus N$, we have $[\mathfrak{h}a] \subseteq N = \mathfrak{h}$, implying that $a \in N$, a contradiction. □

We confront these notions with the generalized eigenspace decomposition with respect to ad_a for an element $a \in L$.

COROLLARY 21.68. *If $\mathfrak{h} = \mathrm{Null}(a)$ is nilpotent, then $\mathfrak{h}$ is a Cartan subalgebra.*

Proof. Recall by Remark 21.59 that $a \in \mathrm{Null}(a) = \mathfrak{h}$, so $\mathrm{Null}(a) \supseteq \mathrm{Null}\,\mathfrak{h} \supseteq \mathfrak{h} = \mathrm{Null}(a)$ by hypothesis; thus, $\mathfrak{h} = \mathrm{Null}(\mathfrak{h})$, implying that $\mathfrak{h}$ is Cartan. □

Remark 21.69. The dimension of $\mathrm{Null}(a)$ equals the multiplicity of 0 in the characteristic polynomial of ad_a, which is at least 1 since $a \in \mathrm{Null}(a)$.

This leads us to a way of constructing Cartan subalgebras. We say that $a \in L$ is **regular** if the dimension of $\mathrm{Null}(a)$ is minimal possible.

THEOREM 21.70. *If a is regular, then $\mathrm{Null}(a)$ is a Cartan subalgebra of L.*

Proof. By Corollary 21.68, it suffices to show that $\mathfrak{h} = \mathrm{Null}(a)$ is nilpotent. By Theorem 21.27 we need to prove that ad_b acts nilpotently on $\mathfrak{h}$ for all $b \in \mathfrak{h}$. Taking a generalized eigenspace decomposition $L = \oplus L_\alpha$ with respect to ad_a and putting $\tilde{L} = \bigoplus_{\mathbf{a}\neq 0} L_{\mathbf{a}}$, we have $L = \mathfrak{h} \oplus \tilde{L}$ as Lie $\mathfrak{h}$-modules. With respect to this decomposition, $A = \mathrm{ad}_a$ partitions as $\left(\begin{smallmatrix} A_1 & 0 \\ 0 & A_2 \end{smallmatrix}\right)$, where A_1 is nilpotent since $\mathfrak{h} = \mathrm{Null}(a)$. The characteristic polynomial of A_1 is λ^m, where $m = \dim \mathfrak{h}$, but λ does not divide the characteristic polynomial of A_2.

Since $\mathfrak{h}$ and $\tilde{A}$ are invariant under $\mathrm{ad}\, b$, the matrix B for ad_b also partitions as $\left(\begin{smallmatrix} B_1 & 0 \\ 0 & B_2 \end{smallmatrix}\right)$, and we need to prove that B_1 is nilpotent. Assume the contrary; then λ^m does not divide the characteristic polynomial of B_1.

We consider elements $a + \nu b \in L$, where $\nu \in F$. Then

$$\mathrm{ad}_{a+\nu b} = \mathrm{ad}_a + \nu\, \mathrm{ad}_b = A + \nu B.$$

Matching the partitions of A and B we see that the characteristic polynomial for $A + \nu B$ has the form $f_1 f_2$, where f_i is the characteristic polynomial of $A_i + \nu B_i$ for $i = 1, 2$. For almost all ν, $A_2 + \nu B_2$ is nonsingular and f_1 is not divisible by λ^m. Hence $\lambda \nmid f_2$ and $\lambda^m \nmid f_1$, implying that $\lambda^m \nmid f_1 f_2$, so, for these ν,

$$\dim_F \mathrm{Null}(a + \nu b) < \dim_F \mathrm{Null}(a),$$

contrary to the regularity of a. Thus B_1 is nilpotent after all. □

Summarizing our previous results, we have the following basic information.

THEOREM 21.71. *Any f.d. semisimple Lie algebra over an algebraically closed field of characteristic 0 has a Cartan subalgebra $\mathfrak{h}$, which is its own nullspace under the corresponding root space decomposition. The restriction of the Killing form to $\mathfrak{h}$ is nondegenerate. In particular, if $h \in \mathfrak{h}$ with $\mathbf{a}(h) = 0$ for all roots $\mathbf{a}$ of $\mathfrak{h}$, then $h = 0$.*

Proof. By Theorem 21.70, we may take $\mathfrak{h} = \mathrm{Null}(a)$ for $a \in L$ regular. Then $\mathfrak{h} = \mathrm{Null}\,\mathfrak{h}$ by Proposition 21.67, so the restriction of the Killing form to $\mathfrak{h}$ is nondegenerate by Corollary 21.65. The last assertion then follows from Remark 21.63, which shows that $\langle h, \mathfrak{h}\rangle = 0$. □

This already has striking consequences.

COROLLARY 21.72. *Suppose $\mathfrak{h}$ is a Cartan subalgebra of the semisimple Lie algebra L of characteristic 0.*

(i) *$\mathfrak{h}$ is Abelian;*

(ii) *ad_h is semisimple for all $h \in \mathfrak{h}$, and consequently, for any root $\mathbf{a}$ of $\mathfrak{h}$,*

$$\mathrm{ad}_h(a) = [ha] = \mathbf{a}(h)a \tag{21.15}$$

for each $a \in L_{\mathbf{a}}$ and $h \in \mathfrak{h}$.

Proof. (i) Suppose $a \in \mathfrak{h}'$. Corollary 21.62(ii) implies that $\mathbf{a}(a) = 0$ for all roots $\mathbf{a}$, yielding $a = 0$.

(ii) Take the Jordan decomposition $\mathrm{ad}_h = \mathbf{s} + \mathbf{n}$ for $\mathbf{s}$ semisimple and $\mathbf{n}$ nilpotent. We aim to show that $\mathrm{ad}_h = \mathbf{s}$. Writing ad_h in upper triangular form with respect to a suitable base of the $L_{\mathbf{a}}$ (cf. Corollary 21.62), we see that $\mathbf{s}$ acts likes the scalar matrix $\mathbf{a}(h)I$ on $L_{\mathbf{a}}$, i.e.,

$$\mathbf{s}a = \mathbf{a}(h)a \tag{21.16}$$

for all $a \in L_{\mathbf{a}}$. For any $a \in L_{\mathbf{a}}$, $b \in L_{\mathbf{b}}$, we get $[ab] \in L_{\mathbf{b}+\mathbf{a}}$, and thus

$$\begin{aligned}\mathbf{s}[ab] &= (\mathbf{b}+\mathbf{a})(h)[ab] \\ &= (\mathbf{b}(h)+\mathbf{a}(h))[ab] = [a, \mathbf{b}(h)b] + [\mathbf{a}(h)a, b] = [a, \mathbf{s}b] + [\mathbf{s}a, b],\end{aligned}$$

so $\mathbf{s}$ acts as a derivation on L. By Theorem 21.54, $\mathbf{s}$ is inner, i.e., of the form ad_w, for suitable $w \in L$. But $[w\mathfrak{h}] = \mathrm{ad}_w(\mathfrak{h}) = \mathbf{s}\mathfrak{h} = 0$ by (i), Proposition 21.67, and Equation (21.16), implying that $w \in N_L(\mathfrak{h}) = \mathfrak{h}$. Hence $h - w \in \mathfrak{h}$, and furthermore $\mathrm{ad}_{h-w} = \mathrm{ad}_h - \mathrm{ad}_w$ has 0 as its only generalized eigenvalue, i.e., $\mathbf{a}(h-w) = 0$ for all roots $\mathbf{a}$. Theorem 21.71 implies that $h - w = 0$, i.e., $h = w$, so $\mathrm{ad}_h = \mathrm{ad}_w = \mathbf{s}$, as desired. □

Suddenly the theory has become much better focused: Lie subalgebras satisfying the conclusion of Corollary 21.72(ii) are called **toral**. In this case, Equation (21.15) shows that the generalized eigenvalues (and eigenvectors) for elements of $\mathrm{ad}\,\mathfrak{h}$ are ordinary eigenvalues (and eigenvectors).

Preliminary properties of roots of Cartan subalgebras.

From now on, until our next level of abstraction in Definition 21.92, we work with a given Cartan subalgebra $\mathfrak{h}$ of a semisimple Lie algebra; by "root" we mean a root of $\mathfrak{h}$. Our objective for the next eight pages is to study the roots via the geometry of the dual space $\mathfrak{h}^*$ with respect to the Killing form. We start with some easy observations.

Remark 21.73. The roots (for $\mathfrak{h}$) span $\mathfrak{h}^*$. (Otherwise there is some nonzero $h \in \mathfrak{h}$ such that $\mathbf{a}(h) = 0$ for all roots $\mathbf{a}$, contrary to Theorem 21.71.)

Remark 21.74. (i) If $\mathbf{a}$ is a root, then $-\mathbf{a}$ is a root and $[L_{\mathbf{a}}L_{-\mathbf{a}}] \neq 0$. (Otherwise $L_{\mathbf{a}}$ would be perpendicular to every $L_{\mathbf{b}}$, by Proposition 21.64, implying that $L_{\mathbf{a}} = 0$.)

(ii) Suppose $e \in L_{\mathbf{a}}$. If $e' \in L_{-\mathbf{a}}$, $\gamma = \langle e', e\rangle$, and $h \in \mathfrak{h}$, then

$$\langle [e'e], h\rangle = \langle e', [eh]\rangle = -\langle e', [he]\rangle = -\mathbf{a}(h)\langle e', e\rangle = -\gamma\mathbf{a}(h).$$

(iii) In (ii), choose $e' \in L_{-\mathbf{a}}$ such that $\gamma = 1$. Then defining $r_{\mathbf{a}} = [ee'] \in \mathfrak{h}$, we have $\langle r_{\mathbf{a}}, h\rangle = \mathbf{a}(h)$ for all $h \in \mathfrak{h}$.

Continuing with this geometric flavor, we introduce an important notion.

Definition 21.75. Given two roots $\mathbf{a}, \mathbf{b}$, define the **a-string** $\mathcal{S}$ of $\mathbf{b}$ to be the set of all roots of the form $\mathbf{b} + k\mathbf{a} : k \in \mathbb{Z}$. Put $L_{\mathcal{S}} = \bigoplus_{k\in\mathbb{Z}} L_{\mathbf{b}+k\mathbf{a}}$.

Any string is finite, since we have only finitely many roots. On the other hand, by inspection, $L_{\mathcal{S}}$ is closed under the adjoint actions of $L_{\mathbf{a}}$ and of $L_{-\mathbf{a}}$, since $[L_{\pm\mathbf{a}}, L_{\mathbf{b}+k\mathbf{a}}] = L_{\mathbf{b}+(k\pm1)\mathbf{a}}$.

Recall $n_{\mathbf{a}} = \dim L_{\mathbf{a}}$. For any $h \in \mathfrak{h}$, let t_h denote the trace of ad_h restricted to $L_{\mathcal{S}}$. By Remark 21.63(i),

$$\begin{aligned}t_h &= \sum_k n_{\mathbf{b}+k\mathbf{a}}(\mathbf{b}(h) + k\mathbf{a}(h)) \\ &= \sum_k n_{\mathbf{b}+k\mathbf{a}}\mathbf{b}(h) + \sum_k n_{\mathbf{b}+k\mathbf{a}}k\mathbf{a}(h) \\ &= \dim L_{\mathcal{S}}\,\mathbf{b}(h) + m\mathbf{a}(h),\end{aligned} \tag{21.17}$$

where $m = \sum_k n_{\mathbf{b}+k\mathbf{a}}k \in \mathbb{Z}$.

LEMMA 21.76. *If $h \in [L_{\mathbf{a}}L_{-\mathbf{a}}]$, then $\mathbf{b}(h) \in \mathbb{Q}\mathbf{a}(h)$ for any root $\mathbf{b}$; in fact $\mathbf{b}(h) = -\frac{m}{\dim L_{\mathcal{S}}}\mathbf{a}(h)$, for m as in Equation (21.17)*

Proof. By Equation (21.3), $\mathrm{ad}_h \in [\mathrm{ad}\,L_{\mathbf{a}}, \mathrm{ad}\,L_{-\mathbf{a}}]$ and thus has trace 0 on $L_{\mathcal{S}}$. Hence, Equation (21.17) yields $\mathbf{b}(h) = -\frac{m}{\dim L_{\mathcal{S}}}\mathbf{a}(h)$. □

PROPOSITION 21.77. *$[L_{\mathbf{a}}, L_{-\mathbf{a}}] = Fr_{\mathbf{a}}$ with $r_{\mathbf{a}}$ as in Remark 21.74(iii) and $\mathbf{a}(r_{\mathbf{a}}) \neq 0$.*

Proof. To prove that $\dim[L_{\mathbf{a}}L_{-\mathbf{a}}] = 1$, we show that the map $[L_{\mathbf{a}}L_{-\mathbf{a}}] \to F$ given by $h \mapsto \mathbf{a}(h)$ is 1:1. Assume that $h \in [L_{\mathbf{a}}L_{-\mathbf{a}}]$ with $\mathbf{a}(h) = 0$. The lemma shows that $\mathbf{b}(h) = 0$ for any root $\mathbf{b}$; we conclude by Theorem 21.71 that $h = 0$, as desired.

Since $r_{\mathbf{a}} \in [L_{\mathbf{a}}L_{-\mathbf{a}}]$ we see that $[L_{\mathbf{a}}L_{-\mathbf{a}}] = Fr_{\mathbf{a}}$ and thus $r_{\mathbf{a}} \neq 0$. Hence, the previous paragraph shows that $\mathbf{a}(r_{\mathbf{a}}) \neq 0$. □

Next we translate the Killing form to the dual space $\mathfrak{h}^*$. Given $f : \mathfrak{h} \to F$, we take the unique $r_f \in \mathfrak{h}$ such that

$$\langle r_f, h\rangle = f(h), \quad \forall h \in \mathfrak{h}$$

(since the Killing form is nondegenerate on $\mathfrak{h}$), and now define the bilinear form

$$\langle f, g\rangle = \langle r_f, r_g\rangle. \tag{21.18}$$

Remark 21.78. (i) Our definition of r_f is consistent with Remark 21.74(iii) when f is a root $\mathbf{a}$. In particular, $\langle \mathbf{a}, \mathbf{b}\rangle$ means $\langle r_{\mathbf{a}}, r_{\mathbf{b}}\rangle$ throughout the sequel.

(ii) Each root $\mathbf{a}$ is nonisotropic, since $\langle \mathbf{a}, \mathbf{a}\rangle = \langle r_{\mathbf{a}}, r_{\mathbf{a}}\rangle = \mathbf{a}(r_{\mathbf{a}}) \neq 0$ by Proposition 21.77.

(iii) By definition, $\mathbf{b}(r_{\mathbf{a}}) = \langle r_{\mathbf{b}}, r_{\mathbf{a}}\rangle = \langle \mathbf{b}, \mathbf{a}\rangle = \langle \mathbf{a}, \mathbf{b}\rangle$ for all roots $\mathbf{a}$ and $\mathbf{b}$.

Let us push this idea further.

THEOREM 21.79. *For any root* $\mathbf{a}$, $\dim L_{\mathbf{a}} = \dim L_{-\mathbf{a}} = 1$, *and* $k\mathbf{a}$ *is not a root whenever* $1 < |k| \in \mathbb{N}$.

Proof. Take $a \in L_{\mathbf{a}}$ and $b \in L_{-\mathbf{a}}$ with $[ab] = r_{\mathbf{a}} \in L_0 = \mathfrak{h}$. Define

$$V = \mathfrak{h} + Fb + \sum_{k>0} L_{k\mathbf{a}},$$

clearly invariant under ad_a and ad_b. Thus in view of Equations (21.3) and (21.17), we have

$$0 = \mathrm{tr}_V(\mathrm{ad}\, r_{\mathbf{a}}) = 0 - \mathbf{a}(r_{\mathbf{a}}) + \sum_{k>0} n_{k\mathbf{a}}\, k\mathbf{a}(r_{\mathbf{a}}),$$

implying that $\sum n_{k\mathbf{a}} k = 1$. Since $n_{\mathbf{a}} \geq 1$, we conclude that $n_{\mathbf{a}} = 1$ and $n_{k\mathbf{a}} = 0$ for all $k > 1$. Likewise $n_{-\mathbf{a}} = 1$, and $n_{-k\mathbf{a}} = 0$ for all $k > 1$. □

Summary 21.80. Let us recapitulate our setup in terms of the Cartan subalgebra $\mathfrak{h}$. The nonzero roots pair off as $\pm\mathbf{a}$, and our root space decomposition has become

$$L = \mathfrak{h} \oplus \bigoplus_{\mathbf{a}\neq 0} (L_{\mathbf{a}} \oplus L_{-\mathbf{a}}), \tag{21.19}$$

where $\mathfrak{h}$ is Abelian, comprised of semisimple elements whose actions on the $L_{\mathbf{a}}$ are given by (21.15); furthermore, for each root $\mathbf{a} \neq 0$, $\dim L_{\mathbf{a}} = 1$, and none of the various $\mathbf{a}$ are integral multiples of another. (This is to be improved in Corollary 21.86.) Remark 21.63 has simplified to:

$$\langle h_1, h_2\rangle = \sum_{\mathbf{a}\neq 0} \mathbf{a}(h_1)\mathbf{a}(h_2), \quad \forall h_1, h_2 \in \mathfrak{h}. \tag{21.20}$$

Lie structure in terms of $sl(2,F)$

Using Lie module theory, we now can reduce much of the structure of our semisimple Lie algebra L to the manageable special case of $sl(2,F)$.

Remark 21.81. For each root $\mathbf{a}$, we define $\hat{L}_{\mathbf{a}} = L_{\mathbf{a}} + L_{-\mathbf{a}} + Fr_{\mathbf{a}}$. Then $\hat{L}_{\mathbf{a}}$ is a Lie subalgebra of L. Explicitly, taking e, e' as in Remark 21.74(iii) so that $r_{\mathbf{a}} = [ee']$, we define

$$e_{\mathbf{a}} = e \in L_{\mathbf{a}}, \qquad f_{\mathbf{a}} = \frac{2}{\langle \mathbf{a}, \mathbf{a}\rangle} e' \in L_{-\mathbf{a}}, \qquad h_{\mathbf{a}} = [e_{\mathbf{a}} f_{\mathbf{a}}] = \frac{2}{\langle \mathbf{a}, \mathbf{a}\rangle} r_{\mathbf{a}} \in \mathfrak{h}.$$

Then, in view of Corollary 21.72(ii)

$$[e_{\mathbf{a}} h_{\mathbf{a}}] = -\frac{2}{\langle \mathbf{a}, \mathbf{a}\rangle}[r_{\mathbf{a}} e_{\mathbf{a}}] = -\frac{2\mathbf{a}(r_{\mathbf{a}})}{\langle \mathbf{a}, \mathbf{a}\rangle} e_{\mathbf{a}} = -2e_{\mathbf{a}},$$

and likewise Remark 21.74(ii) shows that

$$[f_{\mathbf{a}} h_{\mathbf{a}}] = -\frac{-4\mathbf{a}(r_{\mathbf{a}})}{\langle \mathbf{a}, \mathbf{a}\rangle^2} e' = 2f_{\mathbf{a}}.$$

Thus, each $\hat{L}_{\mathbf{a}}$ is isomorphic to $sl(2,F)$.

Remark 21.82. Suppose V is any $\hat{L}_{\mathbf{a}}$-module. If $v \in V$ is an eigenvector of $h_{\mathbf{a}}$, then $e_{\mathbf{a}} v$ and $f_{\mathbf{a}} v$ also are eigenvectors of $h_{\mathbf{a}}$. More precisely, writing $h_{\mathbf{a}} v = \gamma v$, we have

$$h_{\mathbf{a}}(e_{\mathbf{a}} v) = [h_{\mathbf{a}} e_{\mathbf{a}}]v + e_{\mathbf{a}}(h_{\mathbf{a}} v) = 2e_{\mathbf{a}} v + \gamma e_{\mathbf{a}} v = (2+\gamma) e_{\mathbf{a}} v;$$
$$h_{\mathbf{a}}(f_{\mathbf{a}} v) = [h_{\mathbf{a}} f_{\mathbf{a}}]v + f_{\mathbf{a}}(h_{\mathbf{a}} v) = -2f_{\mathbf{a}} v + \gamma f_{\mathbf{a}} v = (-2+\gamma) f_{\mathbf{a}} v.$$

Iterating Remark 21.82, one sees that $e_{\mathbf{a}}^k v$, $k = 1, 2, \ldots$ are eigenvectors (of $h_{\mathbf{a}}$) having distinct eigenvalues, and thus are linearly independent. If $\dim_F V < \infty$, this means there is some k maximal for which $w = e_{\mathbf{a}}^k v \neq 0$; we call such w a **maximal** eigenvector of $h_{\mathbf{a}}$. By definition, $e_{\mathbf{a}} w = 0$ for any maximal eigenvector w.

LEMMA 21.83. *Suppose* $v_0 \in V$ *is a maximal eigenvector of* $h_{\mathbf{a}}$ *having eigenvalue* γ, *and inductively let* $v_k = \frac{1}{k} f_{\mathbf{a}} v_{k-1}$ *for each* $k > 0$. *Then*

(i) $f_{\mathbf{a}} v_k = (k+1) v_{k+1}$;
(ii) $h_{\mathbf{a}} v_k = (\gamma - 2k) v_k$;
(iii) $e_{\mathbf{a}} v_k = (\gamma - k + 1) v_{k-1}$.

Proof. (i) By definition of v_{k+1}.

(ii) $h_{\mathbf{a}} v_{k-1} = (\gamma - 2(k-1))\, v_{k-1}$ by induction, so Remark 21.82 (taking $v = v_{k-1}$) yields

$$h_{\mathbf{a}} v_k = \frac{1}{k} h_{\mathbf{a}} f_{\mathbf{a}} v_{k-1} = \frac{1}{k} \left(-2 + \gamma - 2(k-1)\right) f_{\mathbf{a}} v_{k-1} = (\gamma - 2k) v_k.$$

(iii) $e_{\mathbf{a}} v_{k-1} = (\gamma - k + 2) v_{k-2}$ by induction. Hence

$$\begin{aligned} k e_{\mathbf{a}} v_k &= e_{\mathbf{a}}(f_{\mathbf{a}} v_{k-1}) \\ &= f_{\mathbf{a}}(e_{\mathbf{a}} v_{k-1}) + [e_{\mathbf{a}} f_{\mathbf{a}}] v_{k-1} \\ &= f_{\mathbf{a}}(\gamma - k + 2) v_{k-2} + h_{\mathbf{a}} v_{k-1} \\ &= (\gamma - k + 2)(k-1) v_{k-1} + (\gamma - 2(k-1)) v_{k-1} \quad \text{(by (i),(ii))} \\ &= k(\gamma - k + 1) v_{k-1}, \end{aligned}$$

and we cancel k. □

Remark 21.83′. In the notation of Lemma 21.83, take m maximal such that $v_m \neq 0$. Then $v_{m+1} = 0$, so (iii) yields $0 = e_{\mathbf{a}}v_{m+1} = (\gamma - m)v_m$, implying $\gamma = m$.

PROPOSITION 21.84. *Any simple $\hat{L}_{\mathbf{a}}$-module V has an eigenspace decomposition*

$$V = V_m \oplus V_{m-2} \oplus \cdots \oplus V_{-(m-2)} \oplus V_{-m}, \tag{21.21}$$

where each component $V_{m-2j} = Fv_j$ is a one-dimensional eigenspace of $h_{\mathbf{a}}$ with eigenvalue $m-2j$. In particular, V is determined up to isomorphism by its dimension (which is $m+1$), and all eigenvalues are integers.

Proof. We start with a maximal eigenvector v_0 of $h_{\mathbf{a}}$ having eigenvalue γ. (Thus $e_{\mathbf{a}}v_0 = 0$.) By Remark 21.83′,

$$V_m \oplus V_{m-2} \oplus \cdots \oplus V_{-(m-2)} \oplus V_{-m}$$

is clearly a Lie submodule of V, and thus equals V since V is simple. This describes the action completely, thereby yielding the last assertion. □

In short, we start with any $h_{\mathbf{a}}$-eigenvector of a simple $\hat{L}_{\mathbf{a}}$-module V, apply $e_{\mathbf{a}}$ as many times as needed to get a maximal eigenvector, and then apply $f_{\mathbf{a}}$ as many times as needed to generate V. One can reverse these steps. By inspection, Equation (21.21) defines a simple $\hat{L}_{\mathbf{a}}$-module under the action given in Lemma 21.83; hence (up to isomorphism), there is a unique simple $\hat{L}_{\mathbf{a}}$-module of dimension m for each natural number m.

COROLLARY 21.85. *For any f.d. Lie module M over $\hat{L}_{\mathbf{a}}$, the eigenvalues with respect to $h_{\mathbf{a}}$ are all integers. Furthermore, the number of summands in a direct sum decomposition of M into simple Lie submodules (cf. Weyl's Theorem 21.58) is* $\dim M_0 + \dim M_1$, *where M_j denotes the $h_{\mathbf{a}}$-eigenspace with eigenvalue j.*

Proof. We write M as a direct sum of simple Lie submodules, using Weyl's Theorem, and apply Proposition 21.84 to each simple summand V. The second assertion follows when we note that every such V has a 0-component or a 1-component (each of dimension 1), but not both. □

More about strings of roots.

We obtain deeper information about roots of our given Cartan algebra $\mathfrak{h}$ by applying our setup to $L_{\mathcal{S}}$, defined in Definition 21.75, where $\mathcal{S}$ is an $\mathbf{a}$-string of a root $\mathbf{b}$. $L_{\mathcal{S}}$ is clearly a Lie module over $\hat{L}_{\mathbf{a}} = L_{\mathbf{a}} + L_{\mathbf{a}} + Fr_{\mathbf{a}}$ via the adjoint action (cf. Example 21.15) and thus, as above, has a maximal eigenvector with respect to $h_{\mathbf{a}}$.

COROLLARY 21.86. *Suppose $\mathbf{a} \neq \pm\mathbf{b}$ are any nonzero roots of L, and $\mathcal{S}$ is an $\mathbf{a}$-string of $\mathbf{b}$. Then*

(i) *$L_{\mathcal{S}}$ is a simple Lie module over $L_{\mathbf{a}}$ and thus has the form (21.21).*
(ii) *The eigenvalues in $L_{\mathcal{S}}$ comprise an arithmetic sequence of integers with even difference, the number of terms being the dimension of $L_{\mathcal{S}}$.*
(iii) $\langle h_{\mathbf{b}}, h_{\mathbf{a}} \rangle = \mathbf{b}(h_{\mathbf{a}}) = \frac{2\langle \mathbf{a},\mathbf{b}\rangle}{\langle \mathbf{a},\mathbf{a}\rangle}$ *is an integer.*
(iv) *If $\mathbf{b}$ is a scalar multiple of $\mathbf{a}$, then $\mathbf{b} = \pm\mathbf{a}$.*

Proof. (i) Writing $L_{\mathbf{a}} = Fe_{\mathbf{a}}$ and $L_{-\mathbf{a}} = Ff_{\mathbf{a}}$, we see that any element of $\mathcal{S}$ is obtained by applying $\mathrm{ad}_{e_{\mathbf{a}}}$ and $\mathrm{ad}_{f_{\mathbf{a}}}$ a suitable number of times to $L_{\mathbf{b}}$. Thus, Remark 21.82 shows that the difference of any two $h_{\mathbf{a}}$-eigenvalues is an even integer. It follows that 0 and 1 cannot both appear as eigenvalues for $L_{\mathcal{S}}$, so, using Corollary 21.85 and the fact that the components $L_{\mathcal{S}}$ have dimension 1, we conclude that $L_{\mathcal{S}}$ has the decomposition (21.21); therefore $L_{\mathcal{S}}$ is simple as a Lie module over $\hat{L}_{\mathbf{a}}$.

(ii) The eigenvalues are $\gamma - 2u, \ldots, \gamma - 2, \gamma, \gamma + 2, \ldots, \gamma + 2t$, where $\gamma = \mathbf{b}(h_{\mathbf{a}})$ and $t, u \in \mathbb{N}$, and (21.21) implies that $\gamma - 2u = -(\gamma + 2t)$, yielding

$$\gamma = u - t \in \mathbb{Z}. \tag{21.22}$$

(iii) We saw in (ii) that $\mathbf{b}(h_{\mathbf{a}}) \in \mathbb{Z}$. Also,

$$\mathbf{b}(h_{\mathbf{a}}) = \mathbf{b}\left(\frac{2}{\langle \mathbf{a},\mathbf{a}\rangle} r_{\mathbf{a}}\right) = \frac{2}{\langle \mathbf{a},\mathbf{a}\rangle}\mathbf{b}(r_{\mathbf{a}}) = \frac{2\langle \mathbf{b},\mathbf{a}\rangle}{\langle \mathbf{a},\mathbf{a}\rangle}$$

by Remark 21.78(iii).

(iv) If $\mathbf{b} = \alpha\mathbf{a}$, then $2\alpha = 2\frac{\langle \mathbf{b},\mathbf{a}\rangle}{\langle \mathbf{a},\mathbf{a}\rangle} \in \mathbb{Z}$, implying that $\alpha \in \frac{1}{2}\mathbb{Z}$. But $\mathbf{a} = \alpha^{-1}\mathbf{b}$, so likewise $\alpha^{-1} \in \frac{1}{2}\mathbb{Z}$. This means $\alpha \in \{\pm\frac{1}{2}, \pm 1, \pm 2\}$. But $\alpha \neq \pm 2$ by Theorem 21.79, and likewise $\alpha^{-1} \neq \pm 2$, so we are left with $\alpha = \pm 1$.

□

Remark 21.87. (i) The sequence of nonzero eigenspaces in any string must be "unbroken" since any break would designate the end of a nonzero submodule of the module $L_{\mathcal{S}}$ which, however, is simple.

More explicitly, let $\mathcal{S}$ be an $\mathbf{a}$-string of $\mathbf{b}$ for nonzero roots $\mathbf{a} \neq \pm\mathbf{b}$. Define the **length** m of $\mathcal{S}$ as $\dim L_{\mathcal{S}} - 1$; then $\mathcal{S}$ must have the form

$$\{\mathbf{b} - q\mathbf{a},\ \mathbf{b} - (q-1)\mathbf{a},\ \ldots,\ \mathbf{b},\ \ldots,\ \mathbf{b} + (m-q)\mathbf{a}\}.$$

(ii) If $\mathbf{b} - \mathbf{a}$ is not a root and $\mathbf{b} \neq \pm\mathbf{a}$, then

$$\mathcal{S} = \{\mathbf{b},\ \mathbf{b} + \mathbf{a},\ \ldots,\ \mathbf{b} + m\mathbf{a}\}.$$

In this case, in the notation of (i), $q = 0$ by Corollary 21.86(ii), so $\mathbf{b}(h_{\mathbf{a}}) = \gamma = -m$, and in conjunction with Corollary 21.86 (iii), we get

$$m = -2\frac{\langle \mathbf{a}, \mathbf{b}\rangle}{\langle \mathbf{a}, \mathbf{a}\rangle}. \tag{21.23}$$

One way of interpreting Remark 21.87(ii) is that the length of a string depends only on the inner product, independent of the other properties of L. Here is a quick but very important application.

COROLLARY 21.88. *$[L_{\mathbf{a}}L_{\mathbf{b}}] = L_{\mathbf{b}+\mathbf{a}}$ whenever $\mathbf{a}, \mathbf{b}$, and $\mathbf{b}+\mathbf{a}$ are roots.*

Proof. $[L_{\mathbf{a}}L_{\mathbf{b}}] \subseteq L_{\mathbf{b}+\mathbf{a}}$, which has dimension 1 by Theorem 21.79, so it remains merely to show that $[L_{\mathbf{a}}L_{\mathbf{b}}] \neq 0$. Take $q \in \mathbb{N}$ maximal such that $\mathbf{b}$, $\mathbf{b}-\mathbf{a}$, $\dots$, $\mathbf{b}-q\mathbf{a}$ are roots; then we expand this to a string of roots $\mathcal{S} = \{\mathbf{b}-q\mathbf{a}, \dots, \mathbf{b}-\mathbf{a}, \mathbf{b}, \mathbf{b}+\mathbf{a}, \dots\}$ of L. But since the length of the string depends only on $\mathbf{a}$ and $\mathbf{b}$ (cf. Equation (21.23)), the corresponding root spaces in the Lie subalgebra generated by $L_{\mathbf{b}-q\mathbf{a}}$ and $L_{\mathbf{a}}$ also are nonzero; in particular, the root space $[e_{\mathbf{a}}L_{\mathbf{b}}]$ of $\mathbf{b}+\mathbf{a}$ is nonzero, as desired. □

Example 21.89. Taking $e = e_{\mathbf{a}}$ and $f = f_{\mathbf{a}}$ as defined in Remark 21.81, let us compute $[f[ee_{\mathbf{b}}]]$.

Take $0 \neq v \in L_{\mathbf{b}+(m-q)\mathbf{a}}$, the last eigenspace in the string $\mathcal{S}$. The $h_{\mathbf{a}}$-eigenvalue for v is $-m$ by Remark 21.83′. Furthermore, $e_{\mathbf{b}} = \alpha \operatorname{ad}_f^{m-q} v$ for some $\alpha \in F$. Lemma 21.83(iii) could be rephrased as

$$efv_{k-1} = k(\gamma - k + 1)v_{k-1}.$$

Thus, taking $\alpha \operatorname{ad}_f^{m-q} v$ instead of v, so that $k = m - q$ and $\gamma = m$, we compute $[f[ee_{\mathbf{b}}]]$ as

$$\alpha \operatorname{ad}_f \operatorname{ad}_e \operatorname{ad}_f^{m-q} v = \alpha \operatorname{ad}_f(k(q+1) \operatorname{ad}_f^{m-q-1} v) = (q+1)(m-q)e_{\mathbf{b}}.$$

Recall that the **reflection** $\sigma_{\mathbf{a}}$ in $\mathfrak{h}^*$ is the map

$$\mathbf{v} \mapsto \mathbf{v} - 2\frac{\langle \mathbf{a}, \mathbf{v}\rangle}{\langle \mathbf{a}, \mathbf{a}\rangle}\mathbf{a}.$$

PROPOSITION 21.90. *Any reflection $\sigma_{\mathbf{a}}$ ($\mathbf{a} \in \Phi$) permutes the roots of each $\mathbf{a}$-string. Consequently, each reflection permutes the set of roots.*

Proof. The string has the form $\{\mathbf{b}, \mathbf{b}+\mathbf{a}, \dots, \mathbf{b}+m\mathbf{a}\}$ for $\mathbf{b} \in \Phi$, where $m = -2\frac{\langle \mathbf{a}, \mathbf{b}\rangle}{\langle \mathbf{a}, \mathbf{a}\rangle}$ by (21.23); then $\sigma_{\mathbf{a}}(\mathbf{b}) = \mathbf{b} + m\mathbf{a}$, and likewise

$$\sigma_{\mathbf{a}}(\mathbf{b} + i\mathbf{a}) = \mathbf{b} + (m+i)\mathbf{a} - 2i\mathbf{a} = \mathbf{b} + (m-i)\mathbf{a}.$$

The second assertion follows, since for any roots $\mathbf{b} \neq \pm\mathbf{a}$, $\sigma_{\mathbf{a}}(\mathbf{b})$ belongs to the $\mathbf{a}$-string of $\mathbf{b}$. □

Next we show that the bilinear form of (21.18) is defined over $\mathbb{Q}$.

THEOREM 21.91.

(i) *$\langle \mathbf{a}, \mathbf{a}\rangle > 0$ and $\langle \mathbf{a}, \mathbf{b}\rangle \in \mathbb{Q}$ for all nonzero roots $\mathbf{a}, \mathbf{b}$.*

(ii) *Write ${\mathfrak{h}^*}_0$ for the $\mathbb{Q}$-subspace of $\mathfrak{h}^*$ spanned by the roots. The bilinear form given by (21.18) restricts to a positive form on ${\mathfrak{h}^*}_0$, and $\mathfrak{h}^* = {\mathfrak{h}^*}_0 \otimes_{\mathbb{Q}} F$.*

Proof. (i) First we show that $0 < \langle \mathbf{a}, \mathbf{a}\rangle \in \mathbb{Q}$ for every root $\mathbf{a} \neq \mathbf{0}$. Writing $r = r_{\mathbf{a}} \in \mathfrak{h}$ as in Remark 21.74, recall from (21.18) that

$$\langle \mathbf{a}, \mathbf{a}\rangle = \langle r, r\rangle = \operatorname{tr}(\operatorname{ad}_r^2).$$

Since $\mathfrak{h}$ is Abelian, ad_r is 0 on $\mathfrak{h}$, and by definition acts as scalar multiplication by $\mathbf{b}(r)$ on $L_{\mathbf{b}}$. But $\mathbf{b}(r) \in \mathbb{Q}\mathbf{a}(r)$ by Lemma 21.76, since $r \in [L_{\mathbf{a}}L_{-\mathbf{a}}]$. Thus we can view $\operatorname{ad}_r = \mathbf{a}(r)D$, where D is a diagonal matrix whose entries are rational and not all 0 in view of Remark 21.78. Thus $\operatorname{tr}(\operatorname{ad}_r^2) = \mathbf{a}(r)^2\gamma$ where $\gamma = \operatorname{tr}(D^2)$ is the sum of rational squares, not all 0, so is positive. On the other hand, $\operatorname{tr}(\operatorname{ad}_r^2) = \langle r, r\rangle = \mathbf{a}(r)$ by Remark 21.74(iii), so we get

$$\mathbf{a}(r) = \mathbf{a}(r)^2\gamma,$$

implying that $\mathbf{a}(r) = \gamma^{-1} \in \mathbb{Q}^+$.

Now, for any roots $\mathbf{a}, \mathbf{b}$, Remark 21.78(iii) and Lemma 21.76 imply that $\langle \mathbf{a}, \mathbf{b}\rangle = \mathbf{b}(r_{\mathbf{a}}) \in \mathbb{Q}\mathbf{a}(r) = \mathbb{Q}$.

(ii) If we take a base $\mathbf{a}_1, \dots, \mathbf{a}_t$ of $\mathfrak{h}^*$ consisting of roots and write any root $\mathbf{a} = \sum \alpha_i \mathbf{a}_i$, we get

$$\langle \mathbf{a}, \mathbf{a}_j\rangle = \sum \alpha_i \langle \mathbf{a}_i, \mathbf{a}_j\rangle, \quad 1 \leq j \leq t,$$

a system of t equations in the α_i whose solution (by Cramer's rule) is in $\mathbb{Q}$ since all $\langle \mathbf{a}, \mathbf{a}_j\rangle, \langle \mathbf{a}_i, \mathbf{a}_j\rangle \in \mathbb{Q}$. □

Abstract root systems

Let us take stock of what we have done. We started with a semisimple Lie algebra over an algebraically closed field of characteristic 0. We extracted a root space decomposition with respect to a Cartan subalgebra, possessing a bilinear form defined over $\mathbb{Q}$ (and thus over $\mathbb{R}$). Let us describe this setup formally, putting roots in a more abstract context, in order also to permit later applications in other areas.

Definition 21.92. Suppose a f.d. vector space V over $\mathbb{R}$ is endowed with a positive definite bilinear form. A **root system** of V is a finite set $\Phi \subset V \setminus \{0\}$ satisfying the following properties:

(R1) Φ spans V.

(R2) If $\mathbf{a} \in \Phi$, then also $-\mathbf{a} \in \Phi$, but no other scalar multiple of $\mathbf{a}$ lies in Φ.

(R3) The reflection $\sigma_{\mathbf{a}}(\Phi) = \Phi$ for every $\mathbf{a} \in \Phi$.

The root system is called **crystallographic** if it also satisfies

(R4) If $\mathbf{a}, \mathbf{b} \in \Phi$ with $\mathbf{b} \neq \pm\mathbf{a}$ and $\mathbf{b} - \mathbf{a} \notin \Phi$, then $-2\frac{\langle \mathbf{a},\mathbf{b}\rangle}{\langle \mathbf{a},\mathbf{a}\rangle} \in \mathbb{Z}$.

The elements of Φ are called **roots**. The subgroup of isometries of V generated by the reflections $\{\sigma_{\mathbf{a}} : \mathbf{a} \in \Phi\}$ is called the **Weyl group** $\mathcal{W}$ of V.

The more group-theoretic aspect of this theory is given in Chapter 22 and its exercises.

Remark 21.93. In view of (R1), every element of the Weyl group $\mathcal{W}$ is determined by its action on Φ, so (R3) enables us to identify $\mathcal{W}$ with a subgroup of S_n, where $n = |\Phi|$. In particular, every element of $\mathcal{W}$ has finite order.

Example 21.94. The nonzero roots of a f.d. semisimple Lie algebra with respect to a Cartan subalgebra $\mathfrak{h}$ satisfy the conditions (R1)-(R4) (where $V = \mathfrak{h}^*$) in view of Theorem 21.91, Corollary 21.86(iv), Proposition 21.90, and Remark 21.87(ii) (explicitly Equation (21.23)), respectively.

Our objective at this stage is to develop various geometric properties of root systems based solely on these axiomatic properties.

In view of (R1), any root system Φ of V contains a suitable base $B = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ of V, which we fix temporarily. Writing $v = \sum_{i=u}^{n} k_i \mathbf{a}_i$ for $k_i \in \mathbb{R}$, $k_u \neq 0$, we say v is **negative** or **positive** according to the sign of k_u. (For example, $\mathbf{a}_1 - \mathbf{a}_2$ is positive.) In particular, the elements of B are all positive. This makes V an **ordered vector space** in the sense

that $(V, 0)$ is an ordered Abelian group whose order also respects scalar multiplication; i.e., if $\alpha \in \mathbb{R}^+$ and $v > 0$, then $\alpha v > 0$.

Our focus now is on V as an ordered vector space, although we are about to replace the base B by a nicer base with respect to the ordering. Writing P for the set of positive roots, we have $\Phi = P \cup \{\mathbf{0}\} \cup -P$. Let us define an abstract dependence relation on V by saying $v \in_{dep} S$ (for $v \in V$ and $S \subseteq V$) if v can be written as a linear combination of elements of S, using only positive coefficients. (Digression: This satisfies axioms (AD1), (AD2), and (AD4) of Definition 6.2 of Volume 1, but **not** the Steinitz exchange property (AD3)).

Since Φ is finite, P clearly contains a set S minimal with respect to $\mathbf{a} \in_{dep} S$ for all $\mathbf{a} \in P$. We fix this set S, which is called a **simple root system** and turns out to be particularly important. The following innocuous result is the key to this part of the theory.

LEMMA 21.95. *If $\mathbf{a} \neq \mathbf{b} \in S$ and $m_1, m_2 > 0$, then $m_1\mathbf{a} - m_2\mathbf{b}$ is* **not** *a root.*

Proof. Suppose $m_1\mathbf{a} - m_2\mathbf{b}$ were a root; switching $\mathbf{a}$ and $\mathbf{b}$ if necessary, we may assume that this root is positive and thus can be written as $\sum k_{\mathbf{c}}\mathbf{c}$ for suitable $\mathbf{c} \in S$, with all $k_{\mathbf{c}} > 0$. Thus

$$m_2\mathbf{b} + \sum_{\mathbf{c}\neq\mathbf{a}} k_{\mathbf{c}}\mathbf{c} + (k_{\mathbf{a}} - m_1)\mathbf{a} = 0.$$

But $m_2\mathbf{b} + \sum_{\mathbf{c}\neq\mathbf{a}} k_{\mathbf{c}}\mathbf{c}$ is positive, so $k_{\mathbf{a}} - m_1 < 0$ since V is ordered. Hence

$$\mathbf{a} = (m_1 - k_{\mathbf{a}})^{-1}\Big(m_2\mathbf{b} + \sum_{\mathbf{c}\neq\mathbf{a}} k_{\mathbf{c}}\mathbf{c}\Big),$$

a positive linear combination of other roots of S, contrary to the minimality of S. □

PROPOSITION 21.96. *$\langle \mathbf{a}, \mathbf{b}\rangle \leq 0$ for all $\mathbf{a} \neq \mathbf{b} \in P$.*

Proof. Since every element of P is a positive linear combination of elements of S, we may assume that $\mathbf{a}, \mathbf{b} \in S$. By (R3), $\sigma_{\mathbf{b}}(\mathbf{a}) = \mathbf{a} - 2\frac{\langle \mathbf{a},\mathbf{b}\rangle}{\langle \mathbf{b},\mathbf{b}\rangle}\mathbf{b}$ is a root. This contradicts Lemma 21.95 unless $\langle \mathbf{a}, \mathbf{b}\rangle \leq 0$. □

THEOREM 21.97. *Notation as above, the simple root system S is a base of V and is uniquely determined by the order on V.*

Proof. If $0 \neq v \in V$, then $\pm v \in P$, so v is spanned by S, by definition. We need to prove that S is independent. If $\sum k_i \mathbf{a}_i = 0$ for $\mathbf{a}_i \in S$ and $k_i \neq 0$, then letting $I = \{i : k_i > 0\}$ and $J = \{i : k_i < 0\}$ we have $\sum_{i\in I} k_i \mathbf{a}_i = -\sum_{j\in J} k_j \mathbf{a}_j$, so

$$0 < \left\langle \sum_{i\in I} k_i \mathbf{a}_i, -\sum_{j\in J} k_j \mathbf{a}_j \right\rangle = -\sum_{i\in I, j\in J} k_i k_j \langle \mathbf{a}_i, \mathbf{a}_j\rangle \leq 0$$

(since $k_i > 0$, $k_j < 0$, and $\langle \mathbf{a}_i, \mathbf{a}_j\rangle \leq 0$), a contradiction.

Thus S is a base. It remains to show that S is determined uniquely by the ordering $<$ on V. If S' is another simple root system and $s \in S$, then $s = \sum k_i s'_i$ for $s'_i \in S'$, all $k_i > 0$. But $s'_i = \sum k_{ij} s_j$ with $k_{ij} \geq 0$ and $s_j \in S$, yielding $s = \sum_{i,j} k_i k_{ij} s_j$. Since S is independent, we see that s is some s_{j_0}. Taking i' such that $k_{i'} k_{i'j_0} \neq 0$, we then must have $k_{i'j} = 0$ for all $j \neq j_0$; i.e., s'_i is a multiple of s_{j_0}, so (R2) implies that $s'_i = s_{j_0} = s$. This proves that $S \subseteq S'$ and, symmetrically, $S' \subseteq S$. □

From now on we change our base to be the simple root system $S = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$; i.e., $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the **simple roots**.

COROLLARY 21.98. *If an arbitrary root is written* $\mathbf{a} = \sum k_i \mathbf{a}_i$ *for* $k_i \in \mathbb{R}$, *then either all* $k_i \geq 0$ *or all* $k_i \leq 0$ *(depending on whether* $\mathbf{a}$ *is positive or negative).*

Proof. Any positive root $\mathbf{a}$ is written in this way, by definition of S, and likewise for $-\mathbf{a}$ when $\mathbf{a}$ is negative. □

Writing a positive root $\mathbf{a} = \sum k_i \mathbf{a}_i$ for $k_i \in \mathbb{R}$, we call $\ell = \sum k_i$ the **height** of the root $\mathbf{a}$, denoted height($\mathbf{a}$), and say the **support** of $\mathbf{a}$ is $\{\mathbf{a}_i : k_i \neq 0\}$. If only one $\mathbf{a}_i$ is in the support of $\mathbf{a}$, then $\mathbf{a} = \mathbf{a}_i$ by (R2). Surprisingly, we can get much more information about these coefficients k_i, and in the process learn exactly how to build the root system from the simple roots. The key is to use the reflections $\sigma_{\mathbf{a}_i}$.

PROPOSITION 21.99. *If the positive root* $\mathbf{a} = \sum k_i \mathbf{a}_i$ *is not simple, then* $\sigma_{\mathbf{a}_i}(\mathbf{a})$ *is still positive for each* i; *furthermore,* $\langle \mathbf{a}, \mathbf{a}_i\rangle > 0$ *for some* i, *so* $\sigma_{\mathbf{a}_i}(\mathbf{a})$ *is of lower height than* $\mathbf{a}$.

Proof. The root

$$\sigma_{\mathbf{a}_i}(\mathbf{a}) = \mathbf{a} - 2\frac{\langle \mathbf{a}, \mathbf{a}_i\rangle}{\langle \mathbf{a}_i, \mathbf{a}_i\rangle}\mathbf{a}_i \tag{21.24}$$

has the same coefficients as $\mathbf{a}$ except for $\mathbf{a}_i$, so these all remain positive by Corollary 21.98. It remains to note that $\langle \mathbf{a}, \mathbf{a}_i\rangle > 0$ for some $\mathbf{a}_i$, for otherwise

$$0 < \langle \mathbf{a}, \mathbf{a}\rangle = \left\langle \mathbf{a}, \sum k_i \mathbf{a}_i \right\rangle = \sum k_i \langle \mathbf{a}, \mathbf{a}_i\rangle \leq 0.$$

For this i, the coefficient of $\mathbf{a}_i$ is decreased in $\sigma_{\mathbf{a}_i}(\mathbf{a})$. □

COROLLARY 21.100. *The simple roots are precisely those positive roots of minimal height (which must be* 1*).*

Proof. In the lemma we saw that a nonsimple positive root can always be reflected to a positive root of lower height. Thus, the minimal height is attained precisely by the simple roots, which have height 1. □

The Cartan matrix of a simple root system.

Definition 21.101. The **Cartan numbers** are defined as

$$m_{ij} = 2\frac{\langle \mathbf{a}_i, \mathbf{a}_j\rangle}{\langle \mathbf{a}_j, \mathbf{a}_j\rangle}.$$

The **Cartan matrix** is the matrix (m_{ij}).

Thus $\sigma_{\mathbf{a}_j}(\mathbf{a}_i) = \mathbf{a}_i - m_{ij}\mathbf{a}_j$. These m_{ij} take on a special role in the theory. (The indices here are the reverse of Jacobson [**Jac1**], but that is not significant since we could take the transpose matrix.) Note that $m_{ii} = 2$. Thus our interest focuses on the case $i \neq j$.

Remark 21.101′. Permuting the roots, using a permutation π, would change the Cartan matrix A to $P_\pi A P_\pi^{-1}$, where P_π is the corresponding permutation matrix. We say that such a Cartan matrix is **equivalent** to A.

PROPOSITION 21.102. *The Cartan numbers satisfy the following properties:*

(i) $m_{ij}m_{ji} < 4, \quad \forall i \neq j$.

(ii) *Each* $m_{ij}m_{ji}$ *is integral over* $\mathbb{Z}$. *(In particular, if each* $\langle \mathbf{a}_i, \mathbf{a}_j\rangle \in \mathbb{Q}$, *then* $m_{ij}m_j \in \mathbb{Z}$.)

Proof. (i) $m_{ij}m_{ji} = 4\frac{\langle \mathbf{a}_i, \mathbf{a}_j\rangle^2}{\langle \mathbf{a}_i, \mathbf{a}_i\rangle\langle \mathbf{a}_j, \mathbf{a}_j\rangle}$, but the Cauchy-Schwarz inequality shows that $\langle \mathbf{a}_i, \mathbf{a}_j\rangle^2 < \langle \mathbf{a}_i, \mathbf{a}_i\rangle\langle \mathbf{a}_j, \mathbf{a}_j\rangle$ (since $\mathbf{a}_i$ and $\mathbf{a}_j$ are not proportional, in view of the definition of S).

(ii) Consider the subspace V_{ij} spanned by $\mathbf{a}_i$ and $\mathbf{a}_j$. Clearly $\sigma_{\mathbf{a}_i}$ and $\sigma_{\mathbf{a}_j}$ send V_{ij} to itself, since

$$\sigma_{\mathbf{a}_i}(\mathbf{a}_i) = -\mathbf{a}_i; \qquad \sigma_{\mathbf{a}_i}(\mathbf{a}_j) = \mathbf{a}_j - m_{ji}\mathbf{a}_i;$$
$$\sigma_{\mathbf{a}_j}(\mathbf{a}_i) = \mathbf{a}_i - m_{ij}\mathbf{a}_j; \qquad \sigma_{\mathbf{a}_j}(\mathbf{a}_j) = -\mathbf{a}_j.$$

Thus $\sigma_{\mathbf{a}_i}\sigma_{\mathbf{a}_j}$ restricts to a transformation T on V_{ij} of finite order (by Remark 21.93) whose matrix with respect to the base $\mathbf{a}_i, \mathbf{a}_j$ is

$$\begin{pmatrix} -1+m_{ij}m_{ji} & -m_{ij} \\ m_{ji} & -1 \end{pmatrix}.$$

Thus

$$\operatorname{tr}(T) = -2 + m_{ij}m_{ji},$$

which by Proposition 20.4(i) must be integral over $\mathbb{Z}$. □

Crystallographic root systems.

Now we bring in condition (R4).

Remark 21.103. By (R4), each Cartan number $m_{ij} \in \mathbb{Z}$, since Lemma 21.95 shows that the difference of two simple roots cannot be a root.

Now let us build an inductive procedure.

Remark 21.104. Suppose (R4) also holds, and consider a non-simple positive root $\mathbf{a} = \sum k_i \mathbf{a}_i$; note that all nonzero coefficients are positive by Corollary 21.98, and at least two of the coefficients are nonzero.

We use a two-pronged process to decrease height($\mathbf{a}$). Pick $\mathbf{a}_i \in \operatorname{supp} \mathbf{a}$ such that $\langle \mathbf{a}, \mathbf{a}_i \rangle > 0$; cf. Proposition 21.99.

(i) If $\mathbf{a} - \mathbf{a}_i$ is a root, then it is still positive in view of Corollary 21.98, and so the other coefficients remain positive; we replace $\mathbf{a}$ by $\mathbf{a} - \mathbf{a}_i$ whose height is height($\mathbf{a}$) $- 1$.

(ii) If $\mathbf{a} - \mathbf{a}_i$ is not a root, we replace $\mathbf{a}$ by the root $\sigma_{\mathbf{a}_i}(\mathbf{a})$. Condition (R4) implies that the coefficient of $\mathbf{a}_i$ has been decreased by an integer, so the height has decreased accordingly.

This process can only occur a finite number of times (the largest integer in height($\mathbf{a}$)), and our process can terminate only at a simple root.

We conclude by induction that each k_i is an integer.

Turning this result around, we have an inductive procedure to obtain all roots; namely, we start with all simple roots and repeatedly apply suitable reflections $\sigma_{\mathbf{a}_i}$, adding on simple roots where appropriate. There also is a unique "maximal" root; cf. Exercise 67.

Cartan's classification of semisimple Lie algebras

Our objective is to classify the semisimple Lie algebras (over an algebraically closed field of characteristic 0) in terms of Cartan matrices. In view of Example 21.94, we may use the results obtained above concerning crystallographic root systems, simple roots, and Cartan matrices.

Remark 21.105. For $i \neq j$, Equation (21.23) of Remark 21.87 interprets $-m_{ij}$ as the length of the $\mathbf{a}_i$-string of $\mathbf{a}_j$; in particular, $m_{ij} \leq 0$. Since each $m_{ii} = 2$, all $m_{ij} \in \{0, -1, \pm 2, -3\}$ in light of Proposition 21.96, Proposition 21.102, and Remark 21.103. Also, clearly $m_{ij} = 0$ iff $\mathbf{a}_i$ and $\mathbf{a}_j$ are orthogonal iff $m_{ji} = 0$.

Remark 21.106. Let us determine all the positive roots of a f.d. semisimple Lie algebra from the Cartan matrix and the simple roots; we proceed by induction on height, utilizing strings of roots. We just saw in Remark 21.104 that the roots of height 1 are precisely the simple roots $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Let us consider an arbitrary positive root $\mathbf{a}$ of height ≥ 2. Again by Remark 21.104, there is a positive root $\mathbf{b} = \mathbf{a} - k\mathbf{a}_i$ for some i, suitable $k \in \mathbb{N}^+$. By induction, we already "know" how to write $\mathbf{b} = \sum_{j=1}^n k_j \mathbf{a}_j$. Thus our task boils down to finding those k for which $\mathbf{b} + k\mathbf{a}_i$ is a root.

By Corollary 21.98, any root of the form $\mathbf{b} + k\mathbf{a}_i$ must be positive. The $\mathbf{a}_i$-string of $\mathbf{b}$ starts $\mathbf{b} - q\mathbf{a}_i, \mathbf{b} - (q-1)\mathbf{a}_i, \ldots$, for some $q > 0$, where these are positive of lower height, so we can recognize them by induction on height; in particular, we can determine q. But the length of the string until $\mathbf{b} + k\mathbf{a}_i$ is $q + k$, whereas the length of the full $\mathbf{a}_i$-string of $\mathbf{b}$ is

$$-2\frac{\langle \mathbf{a}_i, \mathbf{b} \rangle}{\langle \mathbf{a}_i, \mathbf{a}_i \rangle} = -\sum_{j=1}^n k_j m_{ij}.$$

So $\mathbf{b} + k\mathbf{a}_i$ is a root iff

$$-\sum_{j=1}^n k_j m_{ij} \geq q + k. \tag{21.25}$$

Remark 21.107. By (21.25) we see that $\mathbf{b} + u\mathbf{a}_j$ is also a root for all $0 \leq u \leq k$. We conclude by induction that if $\mathbf{a}$ has height ℓ there is a sequence of roots

$$\mathbf{a}_{i_1}, \quad \mathbf{a}_{i_1} + \mathbf{a}_{i_2}, \quad \ldots, \quad \mathbf{a}_{i_1} + \mathbf{a}_{i_2} + \cdots + \mathbf{a}_{i_\ell} = \mathbf{a}, \tag{21.26}$$

where each $\mathbf{a}_{i_j}$ is a simple root. For each positive root $\mathbf{a}$, we choose one such specific sequence $i_1, i_2, \ldots, i_\ell$ of indices.

The multiplication table for the semisimple Lie algebra.

We now use this information to recover the structure of semisimple Lie algebras from their root systems. In Chapter 22, we go one step further and see how the theory is encoded into the Weyl groups. When computing in Lie algebras, it is convenient to write $[a_1 a_2 \cdots a_m]$ in place of $[[\cdots [a_1 a_2] \cdots] a_m]$.

THEOREM 21.108. *Suppose $S = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is a simple root system for the semisimple Lie algebra L. As in Remark 21.81, define $e_i \in L_{\mathbf{a}_i}$, $e_i' \in L_{-\mathbf{a}_i}$ with $r_{\mathbf{a}_i} = [e_i e_i']$, and define*

$$f_i = \frac{2e_i'}{\langle \mathbf{a}_i, \mathbf{a}_i \rangle} \qquad \text{and} \qquad h_i = \frac{2r_{\mathbf{a}_i}}{\langle \mathbf{a}_i, \mathbf{a}_i \rangle} = [e_i f_i].$$

Then the following relations hold:

$$[e_i f_j] = \delta_{ij} h_i; \quad [e_i h_j] = -m_{ij} e_i; \quad [f_i h_j] = m_{ij} f_i; \quad [h_i h_j] = 0. \tag{21.27}$$

Furthermore, write any positive root $\mathbf{a} = \mathbf{a}_{i_1} + \cdots + \mathbf{a}_{i_\ell}$ as in (21.26), and

$$x_{\mathbf{a}} = [e_{i_1} e_{i_2} \cdots e_{i_\ell}], \qquad y_{\mathbf{a}} = [f_{i_1} f_{i_2} \cdots f_{i_\ell}].$$

Then $\{h_i : 1 \le i \le n\}$ together with the $x_{\mathbf{a}}$ and $y_{\mathbf{a}}$, taken over all positive roots $\mathbf{a}$, comprise a base of L.

Proof. For $i = j$, this is Remark 21.81, and the analogous computation using Corollary 21.72(ii), Remark 21.74, and Remark 21.78(i) shows that

$$[e_i h_j] = \frac{-2[r_{\mathbf{a}_j} e_i]}{\langle \mathbf{a}_j, \mathbf{a}_j \rangle} = -2m_{ij} e_i; \qquad [f_i h_j] = \frac{-2[r_{\mathbf{a}_j} f_i]}{\langle \mathbf{a}_j, \mathbf{a}_j \rangle} = 2m_{ij} f_i.$$

For $i \ne j$, $[e_i f_j] \in L_{\mathbf{a}_i - \mathbf{a}_j} = 0$ since $\mathbf{a}_i - \mathbf{a}_j$ is not a root by Lemma 21.95.

For the last assertion, we note by Corollary 21.88 that each $x_{\mathbf{a}}, y_{\mathbf{a}} \ne 0$; since the $L_{\mathbf{a}}$ are 1-dimensional (by Theorem 21.79), we see $L_{\mathbf{a}} = F x_{\mathbf{a}}$ and $L_{-\mathbf{a}} = F y_{\mathbf{a}}$. The h_i generate a subspace of $\mathfrak{h}$ of dimension $\dim \mathfrak{h}^* = \dim \mathfrak{h}$, so we are done by the root space decomposition of Summary 21.80. □

In this manner, the Lie structure of L is determined by its root system together with the inner product.

COROLLARY 21.109. *L is a vector space direct sum $\mathfrak{n}_+ \oplus \mathfrak{h} \oplus \mathfrak{n}_-$, where $\mathfrak{n}_+$ and $\mathfrak{n}_-$ are nilpotent Lie subalgebras (generated by the e_i and the f_i, respectively).*

Proof. If nonzero, $[x_{\mathbf{a}} x_{\mathbf{b}}] \in F x_{\mathbf{a}+\mathbf{b}}$, and likewise if nonzero, $[y_{\mathbf{a}} y_{\mathbf{b}}] \in F y_{\mathbf{a}+\mathbf{b}}$, so $\mathfrak{n}_+$ and $\mathfrak{n}_-$ are Lie subalgebras. Also, since there are only finitely many roots, any Lie product of sufficiently many of the elements of $\mathfrak{n}_+$ must be 0. □

Theorem 21.108 also enables us to compute the complete multiplication table of L.

COROLLARY 21.110. *The Lie multiplication table of L (with respect to the base in Theorem 21.108) has rational coefficients.*

Proof. The proof gives an explicit computation of the generators as rational expressions in the Cartan numbers m_{ij}. We start with the hardest part of the proof, which is computing $[e_{j_1} e_{j_2} \cdots e_{j_t}]$ for arbitrary $j_1, \dots, j_t$. Clearly this is 0 unless

$$\mathbf{a}_{j_1}, \quad \mathbf{a}_{j_1} + \mathbf{a}_{j_2}, \quad \dots, \quad \mathbf{a}_{j_1} + \mathbf{a}_{j_2} + \cdots + \mathbf{a}_{j_t} = \mathbf{a}$$

are all roots. Now write $\mathbf{a} = \mathbf{a}_{i_1} + \cdots + \mathbf{a}_{i_\ell}$ as in (21.26). We claim that $t = \ell$ and

$$[e_{j_1} e_{j_2} \cdots e_{j_\ell}] = s[e_{i_1} e_{i_2} \cdots e_{i_\ell}] \tag{21.28}$$

for some rational expression $s \in \mathbb{Q}$ in the m_{ij}, where $j_1, \dots, j_\ell$ is a permutation of $i_1, \dots, i_\ell$.

To see this, first note that $j_1, \dots, j_t$ must be a permutation of $i_1, \dots, i_\ell$, since $\mathbf{a}_1, \dots, \mathbf{a}_n$ are linearly independent. In particular, $t = \ell$, and $i_\ell = j_u$ for some $u \le \ell$; take the largest such u if this simple root repeats. If $u = \ell$, the claim follows by induction on ℓ, so assume that $u < \ell$. Let $f = f_{j_u}$. For each $j \ne j_u$, recalling that $[f e_j] = 0$, we get $\operatorname{ad}_f \operatorname{ad}_{e_j} = \operatorname{ad}_{e_j} \operatorname{ad}_f$ by Remark 21.4, and thus

$$[e_{j_1} e_{j_2} \cdots e_{j_\ell} f] = [e_{j_1} \cdots e_{j_{u-1}} e_{j_u} f e_{j_{u+1}} \cdots e_{j_\ell}]. \tag{21.29}$$

But Example 21.89 (where $\mathbf{b} = \mathbf{a}_{j_1} + \cdots + \mathbf{a}_{j_{n-1}}$) implies that

$$[e_{j_1} \cdots e_{j_{u-1}} e_{j_u} f] = k[e_{j_1} \cdots e_{j_{u-1}}] \tag{21.30}$$

for $0 \ne k \in \mathbb{Z}$, so combining (21.29), (21.30) yields

$$[e_{j_1} e_{j_2} \cdots e_{j_\ell} f] = k[e_{j_1} \cdots e_{j_{u-1}} e_{j_{u+1}} \cdots e_{j_\ell}]. \tag{21.31}$$

Another application of Example 21.89 yields

$$[e_{j_1} e_{j_2} \cdots e_{j_\ell} f e_{j_u}] = k'[e_{j_1} e_{j_2} \cdots e_{j_\ell}] \tag{21.32}$$

for suitable $0 \ne k' \in \mathbb{Z}$. Thus plugging (21.31) into (21.32) yields

$$[e_{j_1} e_{j_2} \cdots e_{j_\ell}] = \frac{k}{k'} [e_{j_1} \cdots e_{j_{u-1}} e_{j_{u+1}} \cdots e_{j_\ell} e_{j_u}].$$

But the right-hand side ends with $e_{j_u} = e_{i_\ell}$, so $j_1, \dots j_{u-1}, j_{u+1}, \dots, j_\ell$ must be a permutation of $i_1, \dots, i_{\ell-1}$. By induction on ℓ,

$$[e_{j_1} \cdots e_{j_{u-1}} e_{j_{u+1}} \cdots e_{j_\ell}] = s'[e_{i_1} \cdots e_{i_{\ell-1}}]$$

for some $s' \in \mathbb{Q}$. Hence,

$$[e_{j_1}e_{j_2}\cdots e_{j_\ell}] = \frac{k}{k'}s'[e_{i_1}\cdots e_{i_{\ell-1}}e_{j_u}] = \frac{k}{k'}s'[e_{i_1}\cdots e_{i_{\ell-1}}e_{i_\ell}],$$

proving (21.28).

The analogous argument works for the products $[f_{j_1}f_{j_2}\cdots f_{j_\ell}]$.

$[e_{\mathbf{a}}h_i]$ can be computed using Theorem 21.108, the Jacobi identity, and induction on length. Explicitly, writing $e_{\mathbf{a}} = [e_{\mathbf{b}}e_j]$, we have

$$[e_{\mathbf{a}}h_i] = [[e_{\mathbf{b}}h_i]e_j] + [e_{\mathbf{b}}[e_jh_i]] = [[e_{\mathbf{b}}h_i]e_j] - m_{ji}[e_{\mathbf{b}}e_j],$$

which can be calculated by induction on height($\mathbf{b}$). (See Exercise 63 for the precise formula.) So it remains to compute $[e_{\mathbf{a}}f_i]$ and $[f_{\mathbf{a}}e_i]$. Again, $[e_{\mathbf{a}}f_i]$ is easy using the Jacobi identity; indeed, writing $e_{\mathbf{a}} = [e_{\mathbf{b}}e_j]$, we see that

$$[e_{\mathbf{a}}f_i] = [[e_{\mathbf{b}}e_j]f_i] = [[e_{\mathbf{b}}f_i]e_j] - [[e_jf_i]e_{\mathbf{b}}] = [[e_{\mathbf{b}}f_i]e_j] + \delta_{ij}[e_{\mathbf{b}}h_j],$$

which is computed by induction and the previous case. Likewise for $f_{\mathbf{a}}e_i$.

□

Next we reduce to simple Lie algebras. We say a simple root system S is **decomposable** if it can be written as $S_1 \cup S_2$ where $\langle S_1, S_2\rangle = 0$.

THEOREM 21.111. *The f.d. semisimple Lie algebra L is simple iff its simple root system is indecomposable.*

Proof. If $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is orthogonal to $\{\mathbf{a}_{i_{k+1}}, \ldots, \mathbf{a}_{i_n}\}$, let

$$I = \mathfrak{h} + \sum_{1\le u\le k} L_{\mathbf{a}_{i_u}};$$

clearly $I \lhd L$, in view of (21.27).

Conversely, if L is not simple, then write $L = L_1 \oplus L_2$; then we can decompose each eigenspace $L_{\mathbf{a}} = L_{1,\mathbf{a}} \oplus L_{2,\mathbf{a}}$. Since each $L_{\mathbf{a}}$ is 1-dimensional, $L_{\mathbf{a}}$ lies in L_1 or L_2, and clearly $L_{-\mathbf{a}}$ lies in the same component (since $[L_{\mathbf{a}}L_{-\mathbf{a}}] \neq 0$), so we see that the simple root system partitions into two orthogonal components. □

Note 21.112. A Cartan matrix A is called **decomposable** if some equivalent Cartan matrix $P_\pi A P_\pi^{-1}$ can be partitioned in the form $\left(\begin{smallmatrix} ? & 0 \\ 0 & ? \end{smallmatrix}\right)$. The Cartan matrix of an indecomposable simple root system is indecomposable.

In conclusion, we have reduced the classification of f.d. semisimple Lie algebras (over an algebraically closed field of characteristic 0) to the determination of indecomposable Cartan matrices. In the next chapter, this classification is completed by means of the celebrated Dynkin diagrams.

Affine Lie algebras

The structure theory of f.d. semisimple Lie algebras is so satisfying that one might wonder what parts can be generalized to infinite dimensions. Perhaps the most natural approach is to focus on those properties of the Cartan matrix that permit us to establish the properties of roots in order to build the Lie algebra as in Theorem 21.108. This leads to the very important class of "affine Lie algebras," or "Kac-Moody algebras," considered independently by V. Kac, I. Kantor, I. Kaplansky, and R. Moody. We develop affine Lie algebras here, leaving the Kac-Moody-Kantor construction for Exercises 71ff. Also see Exercises 22.10 and 22.11.

Definition 21.113. A **generalized Cartan matrix** is an $n \times n$ matrix $A = (m_{ij})$ with integral entries satisfying $m_{ii} = 2$ but $m_{ij} \le 0$ for $i \neq j$, and for which $m_{ij} = 0$ iff $m_{ji} = 0$. The **Lie algebra of a generalized Cartan matrix** A is generated by a set $\{e_i, f_i, h_i : 1 \le i \le n\}$ of $3n$ elements, subject to the relations

$$\text{(21.33)} \qquad \mathrm{ad}_{e_i}^{1-m_{ij}} e_j = 0, \qquad \mathrm{ad}_{f_i}^{1-m_{ij}} f_j = 0, \quad \forall i \neq j;$$
$$[e_i, f_j] = \delta_{ij}h_i, \qquad [e_i, h_j] = -m_{ij}e_j, \qquad [f_i, h_j] = m_{ij}f_i, \qquad [h_i, h_j] = 0;$$

cf. Exercise 72.

Remark 21.105 implies that any Cartan matrix of a f.d. Lie algebra is a generalized Cartan matrix. The Lie algebra of a generalized Cartan matrix is indeed a Lie algebra, which is seen by reviewing the Lie structure over copies of $sl(2, F)$ as in Remark 21.81. It is easy to extend the main features from the finite-dimensional theory such as the root space decomposition, the root system, and the decomposition $\mathfrak{n}_+ \oplus \mathfrak{h} \oplus \mathfrak{n}_-$ of Corollary 21.109. This is summarized in Exercise 73. However, the dimension of the root system is the rank of A, which could be less than n.

From the definition, we can grade the Lie algebra in terms of the length of an element as a word in the e_i, f_i, and h_i, where the e_i have weight $+1$, the f_i have weight -1, and the h_i have weight 0. The theory could then be cast in terms of simple graded Lie algebras.

Generalized Cartan matrices are subdivided into three natural classes by means of a tool from linear programming. A vector $\mathbf{v} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^{(n)}$ is called **positive**, written $\mathbf{v} > 0$, if each $\alpha_j > 0$; likewise we write $\mathbf{v} \ge 0$ if each $\alpha_j \ge 0$. If $\mathbf{x} = (x_1, \ldots, x_n)$, we denote $\sum \alpha_j x_j$ by $\mathbf{v} \cdot \mathbf{x}$.

Remark 21.114. Suppose a generalized Cartan matrix A is indecomposable, as defined in Note 21.112. If $A\mathbf{x} \ge 0$ with $\mathbf{x} \ge 0$ but $\mathbf{x}$ not identically zero, then $\mathbf{x} > 0$. (Indeed, otherwise some $x_j = 0$; we may assume that

$x_1, \dots, x_k = 0$ and $x_{k+1}, \dots, x_n > 0$. For $1 \le i \le k$, each $m_{ij}x_j \le 0$ since $m_{ij} < 0$ for $i \ne j$; but $\sum_{j=1}^n m_{ij}x_j \ge 0$. It follows that $m_{ij} = 0$ whenever $1 \le i \le k$ and $k < j \le n$, implying also that $m_{ji} = 0$ for these i, j. Thus, A is decomposable, a contradiction.)

THEOREM 21.115. *Any indecomposable generalized Cartan matrix A is of one of the three following types:*

(1) *A is nonsingular, with $A\mathbf{w} > 0$ for some $\mathbf{w} > 0$.*
(2) *A has corank 1 (i.e., rank $n-1$) and $A\mathbf{w} = 0$ for some $\mathbf{w} > 0$; furthermore, if $A\mathbf{x} \ge 0$, then $A\mathbf{x} = 0$.*
(3) *$A\mathbf{w} < 0$ for some $\mathbf{w} > 0$; furthermore, if $A\mathbf{x} \ge 0$ with $\mathbf{x} \ge 0$, then $\mathbf{x} = 0$.*

Proof. Consider the cones $\mathcal{C} = \{\mathbf{x} : \mathbf{x} \ge 0\}$ and $\mathcal{C}_A = \{\mathbf{x} : A\mathbf{x} \ge 0\}$ in $\mathbb{R}^{(n)}$. Define the right annihilator $\operatorname{Ann}_r A = \{\mathbf{x} : A\mathbf{x} = 0\}$, a subspace of $\mathbb{R}^{(n)}$ contained in $\mathcal{C}_A$. Conversely, if $\mathcal{C}_A$ contains a subspace V of $\mathbb{R}^{(n)}$ and $\mathbf{x} \in V$, then $-\mathbf{x} \in V$ so $0 \le A(-\mathbf{x}) = -A\mathbf{x} \le 0$, implying that $A\mathbf{x} = 0$. Thus, $\operatorname{Ann}_r A$ is the largest subspace of $\mathbb{R}^{(n)}$ contained in $\mathcal{C}_A$.

Note that $w \in \mathcal{C} \cap \mathcal{C}_A$ iff $w \ge 0$ and $Aw \ge 0$. Suppose first that $\mathcal{C} \cap \mathcal{C}_A \ne 0$. Remark 21.114 implies that $\mathcal{C}_A$ can intersect the boundary of $\mathcal{C}$ only at the origin 0; hence $\mathcal{C}_A$ either is a line passing through the origin or is contained in the union of the origin with the interior of $\mathcal{C}$.

In the first instance, by the first paragraph, $\mathcal{C}_A$ is the right annihilator of A and is 1-dimensional, and we have proved (2).

In the second instance, we have $\mathcal{C}_A \subseteq \{0\} \cup \{\mathbf{x} : \mathbf{x} > 0\}$, which contains no nonzero subspaces, so $\operatorname{Ann}_r A = 0$; i.e., A is nonsingular of type (1).

Finally, we are left with the case $\mathcal{C} \cap \mathcal{C}_A = 0$. Then the cone $\mathcal{C}$ must intersect $-\mathcal{C}_A$ nontrivially, yielding (3). □

The types (1), (2), (3) of Theorem 21.115 are called **finite, affine**, and **indefinite**, respectively, and in the proof are easily seen to be mutually exclusive. The type of A is determined respectively by the existence of $\mathbf{w} > 0$ such that $A\mathbf{w} > 0$, $A\mathbf{w} = 0$, or $A\mathbf{w} < 0$.

COROLLARY 21.116. *Assume that the generalized Cartan matrix A also is symmetric. Then the symmetric bilinear form B on $\mathbb{R}^{(n)}$ defined by A is positive definite iff A has finite type, and is positive semidefinite (of corank 1) iff A has affine type.*

Proof. First suppose A is of finite type; i.e., $A\mathbf{w} > 0$ for some vector $\mathbf{w} > 0$. For any $0 \le \alpha \in \mathbb{R}$, $(A + \alpha I)\mathbf{w} > 0$, so $A + \alpha I$ remains of finite type and in particular is nonsingular. This also means that $-\alpha$ cannot be an eigenvalue of A, and thus the symmetric bilinear form B is positive definite.

If A is of affine type, then taking $\mathbf{w} > 0$ with $A\mathbf{w} = 0$, the same argument shows for all $\alpha > 0$ that $A + \alpha I$ is of finite type and in particular is nonsingular, so the symmetric bilinear form B is positive semidefinite.

Conversely, suppose B is positive semidefinite. Then indefinite type is impossible since otherwise taking $\mathbf{w} > 0$ and $A\mathbf{w} < 0$ we would have $\mathbf{w}^t A\mathbf{w} < 0$, a contradiction. Hence A is of finite type or affine type, depending respectively on whether A is nonsingular or singular. □

Thus, finite type corresponds to the Cartan matrices of f.d. semisimple Lie algebras. Another fact worth noting, whose proof relies on Farkas' theorem from linear inequalities (Exercise 79), is that the transpose matrix A^t has the same type as A; cf. Exercise 81. This gives rise to the **dual affine algebra**.

Physicists favor explicit constructions of affine Lie algebras. We start with two ways of extending Lie algebras. In each case we adjoin a one-dimensional space.

Remark 21.117. If δ is a derivation of a Lie algebra L, then $L \oplus F$ is a Lie algebra under the new Lie multiplication

$$[(a, \alpha), (b, \beta)] = ([ab] + \alpha\delta(b) - \beta\delta(a), 0). \tag{21.34}$$

Indeed, anticommutativity is immediate. The Jacobi identity is seen as follows: Extending ad_a to $L \oplus F$ by $\operatorname{ad}_a(b, \beta) = ([ab] - \beta\delta(a), 0)$, we note that $\operatorname{ad}_{(a,\alpha)} = (\operatorname{ad}_a + \alpha\delta, 0)$, and thus:

$$\begin{aligned}[\operatorname{ad}_{(a,\alpha)}, \operatorname{ad}_{(b,\beta)}] &= ([\operatorname{ad}_a, \operatorname{ad}_b] + \alpha[\delta, \operatorname{ad}_b] - \beta[\delta, \operatorname{ad}_a], 0) \\ &= \operatorname{ad}_{[a,b]} + \operatorname{ad}_{\alpha\delta(b)} - \operatorname{ad}_{\beta\delta(a)};\end{aligned}$$

this verifies Equation (21.3) by means of Equation (21.7).

This construction is known as **extending L by the derivation δ**. For notational convenience, we denote the element $(0, 1)$ by δ; intuitively, our extended algebra is $L \oplus F\delta$.

Remark 21.118. Suppose a bilinear form $\langle\ ,\ \rangle : L \times L \to F$ satisfies the following two equations for all $a, b, c \in L$:

(i) $\langle a, a\rangle = 0$;

(ii) $\langle [ab], c\rangle + \langle [bc], a\rangle + \langle [ca], b\rangle = 0$.

Then $L \oplus F$ is a Lie algebra under the Lie multiplication

$$[(a, \alpha), (b, \beta)] = ([ab], \langle a, b\rangle). \tag{21.35}$$

(Indeed, anticommutativity and the Jacobi identity are immediate from (i) and (ii) respectively, since $[[(a,\alpha),(b,\beta)],(c,\gamma)] = [([ab],\langle a,b\rangle),(c,\gamma)] = ([[ab]c],\langle [ab],c\rangle)$.) We call this construction **extending L by the bilinear form** $\langle\ ,\ \rangle$. For convenience, we denote $(0,1)$ by ϵ, which is central in the new Lie algebra.

We are ready for a major example.

Example 21.119. Let C denote the algebra $\mathbb{C}[\lambda,\lambda^{-1}]$ of Laurent polynomials. Define the **residue** $\mathrm{Res}(f)$ of $f = \sum_{j\geq m} c_j\lambda^j$ to be c_{-1}. (Thus $\mathrm{Res}\, f' = 0$.) Define the bilinear form on C:

$$\langle f,g\rangle = \mathrm{Res}(f'g).$$

We verify the conditions of Remark 21.118:

$\langle f,f\rangle = \mathrm{Res}(ff') = \frac{1}{2}\mathrm{Res}((f^2)') = 0.$

$\langle fg,h\rangle + \langle gh,f\rangle + \langle hf,g\rangle = 0$, since

$$0 = \mathrm{Res}((fgh)') = \mathrm{Res}(fgh') + \mathrm{Res}(f'gh) + \mathrm{Res}(fg'h).$$

Given a f.d. Lie algebra L, taking C as above, define the **Loop algebra** $\mathcal{L}(L) = C\otimes_{\mathbb{C}} L$, endowed with the Lie multiplication given by

$$[f\otimes a, g\otimes b] = fg\otimes [a,b], \qquad \forall f,g\in C \quad a,b\in L.$$

$\mathcal{L}(L)$ also has a bilinear form $\langle f\otimes a, g\otimes b\rangle = \langle f,g\rangle\otimes[a,b]$.

Let $\tilde{\mathcal{L}}$ be the extension $\tilde{\mathcal{L}}(L)\oplus F\epsilon$ of $\mathcal{L}(L)$ by this bilinear form, and let $\hat{\mathcal{L}}$ be the extension of $\tilde{\mathcal{L}}$ by the derivation $\delta = (\lambda\frac{\partial}{\partial\lambda}\otimes 1, 0)$ (i.e., $\delta(\epsilon)=0$). Thus

$$\hat{\mathcal{L}} = \mathcal{L}(L)\oplus F\epsilon\oplus F\delta.$$

$\hat{\mathcal{L}}$ is called the **affinization** of L. It turns out that $\hat{\mathcal{L}}$ is an affine Lie algebra, and its root system extends that of L.

One important instance is called the **Virasoro algebra**, defined in Exercise 78.

This sort of construction has received much recent attention; more generally, C is taken to be the **torus** $F[\lambda_1\lambda_1^{-1},\dots,\lambda_n,\lambda_n^{-1}]$ or even the quantum torus of Exercise 16A.3. When considering such constructions, it is convenient to pass to Lie algebras over fields that are not necessarily algebraically closed, such as the field of rational functions $F(\lambda_1,\dots,\lambda_n)$.

Appendix 21A. The Lie algebra of an algebraic group

In Appendix 19A we constructed the Lie algebra of a Lie group by means of the differential structure. In the case of linear algebraic groups, the same definition can be given much more easily in terms of derivations; we saw in Appendix 10A of Volume 1 that the tangent space of an affine variety can be described using derivations.

Definition 21A.1. Given an algebraic group G, define its **Lie algebra** $\mathrm{Lie}(G) = \{\delta\in \mathrm{Deriv}(\mathcal{A},\mathcal{A}) : \hat{\ell}_g\delta = \delta\hat{\ell}_g,\ \forall g\in G\}$, notation as in Equation 19B.2. Expressed in words, $\mathrm{Lie}(G)$ is the set of derivations of $\mathcal{A}$ that are **left invariant** in the sense that they commute with the left multiplication maps.

Although this definition is wonderfully concise, and is valid in any characteristic, it provides little immediate intuition. Still, one important piece of information can already be garnered from Volume 1.

Remark 21A.2. If $\delta_1,\delta_2\in\mathrm{Lie}(G)$, then the Lie product $[\delta_1,\delta_2]\in\mathrm{Lie}(G)$. Indeed, $[\delta_1,\delta_2]$ is a derivation by Proposition 6B.4 of Volume 1, and clearly is also left invariant if δ_1 and δ_2 are left invariant.

To gain more insight, one turns to the tangent space $T(G)_e$ at e. The idea is given in Exercises A1ff., and in Exercise A6 we recapture the classical examples of Lie algebras; the reader can find more detailed accounts of this approach in [**Hum2**] and [**Sp**].

Perhaps the most intuitive approach is to go in the other direction and consider the formal exponential map

$$e^{a\lambda} = 1 + \sum_{k=1}^{\infty}\frac{a^k}{k!}\lambda^k$$

in the formal power series algebra. Clearly $e^{a\lambda}e^{-a\lambda} = 1$, and we get the power series expansion

$$e^{a\lambda}be^{-a\lambda} = b + [a,b]\lambda + \dots;$$

in the language of Appendix 16A, Lie multiplication is obtained from the deformation of the adjoint with respect to the exponential map. However, one has to work quite hard to make this approach rigorous.

Appendix 21B: General introduction to nonassociative algebras

Although we jumped directly into the theory of Lie algebras in order to get directly to the theory of f.d. Lie algebras, it is enlightening to view the Lie theory in the general context of nonassociative algebras. Our point of departure is to define algebras in terms of generators and relations, starting with the "free" (nonassociative) algebra. We define a **magma** to be a set S with a binary operation $S \times S \to S$; this is the nonassociative version of a monoid. First we construct the "free" magma, mimicking Definition 17.2.

Definition 21B.1. Take an alphabet $X = \{x_i : i \in I\}$, and define **nonassociative words** by induction on **degree**. Each x_i is a nonassociative word of degree 1, and inductively, if w_1, w_2 are nonassociative words of respective degrees u_1, u_2, then $(w_1 w_2)$ is a nonassociative word of degree $u_1 + u_2$. Customarily, one deletes the outermost set of parentheses. For example, $(x_4x_1)((x_2(x_3x_4))x_2)$ is a nonassociative word of degree 6.

The **free magma** is the set of nonassociative words with multiplication given by restoring the outer set of parentheses in each multiplicand and juxtaposing. Note that we do *not* include the blank word, since we want to study algebras without 1.

Sometimes it is useful to refine the notion of degree, saying that $\deg_i x_j = \delta_{ij}$ (the Kronecker delta), and inductively defining $\deg_i(w_1w_2) = \deg_i w_1 + \deg_i w_2$. In other words, $\deg_i w$ is the number of times x_i appears in the word w. (Clearly $\deg w = \sum_i \deg_i w$.)

Assume throughout this discussion that C is a given commutative, associative ring. (Usually $C = \mathbb{Z}$ or C is a field.)

Definition 21B.2. The **free nonassociative C-algebra** $\mathcal{F}$ (on the alphabet X) is defined as the "magma algebra", i.e., the free C-module whose base is the free magma, endowed with multiplication extended from the magma multiplication via distributivity. The elements of $\mathcal{F}$ are called **(nonassociative) polynomials**.

Thus, any polynomial can be written uniquely in the form $f = \sum c_w w$, summed over all nonassociative words w, where $c_w \in C$ and almost all $c_w = 0$. We say that x_i **occurs** in f if $\deg_i w > 0$ for some w such that $c_w \neq 0$; i.e., x_i appears nontrivially in some word of f. We write $f = f(x_1, \ldots, x_t)$ to indicate that the letters $x_1, \ldots, x_t$ occur in f.

Analogously to Remark 17.34, we have the following "freeness" property:

Remark 21B.3. For any (nonassociative) algebra A and $\{a_i : i \in I\}$, one can define the **substitution homomorphism** $\mathcal{F} \to R$, sending $x_i \mapsto a_i$. The image of a polynomial $f(x_1, \ldots, x_t)$ under this substitution is denoted $f(a_1, \ldots, a_t)$.

For example, there is a homomorphism from $\mathcal{F}$ to the free associative algebra, obtained by erasing all parentheses. Remark 21B.3 shows us that any nonassociative algebra of cardinality $\aleph$ can be obtained by imposing suitable relations on the free nonassociative algebra on an alphabet of $\aleph$ letters. There is a special kind of relation that plays a key role, previewing Chapter 23.

Definition 21B.4. Suppose $f(x_1, \ldots, x_t)$ is a (nonassociative) polynomial. We say that $f = 0$ **holds identically** (in A) if $f(a_1, \ldots, a_t) = 0$, $\forall a_i \in A$. Sometimes, to avoid subscripts, we use x, y, z instead of x_1, x_2, x_3. In this case, we also say f is an **identity** of A, although a more precise usage of "identity" is formulated in Chapter 23. Likewise, $f = g$ **holds identically** in A for polynomials f and g if $f - g = 0$ holds identically in A.

Example 21B.5. (i) An algebra A is commutative iff $xy = yx$ holds identically, or equivalently, the Lie commutator $[x, y] = 0$ is an identity of A.

(ii) Analogously, we define the **associator** $[x, y, z] = (xy)z - x(yz)$; an algebra A is associative iff $(xy)z = x(yz)$ holds identically, iff $[x, y, z] = 0$ is an identity of A.

(iii) A (nonassociative) algebra is a Lie algebra iff it satisfies the two identities (anticommutativity and Jacobi) of Definition 21.1.

(iv) An important identity involving Lie brackets in associative algebras was given in Equation (21.2).

Many important classes of algebras are defined in terms of identities. Rather than embark now on a thorough study of identities of algebras, we content ourselves with a few observations as a foretaste of Chapter 23.

Definition 21B.6. A polynomial $f(x_1, \ldots, x_t)$ is **linear in** x_i if, writing $f = \sum c_w w$, we have $\deg_i w = 1$ for each word occurring in f; the polynomial $f(x_1, \ldots, x_t)$ is **multilinear** if f is linear in x_i for $1 \leq i \leq t$.

For example, the associator and the Lie commutator are multilinear. For the time being, we leave the easy verifications of the next assertion to the reader.

Remark 21B.7. (i) Given C-algebras A and H, $A \otimes_C H$ is an algebra over C with respect to the multiplication

$$(a_1 \otimes h_1)(a_2 \otimes h_2) = a_1 a_2 \otimes h_1 h_2,$$

the products taken in A and H respectively; cf. Remark 18.22.

(ii) Suppose H is a commutative associative C-algebra. Then $A \otimes_C H$ is an H-algebra under the operation $h(a_1 \otimes h_1) = a_1 \otimes h h_1$, as in Remark 18.27; furthermore, if a multilinear (not necessarily associative) polynomial f is an identity of A, then f is also an identity of $A \otimes_C H$.

One important instance of (ii) is when C is a field and H is the algebraic closure of C; this reduces much of the theory to algebras over algebraically closed fields.

Remark 21B.8. There is a technique called **linearization** for reducing any identity to a multilinear identity. This process, foreshadowed earlier in Remark 21.2, is described easily when our polynomial f is quadratic in x; we take another indeterminate x' and define

$$g(x, x', \dots) = f(x + x', \dots) - f(x, \dots) - f(x', \dots).$$

Linearization is described in more detail in Chapter 23.

Some important classes of nonassociative algebras

Having gone through the Lie theory in some detail, let us describe some other nonassociative algebras. All of the algebras considered here are obtained by weakening the identity of associativity. We say that an algebra A has a **multiplicative unit** 1 if $1a = a1 = a$ for all $a \in A$. Although we saw that a nontrivial Lie algebra cannot have a multiplicative unit, all of the algebras we consider here are compatible with multiplicative units. We write a^2 for aa.

Alternative algebras.

Our first condition is very close to associativity.

Definition 21B.9. An algebra A is **alternative** if

$$x^2 y = x(xy), \qquad yx^2 = (yx)x \tag{21B.1}$$

hold identically in A.

Remark 21B.10. Using associators, we have the identities $[x, x, y] = 0$ and $[y, x, x] = 0$. Linearizing (i.e., substituting x_3 for y and $x_1 + x_2$ for x), we get

$$[x_1, x_2, x_3] + [x_2, x_1, x_3] = 0, \qquad [x_3, x_1, x_2] + [x_3, x_2, x_1] = 0 \tag{21B.2}$$

holding identically, and since two transpositions generate S_3, we see that

$$[x_{\pi 1}, x_{\pi 2}, x_{\pi_3}] = (\operatorname{sg} \pi)[x_1, x_2, x_3]$$

holds identically for any permutation π of $\{1, 2, 3\}$. This justifies the name "alternative."

In particular, $[x, y, x] = -[x, x, y] = 0$, so we have the **flexible law** $(xy)x = x(yx)$, enabling us to write xyx without ambiguity. Similarly,

$$(x^2 y)x = \big(x(xy)\big)x = x\big((xy)x\big) = x(xyx)$$

holds identically.

Much of the alternative theory involves computing identities that are formal consequences of (21B.2) and thus "trivial," but not at all obvious to the uninitiated. See Exercise B9 for an instructive example (Moufang's identities), which leads to the lovely result of Artin (Exercise B11), that every subalgebra generated by two elements is associative. Alternative algebras come up naturally in the following notion.

Composition algebras.

Definition 21B.11. As in the associative theory, we define an **involution** of an algebra to be an anti-automorphism of order 2. Suppose $(A, *)$ is an F-algebra with multiplicative unit 1 and with involution fixing F. We say that $(A, *)$ is a **composition algebra (with involution)** if $aa^* = a^* a \in F$ for each $a \in A$ such that the quadratic form $Q(a) = a^* a$ gives rise to a nondegenerate bilinear form,

$$\langle a, b \rangle = Q(a + b) - Q(a) - Q(b) = a^* b + b^* a.$$

Remark 21B.12. Several observations are in order. The quadratic form Q is multiplicative since

$$Q(ab) = (ab)^* ab = b^* a^* ab = b^* Q(a) b = Q(a) Q(b).$$

Also $Q(a) = Q(a^*) = Q(-a)$.

Defining $T(A) = a + a^* = \langle a, 1 \rangle \in F$, we see at once that any $a \in A$ is quadratic over F, satisfying the equation $a^2 - T(A)a + Q(a) = 0$.

Composition algebras arise in connection with a result of Hurwitz; cf. Exercise B20. Also, Hamilton's quaternion algebra $\mathbb{H}$ (Example 14.33(v)) is a composition algebra. This can be extended to the following construction.

Definition 21B.13 (Cayley-Dickson). Given a composition algebra $(A,*)$ and $0 \neq \nu \in F$, we define the **ν-double** $\mathcal{A} = A \oplus A$ with multiplication as follows:

$$(a,b)(c,d) = (ac + \nu d^* b,\ bc^* + da).$$

Then we can identify A with $A \oplus 0 \subset \mathcal{A}$, and $(1,0)$ is the multiplicative unit of $\mathcal{A}$. (In particular, F is identified with $F \oplus 0 \subset \mathcal{A}$.)

Remark 21B.14. The involution $(*)$ extends to an involution on the ν-double $\mathcal{A}$ via $(a,b)^* = (a^*, -b)$; indeed $(^*)^2 = 1_{\mathcal{A}}$ and

$$\begin{aligned}((a,b)(c,d))^* &= (ac + \nu d^* b,\ bc^* + da)^* \\ &= (c^* a^* + \nu b^* d,\ -bc^* - da) = (c^*, -d)(a^*, -b) = (c,d)^*(a,b)^*.\end{aligned}$$

The quadratic form Q extends to a natural quadratic form on $(\mathcal{A},*)$. Indeed, given $\mathbf{a} = (a,b) \in \mathcal{A}$, define $Q(\mathbf{a})$ to be

$$\mathbf{a}^*\mathbf{a} = (a^*, -b)(a,b) = (a^*a - \nu b^* b, -ba + ba^{**}) = (Q(a) - \nu Q(b), 0) \in F.$$

Thus, $(\mathcal{A},*)$ is a composition algebra if the corresponding bilinear form is nondegenerate.

Example 21B.15. (i) Any field F of characteristic $\neq 2$ is itself a (commutative and associative) composition algebra, taking $(*)$ to be the identity.

(ii) The quadratic extension $\mathcal{F} = F(\sqrt{\nu}) = F \oplus F\sqrt{\nu}$ is the ν-double of F. The quadratic form Q is the norm of the quadratic extension, and is nonisotropic iff there is no solution to $\alpha^2 - \nu\beta^2 = 0$ in F; i.e., ν is not a square in F, which is the case iff $\mathcal{F}$ is a field. In particular, the field $\mathbb{C}$ is the (-1)-double of $\mathbb{R}$, and thus can be viewed as a composition algebra.

(iii) $\mathbb{H} = \mathbb{C} \oplus \mathbb{C}j$ is the (-1)-double of $\mathbb{C}$, and so is a composition algebra.

(iv) (Our main interest here) The (-1)-double of $\mathbb{H}$ is called the **octonion algebra** $\mathbb{O}$.

(v) More generally, the ν-double of a generalized quaternion algebra (α,β) (cf. Exercise 14.9) is called a **Cayley algebra** or **generalized octonion algebra**. The **split octonion algebra** is the 1-double of $(1,1)$.

By Exercise B14, any composition algebra $(A,*)$ is simple as a ring with involution.

Remark 21B.16. For any $0 \neq \mathbf{a} = (a,b) \in \mathbb{O}$, $Q(\mathbf{a}) = Q(a) + Q(b) > 0$, seen by taking $\nu = -1$ in Remark 21B.14. In particular, the bilinear form

of Q on $\mathbb{O}$ is nondegenerate, so $\mathbb{O}$ is a composition algebra. Furthermore, every nonzero element of $\mathbb{O}$ is invertible, since

$$\mathbf{a}^{-1} = \frac{1}{Q(a)}\mathbf{a}^*.$$

$\mathbb{O}$ is an alternative algebra, in view of the next proposition.

Remark 21B.17. Suppose $(A,*)$ is a composition algebra. To prove that A is alternative, it suffices to verify that $(a^*, a, b) = 0$ holds for all a, b in A. (Indeed, $(a + a^*, a, b) = 0$ since $a + a^* \in F$, so we would conclude that

$$(a,a,b) = (a + a^*, a, b) - (a^*, a, b) = 0,$$

yielding the first identity in (21B.1); the other is by symmetry, since $(b, a, a^*) = -(a, a^*, b^*)^* = 0$.)

PROPOSITION 21B.18. *Suppose the composition algebra $(\mathcal{A},*)$ is the ν-double of $(A,*)$. If A is associative, then $\mathcal{A}$ is alternative.*

Proof. Take $(a,b) \in \mathcal{A}$; then $(a,b)^* = (a^*, -b)$. By Remark 21B.17, one needs to show that $((a^*,-b),(a,b),(c,d)) = 0$ for all a,b,c,d in A. But

$$\begin{aligned}&\big((a^*,-b)(a,b)\big)(c,d) = (Q(a) - \nu Q(b), 0)(c,d) \\ &\quad = \big((Q(a) - \nu Q(b))c,\ d(Q(a) - \nu Q(b))\big),\end{aligned}$$

$$\begin{aligned}&(a^*,-b)\big((a,b)(c,d)\big) = (a^*,b)(ac + \nu d^* b, da + bc^*) \\ &\quad = \big(a^*(ac) + \nu a^*(d^* b) - \nu(a^* d^* + cb^*)b,\ (da + bc^*)a^* - b(c^* a^* + \nu b^* d)\big) \\ &\quad = \big(a^* ac - \nu cb^* b,\ daa^* - \nu bb^* d\big) \\ &\quad = \big((Q(a) - \nu Q(b))c,\ d(Q(a) - \nu Q(b))\big),\end{aligned}$$

and their difference is 0. □

It remains to show that $\mathbb{O}$ is nonassociative by reversing direction.

PROPOSITION 21B.19. *Suppose $(\mathcal{A},*)$ is the ν-double of $(A,*)$. If $\mathcal{A}$ is associative, then A must be commutative.*

Proof. $(0, ab) = ((a,0)(b,0))(0,1) = (a,0)((b,0)(0,1)) = (a,0)(0,b) = (0, ba)$, implying that $ab = ba$ for all $a, b \in A$. □

In particular, $\mathbb{O}$ is nonassociative and plays a key role in the theory of alternative algebras. Zorn showed that every simple nonassociative, alternative algebra is isomorphic to a Cayley algebra. His proof is by a careful study of idempotents and matrix units, but can be streamlined using generic methods; cf. Exercise B22. As we shall see, $\mathbb{O}$ also ties in to Lie algebras and Jordan algebras.

Jordan algebras.

One can carry out an analog of Lie theory using symmetric matrices as motivation rather than antisymmetric matrices. We give just a brief sketch of their structure, providing some more details in the exercises.

Definition 21B.20. A **(linear) Jordan algebra** over a field F of characteristic $\neq 2$ is a vector space together with a binary multiplication $\circ$ that is distributive over addition and satisfies the following two identities:

(i) (commutativity) $x \circ y = y \circ x$;

(ii) $(x^2 \circ y) \circ x = x^2 \circ (y \circ x)$.

Example 21B.21. (i) If R is an associative algebra with $\frac{1}{2}$, then R^+ is a Jordan algebra under the multiplication $a \circ b = \frac{1}{2}(ab + ba)$. Any Jordan subalgebra of a Jordan algebra of this form is called **special**.

We have introduced $\frac{1}{2}$ in order for the Jordan multiplication to agree with the associative multiplication:

$$1 \circ a = \frac{1}{2}(1a + a1) = \frac{1}{2}2a = a,$$

and likewise $a \circ a = \frac{1}{2}(aa + aa) = a^2$.

(ii) If $(R, *)$ is an associative algebra with involution, then the symmetric elements comprise a special Jordan subalgebra $\mathcal{S}(R, *)$ of R^+.

The assumption $\frac{1}{2} \in F$ is bothersome at times, and there is a characteristic-free version of the theory (cf. Exercise B31), inspired by the easily verified equality

$$ara = 2(a \circ (a \circ r)) - a^2 \circ r \tag{21B.3}$$

for all $a, r \in R^+$. Defining $U_a(r) = ara$, we could replace the Jordan composition $\circ$ by quadratic operators U_a and replace Definition 21B.20 by the appropriate identities; cf. Exercise B33. This provides us a way of defining abstract Jordan algebras in terms of the U_a also in characteristic 2; Jordan algebras so defined are called **quadratic Jordan algebras**; cf. Exercise B34. The quadratic concepts often motivate the Jordan theory, and the linear notation is rarely used in the literature. Another useful generalization involves triple products; cf. Exercise B41.

Recall that the Jordan **ideals** of a Jordan algebra J are the submodules I of J satisfying $a \circ r \in I$ for all $a \in I$, $r \in J$; these are the kernels of Jordan homomorphisms. J is **simple** if J has no proper nonzero Jordan ideals. This leads us to the classification question for simple Jordan algebras. Just as in the Lie case, we have some classical examples, all of which are subsumed in theorems of Herstein [**Hers3**]. Actually these theorems and proofs often are easier than the analogous Lie results. Here is a sample.

THEOREM 21B.22 (HERSTEIN). *Assume that R is an associative semiprime algebra, with $\frac{1}{2} \in R$. Any Jordan ideal $I \neq 0$ of R^+ contains a nonzero ideal of R. In particular, if R is simple, then R^+ is simple as a Jordan algebra.*

Proof. Suppose $0 \neq a, b \in I$, and $c = ab + ba \in I$. First we claim $Rc \subseteq I$. Indeed, for all r in R, Equation (21.2) shows that

$$\begin{aligned}[r, c] = [r, ab] + [r, ba] &= [r, a]b + a[r, b] + [r, b]a + b[r, a] \\ &= 2([r, a] \circ b + a \circ [r, b]) \in I.\end{aligned}$$

But clearly $r \circ c \in I$, so $rc = (\frac{1}{2}[r, c] + r \circ c) \in I$, as claimed.

It follows that $rcs + src = 2(rc) \circ s \in I$, for all $r, s \in R$. But $src \in Rc \subseteq I$, so $rcs \in I$. This proves that $RcR \subseteq I$, so we are done unless $c = 0$; i.e., $ab + ba = 0$ for all $a, b \in I$. In particular, $a^2 = 0$; hence, the RHS of Equation (21B.3) is 0, yielding $aRa = 0$. We conclude that $a = 0$, a contradiction. □

This gives us the simple Jordan analog of A_n, namely $M_n(F)^+$. Moreover, Herstein proved that $\mathcal{S}(R, *)$ is simple whenever R is a simple algebra with involution; cf. Exercise B39. This result subsumes Theorem 21B.22, since R^+ can also be viewed as $(R \oplus R^{\mathrm{op}}, \circ)$ where $\circ$ is the exchange involution. Thus, R^+ and $\mathcal{S}(R, *)$ gives us the simple Jordan analogs of A_n, B_n, C_n, and D_n for suitable associative R with involution.

There is another general class of simple Jordan algebras. Note that in any Jordan algebra (in characteristic $\neq 2$),

$$a \circ b = \frac{1}{2}\left((a + b)^2 - a^2 - b^2\right). \tag{21B.4}$$

Example 21B.23. Suppose (V, Q) is a nondegenerate quadratic space over a field F. Define a squaring operation on $F \oplus V$ by

$$(\alpha + v)^2 = \alpha^2 + Q(v) + 2\alpha v.$$

We then define $\circ$ by (21B.4). Also, we see for all $v, w \in V$ that

$$v \circ w = \frac{1}{2}(Q(v + w) - Q(v) - Q(w)) = \langle v, w \rangle,$$

where $\langle\ ,\ \rangle$ denotes the bilinear form corresponding to Q. Writing $a = \alpha + v$ and $b = \beta + w$, we have $a \circ b = \langle v, w\rangle + \alpha w + \beta v + \alpha\beta$; we verify that we have a Jordan algebra structure $J(V, Q)$ on $F \oplus V$, by checking Definition 21B.20:

(i) is obvious. To see (ii), we have

$$\begin{aligned} a^2 \circ (b \circ a) &= (a \circ b) \circ a^2 = (a \circ b) \circ (\alpha^2 + Q(v) + 2\alpha v) \\ &= (\alpha^2 + Q(v))a \circ b + 2\alpha(v \circ b) \circ v + 2\alpha^2(b \circ v) \end{aligned}$$

whereas

$$\begin{aligned} (a^2 \circ b) \circ a &= ((\alpha^2 + Q(v) + 2\alpha v) \circ b) \circ a \\ &= ((\alpha^2 + Q(v))b) \circ a + 2\alpha(v \circ b) \circ a \\ &= (\alpha^2 + Q(v))a \circ b + 2a(v \circ b) \circ v + 2\alpha^2(v \circ b). \end{aligned}$$

Any element $a = \alpha + v \in I$ is algebraic of degree 2, since

$$a^2 - 2\alpha a = Q(v) - \alpha^2 \in F. \tag{21B.5}$$

In particular, a is invertible unless $Q(v) = \alpha^2$. See Exercise B36 for more information of this sort.

$J(V, Q)$ is simple if $\dim V_F > 1$. Indeed, for any Jordan ideal I, we claim that $I \cap V \neq \emptyset$. Take $\alpha + v \in I$ for $\alpha \in F$ and $v \in V$. If $\alpha = 0$, we are done. If $\alpha \neq 0$, we can take $0 \neq w \in v^\perp \subset V$, implying $\alpha w \ = w \circ (\alpha + v) \in I$. Hence $w \in I \cap V$.

So take $0 \neq v \in I \cap V$, and taking v' with $\langle v, v'\rangle = 1$ yields $1 = v \circ v' \in I$.

Note that all the simple Jordan algebras thus far constructed are special; Example 21B.23 is the set of symmetric elements of the Clifford algebra of Example 18.40(iii). A non-special Jordan algebra is called **exceptional**. There is one very important example of an exceptional simple Jordan algebra, which we get from alternative algebras, as follows:

Suppose A is an alternative algebra over a field F. Then we can define the matrix algebra $M_n(A) \cong A \otimes_F M_n(F)$ in view of Remark 18.22′, but it need not be alternative. (In fact, if $M_n(A)$ is alternative and simple with $n \neq 2$, then A is associative.) If A also has an involution $(*)$, then $(*)$ extends to the involution $(*) \otimes (t)$ of $M_n(A)$ where (t) is the matrix transpose; i.e., $(a_{ij})^* = (a^*_{ji})$.

As before, let $\mathcal{S}(M_n(A), *)$ denote the algebra of symmetric elements under the Jordan multiplication $\circ$ given by $x \circ y = \frac{1}{2}(xy + yx)$. It turns out that $\mathcal{S}(M_n(A), *)$ is a Jordan algebra iff $n \leq 3$. We only get something new when $n = 3$ and A is a Cayley algebra; in this case, $\mathcal{S}(M_3(A), *)$ is

called an **Albert algebra**. In particular, $\mathcal{S}(M_3(\mathbb{O}, *)$ is a Jordan division algebra; the proof is a careful but straightforward computational verification of Definition 21B.20(ii). When the base field F is not algebraically closed, this class is expanded slightly by using cubic norm forms; the full story is given in the monograph of Jacobson [**Jac6, Chapter 2**]. Albert-Paige proved that Albert algebras are exceptional. This is seen most readily via Glennie's identity (Exercise B37), which holds for all special Jordan algebras but fails in Albert algebras.

By a deep theorem of Zelmanov, Albert algebras are the only exceptional simple Jordan algebras. This completes the classification of simple Jordan algebras: $\mathcal{S}(A, *)$ for $(A, *)$ simple alternative with involution, $J(V, Q)$, where Q is a quadratic form on V, and Albert algebras. This classification is the main result in Jacobson [**Jac6**].

Remark 21B.24. We say that an algebra A has **degree** $\leq d$ if each element of A is algebraic of degree $\leq d$ over F. For example, by Remark 21B.12, every noncommutative composition algebra has degree 2. By Equation (21B.5), any Jordan algebra $J(V, Q)$ of a quadratic form has degree 2. Jacobson [**Jac6, Chapter 2**] verifies that any Albert algebra has degree 3. Thus, the simple f.d. Jordan algebras of degree ≥ 4 are all of the form $\mathcal{S}(R, *)$ for R f.d. simple associative.

The Lie algebra of a Jordan algebra.

At first glance, Jordan algebras have more in common with associative algebras than do Lie algebras. Perhaps this is because $\mathcal{S}(R, *)$, for R an associative ring, contains the unit element 1, which plays a key role at times. On the other hand, there is an amazing connection between Jordan algebras and Lie algebras, discovered by Koecher, Tits and Kantor, by means of which we can exhibit all of the exceptional Lie algebras (to be defined in Chapter 22). The starting point is that $\operatorname{Deriv} A$ is a Lie algebra for any algebra A, by Example 21.7; here we take A alternative or Jordan.

Remark 21B.25. For a Jordan algebra J, let $\ell_a \colon J \to J$ denote the map $b \mapsto a \circ b$. Let us write $\tilde{J}$ for $\{\ell_a : a \in J\} \subseteq \operatorname{End} J$; then $\tilde{J}$ satisfies the identity

$$[[\ell_a \ell_b]\ell_c] = \ell_{a\circ(b\circ c) - b\circ(a\circ c)},$$

as seen by direct verification (given in [**Jac2, pp. 53, 54**]). It follows easily from the Jacobi identity that $\tilde{J} + [\tilde{J}, \tilde{J}]$ is a Lie subalgebra of $(\operatorname{End} J)^-$, called the **Lie multiplication algebra** of J, in which $[\tilde{J}, \tilde{J}]$ is a Lie ideal. (See Exercise B32.)

When J is the Albert algebra, Chevalley-Schafer showed that the Lie algebra corresponding to D_4 can be described as Deriv J, and the Lie multiplication algebra of J is the exceptional simple Lie algebra, called E_6; these examples lead to the beautiful unification of all the f.d. exceptional split Lie algebras over an algebraically closed field, due to Tits, to be described in Example 22.2. Kantor found a deep link between the Lie and Jordan theories by means of graded algebras.

Appendix 21C: Enveloping algebras of Lie algebras

As we have seen, representation theory involves the transfer of one algebraic structure (e.g., fields, groups, or Lie algebras) to another (associative algebras, often matrix algebras). This idea can be expressed as a universal, in the sense of Definition 8.5 of Volume 1. We have seen this with group representations in Remark 19.15, where the map $G \hookrightarrow \mathrm{Unit}(C[G])$ is viewed as a universal from a group G to the functor F (sending an algebra to its multiplicative group of units). In this appendix, we describe the analogous universal construction for Lie algebras.

Just as any faithful representation of a group is an injection into $\mathrm{End}_F V$, any faithful representation of a Lie algebra is an injection into $(\mathrm{End}_F V)^-$, leading to the following important notion.

Definition 21C.1. An associative algebra R is called an **enveloping algebra** of a Lie algebra L when L is isomorphic to a Lie subalgebra of R^- that generates R as an (associative) algebra.

To see that any Lie algebra has an enveloping algebra, we turn to the generic construction.

Universal enveloping algebras.

Definition 21C.2. Let $F: C\text{-}\mathbf{Alg} \to C\text{-}\mathbf{Lie}$ be the functor sending any associative algebra R (with 1) to its corresponding Lie algebra R^-. The universal $(U(L), \nu_L)$ (cf. Definition 8.5 of Volume 1) from a Lie algebra L to the functor F is called the **universal enveloping algebra** of L.

There are two facets to this definition. First, the "universal" aspect: $\nu_L: L \to U(L)^-$ should have the property that for any associative algebra R and any Lie homomorphism $f: L \to R^-$, there is a unique (associative) algebra homomorphism $g: U(L) \to R$ "lifting" f, i.e., such that $f = g\nu_L$. It follows at once that any enveloping algebra of L must contain a homomorphic image of $U(L)$.

Second, the "enveloping" aspect: ν_L should be 1:1. Although this will be seen to hold for Lie algebras, we shall see that the analog for Jordan algebras is false! Often the canonical map $\nu_L: L \to U(L)$ is understood.

Remark 21C.2′. Taking $R = M_n(F)$, any Lie representation $L \to gl_n(F)$ extends naturally to an algebra homomorphism $U(L) \to M_n(F)$. In particular, the trivial representation $L \to 0$ extends to the homomorphism $U(L) \to F$ sending $L \to 0$, whose kernel, $U(L)L$, is called the **augmentation ideal** of $U(L)$.

By Proposition 8.6 of Volume 1, all constructions of $U(L)$ are isomorphic, so we choose the most convenient one:

Example 21C.3. Let $T(L) = C \oplus L \oplus L^{\otimes 2} \oplus \cdots$ be the (associative) tensor algebra of L (viewed as a C-module), and let $\mathcal{I} \lhd T(L)$ be the ideal generated by all

$$\{a \otimes b - b \otimes a - [ab] : a, b \in L\}.$$

Let $U = T(L)/\mathcal{I}$. By construction, the images of $a \otimes b - b \otimes a$ and $[ab]$ are equal. The canonical Lie homomorphism $\nu_L: L \to U(L)^-$ is obtained by defining $\nu_L(a)$ to be the coset of a in U. Clearly, $\nu_L(L)$ generates U as an associative algebra, since L generates $T(L)$.

Next we check that U satisfies the universal property. For any Lie homomorphism $f: L \to R^-$ there is a unique homomorphism $T(L) \to R^-$ satisfying $a \mapsto f(a)$, whose kernel must contain $\mathcal{I}$, so Noether's Theorem yields the desired (unique) homomorphism $T(L)/\mathcal{I} \to R^-$ lifting f.

Having $U = U(L)$ in hand, we now simplify the notation by suppressing tensor products, and just write ab for the product of a and b in U. *Also, from now on, for convenience, we suppose C is a field F.* Taking a base $\{b_i : i \in I\}$ of L over F, we see that $U = \sum_{k \in \mathbb{N}} U_k$, where each U_k is spanned by

$$\{b_{i_1} b_{i_2} \cdots b_{i_k} : i_1 < i_2 < \cdots < i_k\}. \tag{21C.1}$$

We can get considerable information by studying the ring-theoretic structure of $U(L)$, especially in terms of positive filtrations, in the sense of Definition 7.29 of Volume 1.

Remark 21C.4. $U(L)$ has a positive $\mathbb{Z}$-filtration, where the k component is $\sum_{j \le k} U_j$. Thus, the k-component $\mathrm{gr}(U)_k$ of the associated graded algebra $\mathrm{gr}(U)$ is

$$\sum_{j \le k} U_j \Big/ \sum_{j \le k-1} U_j.$$

Since $b_ib_j = b_jb_i + [b_ib_j]$ in U_2 and $[b_ib_j] \in U_1$, we see in $\mathrm{gr}(U)$ that

$$\bar{b}_i\bar{b}_j = \bar{b}_j\bar{b}_i, \tag{21C.2}$$

writing $\bar{\ }$ for the homomorphic image of an element of U_k in $\mathrm{gr}(U)$. Since the $\bar{b}_i$ generate $\mathrm{gr}(U)$, we see that $\mathrm{gr}(U)$ is commutative, and since its relations are generated only by those of the form (21C.2), we conclude that $\mathrm{gr}(U)$ has no other relations than those implied by the fact that its generators commute; in other words, $\mathrm{gr}(U)$ is isomorphic to the commutative polynomial algebra $F[\lambda_i : \lambda_i \in I]$. For example, if $|I| = n$, then

$$\mathrm{gr}(U) \cong F[\lambda_1, \dots, \lambda_n].$$

It follows easily that $U(L)$ is a domain whose Jacobson radical is 0. Moreover, if L is f.d., then $U(L)$ is a Noetherian domain (cf. Exercise 7.14 of Volume 1), and $\mathrm{GKdim}(U(L)) \in \mathbb{N}$ by Corollary 17.49′. Even when L is infinite-dimensional, $U(L)$ is often an Ore domain; cf. Exercise C4.

Next we check that ν_L is 1:1; this fact is the celebrated Poincaré-Birkhoff-Witt (PBW) Theorem:

THEOREM 21C.5. *The map $\nu_L\colon L \to U(L)^-$ is 1:1, for any Lie algebra L.*

Proof. There are a variety of proofs, one of the most direct of which is based on the regular representation; cf. Exercise C1. Here is Bergman's quick, intuitive proof using his method of resolving ambiguities (Definition 17B.13). We take our base $\{b_i : i \in I\}$ of L and, well-ordering the index set I, consider the reductions $\{\rho_{ij} : i > j\}$ on the free algebra $F\{b_i : i \in I\}$, given by

$$\rho_{ij}(b_ib_j) = b_jb_i + [b_ib_j]$$

and fixing all other monomials. Since the reductions all have degree 2, the only ambiguities are overlap ambiguities; namely, on $b_ib_jb_k$, for $i > j > k$, we could apply ρ_{ij} or ρ_{jk} to get respectively

$$b_jb_ib_k + [b_ib_j]b_k; \quad b_ib_kb_j + b_i[b_jb_k].$$

But we can reduce $b_jb_ib_k + [b_ib_j]b_k$ further by applying ρ_{ik} and then ρ_{jk} to get

$$b_kb_jb_i + [b_ib_j]b_k + b_j[b_ib_k] + [b_jb_k]b_i, \tag{21C.3}$$

and we can reduce $b_ib_kb_j + b_i[b_jb_k]$ further by applying ρ_{ik} and then ρ_{ij} to get

$$b_kb_jb_i + b_i[b_jb_k] + [b_ib_k]b_j + b_k[b_ib_j]. \tag{21C.4}$$

The difference between (21C.3) and (21C.4) is

$$[b_ib_j]b_k + b_j[b_ib_k] + [b_jb_k]b_i - (b_i[b_jb_k] + [b_ib_k]b_j + b_k[b_ib_j]). \tag{21C.5}$$

But writing $[b_ib_j] = \sum_\ell \alpha_{ij\ell}b_\ell$ gives us $[b_ib_j]b_k - b_k[b_ib_j] = \sum \alpha_{ij\ell}(b_\ell b_k - b_kb_\ell)$; applying the ρ_{kl} for $k > \ell$ and $\rho_{\ell k}$ for $\ell > k$ shows that $[b_ib_j]b_k - b_k[b_ib_j]$ reduces to $\sum_\ell \alpha_{ij\ell}[b_\ell b_k] = [[b_ib_j]b_k]$. Thus, (21C.5) reduces to

$$[[b_ib_j]b_k] + [b_j[b_ib_k]] + [[b_jb_k]b_i],$$

which is 0 by the Jacobi identity, so both sides in (21C.3) have a common reduction; thus, by the Diamond Lemma (Proposition 17B.5), the reduction procedure is reduction-unique. But this reduction procedure does not act on $\nu_L(L) = U_1$, so this proves $\ker \nu_L = 0$, as desired. □

Remark 21C.6. There is a 1:1 natural correspondence between Lie representations of L and $U(L)$-modules, because of the universal property of $U(L)$. Hence, we can translate many of our earlier Lie-theoretic results to ring-theoretic properties of the $U(L)$. For example, in characteristic 0, Weyl's Theorem (21.58) says any $U(L)$-module that is f.d. over F is completely reducible. Further development of the Lie module theory can be found in [**Jac1, Chapters VII and VIII**] and [**Hum1, Chapter VI**].

Although here we have been dealing with associative algebras with 1, at times one also needs enveloping algebras without 1; in this situation, one just defines $T(L)$ without the initial copy of C and proceeds exactly as before.

Quantized enveloping algebras.

We discussed quantized algebras, also called **quantum algebras**, in Appendix 16A. These algebras are often called **quantum groups**, probably because they arise as coordinate algebras of algebraic groups, and over the last few years many treatises have been written about them. Quantized enveloping algebras enjoy many of the same ring-theoretic properties as the usual enveloping algebras, and have become an important tool in Lie theory. Since the theory of semisimple Lie algebras devolves from $sl(2, F)$, our main interest lies in the following example. We use the variant in Jantzen [**Jan**], which seems to be an improvement over the original formulas defining $[ef]$.

Definition 21C.7. We fix $q \neq 0$ in F, not a root of 1. The **quantized enveloping algebra** $U_q(sl(2, F)) = F\{e, f, k, k^{-1}\}$ is defined as the associative algebra with relations

$$ke = q^2ek; \quad kf = q^{-2}fk; \quad [ef] = \frac{k - k^{-1}}{q - q^{-1}}. \tag{21C.6}$$

Although the appearance of k^{-1} makes life more complicated, U has a commutative subalgebra $F[k, k^{-1}]$ that is invariant under conjugation by e and f. Often it is convenient to consider the generic case, taking $F = K(q)$, where q is a commuting indeterminate over a field K.

Remark 21C.8. In order to make computations manageable, we look for a good base of $U_q(sl(2, F))$. Toward this end, we refer to the reduction techniques of Appendix 17B. We consider the relations as reductions that move e to the left and/or f to the right:

$$ke \mapsto q^2ek; \quad k^{-1}e \mapsto q^{-2}ek^{-1}; \quad fk \mapsto q^2kf; \quad fk^{-1} \mapsto q^{-2}k^{-1}f;$$
$$fe \mapsto ef - \frac{k - k^{-1}}{q - q^{-1}}.$$

The only ambiguities that can arise are on words fke or $fk^{-1}e$. But

$$(fk)e \mapsto q^2kfe \mapsto q^2k\left(ef - \frac{k - k^{-1}}{q - q^{-1}}\right) \mapsto q^4ekf - q^2k\frac{k - k^{-1}}{q - q^{-1}},$$

whereas $f(ke) \mapsto fq^2ek \mapsto q^2\left(ef - \frac{k-k^{-1}}{q-q^{-1}}\right)k \mapsto q^4ekf - q^2k\frac{k-k^{-1}}{q-q^{-1}}$, which is the same; likewise, the other ambiguity has a common reduction. Hence, our system is reduction-unique, by Newman's Diamond Lemma (Proposition 17B.5).

The base is comprised of the elements $\{e^sk^{\pm t}f^u : s, t, u \in \mathbb{N}\}$.

Remark 21C.9. $U_q(sl(2, F))$ is $\mathbb{Z}$-graded, where $\deg(e) = 1$, $\deg(f) = -1$, and $\deg(k) = \deg(k^{-1}) = 0$, since the defining relations in (21C.6) are homogeneous. Note that the base in Remark 21C.7 is also homogeneous. Hence, most properties can be verified by checking the base elements. U is a noncommutative Noetherian domain.

One can generalize this example to quantum enveloping algebras for any semisimple Lie algebra, by means of quantizing the Cartan matrix of Theorem 21.108. This is done elegantly in Jantzen [**Jan, Chapter 4**]; cf. Exercise C20. Quantum algebras have important applications to physics, as to be described in Chapter 26.

The universal enveloping algebra of a Jordan algebra.

Since the Jordan representation theory likewise leads us to $M_n(F)^+$ and to universals, let us consider for a moment the Jordan analog $U(J)$ of the universal enveloping algebra. As in Example 21C.3, we can build $U(J)$ directly. Unfortunately, the natural map $J \to U(J)^+$ need not always be 1:1.

Indeed, if it were, then all Jordan algebras would be special, but we already have the exceptional class of Albert algebras obtained in Appendix 21B. This discussion may be continued in the context of identities of algebras, as noted in Example 23B.1.

Chapter 22

Dynkin Diagrams (Coxeter-Dynkin Graphs and Coxeter Groups)

This chapter has two main topics. The first is the theory of (Coxeter-) Dynkin diagrams, which enables us to complete the classification of the (split) simple f.d. Lie algebras over an algebraically closed field of characteristic 0, initiated in Chapter 21. We also consider briefly the affine case (defined in Theorem 21.115). Dynkin diagrams have a decidedly geometric flavor; they are tied to quadratic forms, which arise in several mathematical contexts, so they merit a separate treatment.

The exposition of f.d. semisimple Lie algebras in Chapter 21 utilized the Weyl group. In the second part of this chapter, we forget about the underlying Lie algebra and focus on the property that the Weyl group is generated by reflections, leading to the notion of **Coxeter group**. Coxeter groups are defined in terms of generators and relations, and have become quite pervasive in recent research. Although the setting of Coxeter groups is much more general than that of the Weyl group of a semisimple Lie algebra, we still tread the same path and get a **Coxeter graph** much like the Dynkin diagram, together with an attached bilinear form. The formulation is so simple that it has a wealth of other applications, one of which is given in Appendix 25C.

Dynkin diagrams

In view of Theorems 21.91, 21.108, and 21.111, we can determine the (split) simple f.d. Lie algebras by classifying the indecomposable simple root systems by means of their associated positive definite bilinear forms.

This can be done geometrically. Given an indecomposable simple root system $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ of $V = \mathbb{R}^{(n)}$, we normalize each vector to have length 1 by defining the vectors

$$\mathbf{e}_i = \frac{\mathbf{a}_i}{\sqrt{\langle \mathbf{a}_i, \mathbf{a}_i \rangle}}.$$

Thus, $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a base of V. Furthermore, $\forall i \neq j$, and

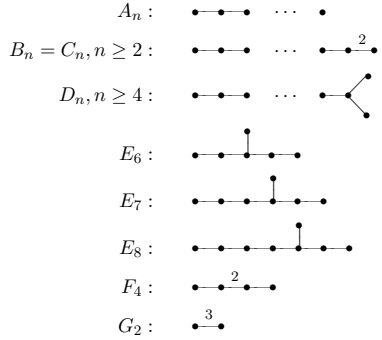
$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle \leq 0, \qquad 4\langle \mathbf{e}_i, \mathbf{e}_j \rangle^2 = m_{ij} m_{ji} \in \{0, 1, 2, 3\}; \tag{22.1}$$

cf. Proposition 21.102 and Equation (21.24). As in Jacobson [**Jac1**], a base of normal vectors satisfying conditions (22.1) is called a **configuration**.

Definition 22.1. The **Dynkin diagram** (also called the **Coxeter-Dynkin graph**) of a given configuration $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ consists of the undirected graph with vertices $\mathbf{e}_1, \dots, \mathbf{e}_n$, where vertices $\mathbf{e}_i \neq \mathbf{e}_j$ are connected by an edge iff $\langle \mathbf{e}_i, \mathbf{e}_j \rangle < 0$. In this case, the **multiplicity** of the edge is defined to be $4\langle \mathbf{e}_i, \mathbf{e}_j \rangle^2$ (which by Equation (22.1) is some number from 1 to 3); we write the multiplicity above the edge when it is 2 or 3.

Since any Dynkin diagram can be viewed as the disjoint union of its connected components, each of which is a Dynkin diagram for its corresponding root system, *we assume that all our Dynkin diagrams are connected*.

Example 22.2. Jumping ahead a bit, here is the celebrated list of (unweighted) connected Dynkin diagrams arising from positive definite bilinear forms:



The subscript always denotes the number of vertices.

We aim to show that this list comprises the only possible connected Dynkin diagrams for positive definite bilinear forms. To finish the classification, it then remains to construct a simple f.d. Lie algebra for every Dynkin diagram. In Exercises 1–4, the diagrams A_n, B_n, C_n, and D_n are realized as Dynkin diagrams of the classical Lie algebras of the same name. (The correspondence is not precisely 1:1, for the Lie algebras B_n and C_n have the same Dynkin diagram. These will be differentiated according to **weights** of roots, to be defined shortly.)

The Lie algebras for the remaining Dynkin diagrams E_6, E_7, E_8, F_4, and G_2, called **exceptional**, can be constructed via Remark 21B.25. Taking any alternative algebra A of degree 2 (i.e., a composition algebra) and any Jordan algebra J of degree ≤ 3, both defined over F, define the Lie algebra

$$L = \mathrm{Deriv}(A) \oplus \mathrm{Deriv}(J) \oplus (A_0 \otimes J_0),$$

where A_0 and J_0 denote the respective subspaces (in A and J) of elements having trace 0; define the Lie multiplication extending that of the Lie algebra $\mathrm{Deriv}(A) \oplus \mathrm{Deriv}(J)$ (with $[\mathrm{Deriv}(A), \mathrm{Deriv}(J)] = 0$) via the following rules, $\forall a \in A_0$, $v \in J_0$:

$$[\delta, a \otimes v] = \delta(a) \otimes v, \quad \forall \delta \in \mathrm{Deriv}\, A,\ a \in A,\ v \in J;$$
$$[\Delta, a \otimes v] = a \otimes \Delta(v), \quad \forall \Delta \in \mathrm{Deriv}\, J,\ a \in A,\ v \in J;$$

$[a \otimes v, b \otimes w]$ is described in Exercise 9. If A is the split octonion algebra $\mathcal{O}$ and J is respectively $\mathbb{C}$, $(M_3(\mathbb{C}), *)$ where $(*)$ is the transpose involution and $\mathcal{S}$ is defined in Example 21B.21(ii), $M_3(\mathbb{C})^+$, $\mathcal{S}(M_6(\mathbb{H}, *))$ where $(*)$ is the symplectic involution of Example 14.33, or $\mathcal{S}(M_3(\mathcal{O}, *))$, then L is respectively the (exceptional) Lie algebra for G_2, F_4, E_6, E_7, or E_8. Details can be found in Jacobson [**Jac3, pp. 89–103**], which contains more constructions as well as a description for the case of a non-algebraically closed base field. Their dimensions are respectively 14, 52, 78, 133, and 248 (as can be seen by counting their roots; see Samelson [**Sam, pp. 84–86**] for a description of their root systems).

We turn toward our main goal of deriving the list of Example 22.2.

Remark 22.3. By Equation (22.1), if $\langle \mathbf{e}_i, \mathbf{e}_j \rangle \neq 0$, then $\langle \mathbf{e}_i, \mathbf{e}_j \rangle \leq -\frac{1}{2}$, equality holding iff i and j are connected by an edge of multiplicity 1.

LEMMA 22.4. *The number ℓ of edges in a Dynkin diagram is less than n (not counting multiplicity).*

Proof. $0 < \langle \sum_{i=1}^n \mathbf{e}_i, \sum_{j=1}^n \mathbf{e}_j \rangle = \sum_{i=1}^n \langle \mathbf{e}_i, \mathbf{e}_i \rangle + 2 \sum_{i<j} \langle \mathbf{e}_i, \mathbf{e}_j \rangle \leq n - \ell$ by Remark 22.3, implying that $\ell < n$. □

We impose more restrictions on a Dynkin diagram by using Lemma 22.4 in conjunction with the following very useful observation:

Remark 22.5. If we erase a certain number of vertices from a Dynkin diagram as well as all edges emanating from these vertices, then we still are left with a Dynkin diagram (of the configuration obtained by removing these vectors); we call this a **subdiagram**.

LEMMA 22.6. *A Dynkin diagram cannot contain a cycle.*

Proof. Consider the subdiagram consisting of the vertices of the cycle; the number of edges would be at least the number of vertices, contradicting Lemma 22.4. □

PROPOSITION 22.7. *Each vertex in a Dynkin diagram has degree ≤ 3, counting multiplicities of edges.*

Proof. Otherwise, we would have a subdiagram consisting of some vertex $\mathbf{e}$ connected to four other vertices, say $\mathbf{e}_1, \ldots, \mathbf{e}_4$; in view of Remark 22.3,

$$\sum_{i=1}^4 4\langle \mathbf{e}, \mathbf{e}_i \rangle^2 \geq 4. \tag{22.2}$$

Note that $\mathbf{e}_1, \ldots, \mathbf{e}_4$ are orthogonal since otherwise, if some $\mathbf{e}_i$ were not orthogonal to $\mathbf{e}_j$, there would be the cycle $\mathbf{e}\mathbf{e}_i\mathbf{e}_j$. Thus, in the vector space

$$V = \bigoplus_{i=1}^4 F\mathbf{e}_i + F\mathbf{e},$$

$\{\mathbf{e}_1, \ldots, \mathbf{e}_4\}$ can be expanded to an orthonormal base $\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_4\}$. Since

$$\mathbf{e} = \langle \mathbf{e}, \mathbf{e}_0 \rangle \mathbf{e}_0 + \sum_{i=1}^4 \langle \mathbf{e}, \mathbf{e}_i \rangle \mathbf{e}_i,$$

we have $\langle \mathbf{e}, \mathbf{e}_0 \rangle \neq 0$ (since $\mathbf{e}$ is independent of $\mathbf{e}_1, \ldots, \mathbf{e}_4$), and

$$1 = \langle \mathbf{e}, \mathbf{e} \rangle = \Big\langle \mathbf{e}, \langle \mathbf{e}, \mathbf{e}_0 \rangle \mathbf{e}_0 + \sum_{i=1}^4 \langle \mathbf{e}, \mathbf{e}_i \rangle \mathbf{e}_i \Big\rangle = \langle \mathbf{e}, \mathbf{e}_0 \rangle^2 + \sum_{i=1}^4 \langle \mathbf{e}, \mathbf{e}_i \rangle^2,$$

implying that $\sum_{i=1}^4 \langle \mathbf{e}, \mathbf{e}_i \rangle^2 < 1$, contradicting (22.2). □

COROLLARY 22.8. *The only Dynkin diagram with a triple edge is G_2.*

Proof. Each of the two vertices of the triple edge has degree ≤ 3, and thus cannot belong to any other edge; hence, the graph has to terminate at both sides, and is G_2. □

Having disposed of triple edges, we need only consider single and double edges. The Dynkin diagram A_n (for some n) is also called a **chain**.

LEMMA 22.9. *Suppose a Dynkin diagram $\mathcal{D}$ has a subdiagram that is a chain with vertices $p, p+1, \ldots, q$ for suitable $p < q$. Then, letting $\mathbf{e} = \sum_{i=p}^{q} \mathbf{e}_i$, we have a configuration*

$$\{\mathbf{e}_1, \ldots, \mathbf{e}_{p-1}, \mathbf{e}, \mathbf{e}_{q+1}, \ldots, \mathbf{e}_n\}$$

obtained from $\mathcal{D}$ by contracting the chain to a point.

Proof. Note for $p \leq i, j \leq q$ that $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ unless $i = j$ or $|i - j| = 1$; in the latter case $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = -\frac{1}{2}$. Hence

$$\begin{aligned}\langle \mathbf{e}, \mathbf{e} \rangle &= \sum_{i,j=p}^{q} \langle \mathbf{e}_i, \mathbf{e}_j \rangle \\ &= \sum_{i=p}^{q} \langle \mathbf{e}_i, \mathbf{e}_i \rangle + \left(\sum_{i=p}^{q-1} \langle \mathbf{e}_{i+1}, \mathbf{e}_i \rangle + \langle \mathbf{e}_i, \mathbf{e}_{i+1} \rangle \right) \\ &= 1 + q - p - 2(q-p)\frac{1}{2} = 1.\end{aligned}$$

Furthermore, any other vertex i not in $\{p, \ldots, q\}$ is connected to at most one j in $\{p, \ldots, q\}$ (since otherwise we would have a cycle), and in this case $\langle \mathbf{e}_i, \mathbf{e} \rangle = \langle \mathbf{e}_i, \mathbf{e}_j \rangle$, yielding the desired conclusion. □

This technique has some powerful applications.

PROPOSITION 22.10. *Assume that $\mathcal{D}$ is a Dynkin diagram without a triple edge (i.e., $\mathcal{D} \neq G_2$).*

(i) *$\mathcal{D}$ cannot have two double edges.*
(ii) *If $\mathcal{D}$ has a double edge, then any vertex not on this double edge has degree 2.*
(iii) *At most one vertex of $\mathcal{D}$ has degree 3, unless there is a double edge.*

Proof. (i) Otherwise, the two nearest double edges are joined by a chain, which we shrink to a point to get a vertex of degree 4, contrary to Proposition 22.7.

(ii) Otherwise, the double edge and the nearest vertex (not on this double edge) of degree 3 are connected by a chain, which we could shrink to a point, and thus get a vertex of degree 4.

(iii) Otherwise, two vertices of degree 3 are joined by a chain, which we shrink to a point, and thus get a vertex of degree 4. □

We are ready for the classification.

THEOREM 22.11. *Any connected Dynkin diagram $\mathcal{D}$ has the form A_n, $B_n = C_n$, D_n, E_6, E_7, E_8, F_4, or G_2 of Example 22.2.*

Proof. We saw in Corollary 22.8 that if $\mathcal{D}$ has a triple edge, then $\mathcal{D} = G_2$. Thus, we may assume that $\mathcal{D}$ has no triple edges.

First assume that $\mathcal{D}$ has a double edge and n vertices. By Proposition 22.10, there is only one double edge, and all vertices not on this double edge have degree 2; thus, $\mathcal{D}$ looks like a chain except that the vertices p and $p+1$ are connected by a double edge for some p. Put

$$\mathbf{e} = \sum_{i=1}^{p} i\mathbf{e}_i, \qquad \mathbf{e}' = \sum_{j=1}^{n-p} j\mathbf{e}_{n+1-j}.$$

Then

(22.3)
$$\langle \mathbf{e}, \mathbf{e} \rangle = \sum_{i=1}^{p} i^2 - \sum_{i=1}^{p-1} 2i(i+1)\frac{1}{2} = p^2 - \sum_{i=1}^{p-1} i = p^2 - \frac{p(p-1)}{2} = \frac{1}{2}p(p+1);$$

likewise

(22.4)
$$\langle \mathbf{e}', \mathbf{e}' \rangle = \frac{1}{2}(n-p)(n-p+1).$$

Also $\langle \mathbf{e}, \mathbf{e}' \rangle = \langle p\mathbf{e}_p, (n-p)\mathbf{e}_{p+1} \rangle = p(n-p)\langle \mathbf{e}_p, \mathbf{e}_{p+1} \rangle$, where, by definition of double edge, $\langle \mathbf{e}_p, \mathbf{e}_{p+1} \rangle^2 = \frac{1}{2}$, so the Cauchy-Schwarz inequality yields

$$\frac{1}{2}p^2(n-p)^2 = \langle \mathbf{e}, \mathbf{e}' \rangle^2 < \langle \mathbf{e}, \mathbf{e} \rangle \langle \mathbf{e}', \mathbf{e}' \rangle = \frac{1}{4}p(p+1)(n-p)(n-p+1)$$

by Equations (22.3) and (22.4), and thus

$$p(n-p) < \frac{1}{2}(p+1)(n-p+1),$$

or $2p(n-p) < (p+1)(n-p) + p + 1$, in turn implying that $pn - n - p^2 < 1$, i.e.,

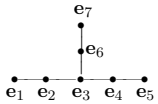
$$(p-1)(n-p-1) < 2.$$

If $p = 1$ or $p = n - 1$, then $\mathcal{D} = B_n$ (or its mirror image). Otherwise, each factor is 1; i.e., $p-1 = 1 = n-p-1$, implying $p = 2$ and $n = 4$, yielding F_4.

It remains to consider the case for which $\mathcal{D}$ has no double edges. Since chain is some A_n, we may assume that $\mathcal{D}$ is not a chain; then Proposition 22.10(iii) says that there is precisely one point of degree 3 with branches extending in three directions.

Although we could conclude the proof with the argument given below in the proof of Proposition 22.25, the ad hoc argument presented here is rather quick. We exclude certain subdiagrams by producing an isotropic vector $\mathbf{e}$, which is impossible. Recall that $\langle \mathbf{e}_i, \mathbf{e}_i \rangle = 1$, $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = -\frac{1}{2}$ for adjacent vertices, and all other $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$.

FIRST EXCLUSION:



Let $\mathbf{e} = 3\mathbf{e}_3 + 2(\mathbf{e}_2 + \mathbf{e}_4 + \mathbf{e}_6) + \mathbf{e}_1 + \mathbf{e}_5 + \mathbf{e}_7$; then

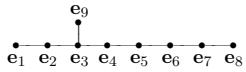
$$\langle \mathbf{e}, \mathbf{e} \rangle = 9 + 3 \cdot 4 + 3 \cdot 1 - \frac{2}{2}(2 \cdot 3 \cdot 3 + 2 \cdot 3) = 24 - 24 = 0.$$

SECOND EXCLUSION:

Let $\mathbf{e} = \mathbf{e}_1 + 2\mathbf{e}_2 + 3\mathbf{e}_3 + 4\mathbf{e}_4 + 3\mathbf{e}_5 + 2\mathbf{e}_6 + \mathbf{e}_7 + 2\mathbf{e}_8$; then

$$\langle \mathbf{e}, \mathbf{e} \rangle = 2(1+4+9) + 16 + 4 - \frac{2}{2}(2+6+12+12+6+2+8) = 48 - 48 = 0.$$

THIRD EXCLUSION:



Let $\mathbf{e} = 2\mathbf{e}_1 + 4\mathbf{e}_2 + 6\mathbf{e}_3 + 5\mathbf{e}_4 + 4\mathbf{e}_5 + 3\mathbf{e}_6 + 2\mathbf{e}_7 + \mathbf{e}_8 + 3\mathbf{e}_9$. Then

$$\begin{aligned}\langle \mathbf{e}, \mathbf{e} \rangle &= 4 + 16 + 36 + 25 + 16 + 9 + 4 + 1 + 9 \\ &\qquad - (8 + 24 + 30 + 20 + 12 + 6 + 2 + 18) \\ &= 120 - 120 = 0.\end{aligned}$$

Let us utilize these inclusions to describe our Dynkin diagram with three branches. The shortest branch has length 1, since otherwise there would be a subdiagram matching the first exclusion. Let t be the length of the next shortest branch; then $t \leq 2$, since otherwise there would be a subdiagram matching the second exclusion.

If $t = 1$, then we have D_n.

If $t = 2$, then the third branch has length < 5 by the third exclusion, so we are left with E_6, E_7, or E_8. □

Weights in Dynkin diagrams.

So far, our classification of f.d. simple Lie algebras has led us to a Cartan matrix or, equivalently, by Theorem 21.91, a positive definite bilinear form, and from this we obtained the Dynkin diagram. Conversely, how can we recover the Cartan numbers of Definition 21.101 from the Dynkin diagram? We need a bit more numerical information. For each simple root $\mathbf{a}$, we introduce the **weight** $\langle \mathbf{a}, \mathbf{a} \rangle$ and write it over the corresponding vertex of the Dynkin diagram. When an edge connects two roots of unequal weights, then we direct the edge from the heavier vertex to the lighter vertex. Note that

$$\frac{m_{ij}}{m_{ji}} = \frac{\langle \mathbf{a}_i, \mathbf{a}_i \rangle}{\langle \mathbf{a}_j, \mathbf{a}_j \rangle} \qquad \text{whenever} \qquad m_{ij} \neq 0. \tag{22.5}$$

Since in this case $m_{ij}m_{ji}$ is the multiplicity of the edge joining i and j, the Dynkin diagram together with the weights provide the Cartan matrix (and thus the multiplication table of the corresponding Lie algebra). Thus, we would like to find all possible weights for a given Dynkin diagram. Rescaling all the weights by the same constant factor clearly does not affect the m_{ij}, so we rescale the weights such that the smallest weight is 1.

Equation (22.5) shows that two adjacent vertices e_i and e_j have the same weight iff $m_{ij} = m_{ji}$, which by Equation (22.1) means $m_{ij} = m_{ji} = 1$ (since by hypothesis $m_{ij} \neq 0$); i.e., e_i and e_j are connected by a single edge. Thus, weights are irrelevant in the Dynkin diagrams A_n, D_n, E_6, E_7, and E_8.

When there is a multiple edge, we rely on Equation (22.5) to yield the ratio of the weights at its two ends. In G_2, where there is a triple edge, $m_{12}m_{21} = 3$, so $m_{12} \in \{1, 3\}$. We may as well choose the first weight to be 3 and the second to be 1, and again there is no ambiguity. Likewise in F_4 the product $m_{ij}m_{ji}$ at the double edge is 2, so again by symmetry one could take weights 2,2,1,1.

So the possible ambiguity is for the Dynkin diagram $B_n = C_n$, which could be $2, 2, \ldots, 2, 1$ or $1, 1, \ldots, 1, 2$; the former corresponds to the Lie algebra B_n and the latter to C_n.

Summary 22.12. Over an algebraically closed field of characteristic 0, any f.d. semisimple Lie algebra is a finite direct sum of Lie algebras of type A_n, $B_n (n \geq 2)$, $C_n (n \geq 3)$, $D_n (n \geq 4)$, E_6, E_7, E_8, F_4, and G_2, and each of these weighted Dynkin diagrams corresponds to a unique simple f.d. Lie algebra up to isomorphism.

Over non-algebraically closed fields F of characteristic 0, one can tensor by the algebraic closure and then use the preceding theory to obtain the appropriate Dynkin diagram. But one Dynkin diagram could correspond to infinitely many nonisomorphic simple Lie algebras over F.

Dynkin diagrams of affine Lie algebras.

Generalized Cartan matrices of affine type (cf. Theorem 21.115) also define "extended" Dynkin diagrams, corresponding to bilinear forms that need only be positive semidefinite. These extended Dynkin diagrams can be classified by means of the following analog of Proposition 22.10.

COROLLARY 22.13. *Suppose $\mathcal{D}$ is the extended Dynkin diagram of a simple affine Lie algebra.*

(i) *If any single vertex is erased, the remaining subdiagram is a disjoint union of Dynkin diagrams (of finite type).*

(ii) *(For $n \geq 3$) If $\mathcal{D}$ contains a cycle, then $\mathcal{D}$ is that cycle, and there are no multiple edges.*

Proof. (i) Cancelling a vertex corresponds to taking a principal submatrix, which we must prove is of finite type. For convenience, let A' denote the matrix obtained by cancelling the first row and column. Since the symmetric bilinear form arising from A' remains positive semidefinite, A' has either affine or finite type, and we shall eliminate affine type by showing that it leads to a contradiction.

Suppose $\mathbf{x} > 0$ and $A'\mathbf{x}' = 0$, where $\mathbf{x}'$ is the vector obtained by cancelling the first component of $\mathbf{x}$. By hypothesis, $A\mathbf{x} \geq 0$, so Theorem 21.115(2) implies that $A\mathbf{x} = 0$, in turn yielding $a_{1j} = 0$ for all $j > 1$. But then also $a_{j1} = 0$ for all $j > 1$, implying that A is decomposable, a contradiction.

(ii) If on the contrary $\mathcal{D}$ properly contains a cycle, then the subdiagram obtained by erasing all other vertices contains that cycle, which is impossible since the subdiagram must be of finite type (by (i)). Thus, $\mathcal{D}$ is a cycle. We label its vertices $\mathbf{e}_1, \ldots, \mathbf{e}_n$, with $\langle \mathbf{e}_i, \mathbf{e}_{i+1} \rangle \leq -\frac{1}{2}$, subscripts mod n, equality holding iff the corresponding edge has multiplicity 1. Thus, for

$\mathbf{e} = \sum_{i=1}^n \mathbf{e}_i$,

$$0 \leq \langle \mathbf{e}, \mathbf{e} \rangle = \sum_{i=1}^n \langle \mathbf{e}_i, \mathbf{e}_i \rangle + 2 \sum_{i=1}^n \langle \mathbf{e}_i, \mathbf{e}_{i+1} \rangle \leq n - 2\frac{n}{2} = 0,$$

implying that each $\langle \mathbf{e}_i, \mathbf{e}_{i+1} \rangle = -\frac{1}{2}$, as desired. □

Note that the "exclusions" in the proof of Theorem 22.11 no longer apply in the affine case, and indeed they do occur in certain extended Dynkin diagrams. A complete list of extended Dynkin diagrams for simple affine Lie algebras is given in Kac [**Kacv4, p. 54**].

Reflection groups

Now we switch gears and consider various classes of groups that are motivated by the Lie theory but arise in many other situations. Our exposition follows the beautiful treatment by Humphreys [**Hum3**]. Write $V = \mathbb{R}^{(n)}$, endowed with a positive definite symmetric bilinear form. $O(V)$ denotes the group of orthogonal transformations of V. Recall from Definition 0A.7 of Volume 1 that the **(hyperplane) reflection** $\sigma_b \colon V \to V$ **with respect to** $b \in V$ is defined by

$$\sigma_b(v) = v - \frac{2\langle v, b \rangle}{\langle b, b \rangle} b.$$

Clearly $\sigma_b \in O(V)$; also see Exercise 11.

Definition 22.14. A **reflection group** is a subgroup of $O(V)$ that is generated by hyperplane reflections.

$O(V)$ itself is a reflection group, by Dieudonné's Theorem (proved in Exercise 0A.16 of Volume 1). The Weyl group of a split semisimple Lie algebra, generated by reflections with respect to simple roots, is our current motivating example of a reflection group. The basic question here is, "How much of the preceding theory is encapsuled in the reflection group?"

To address this question, our first task is to associate a root system Φ of vectors (cf. Definition 21.92) to any reflection group $\mathcal{W}$.

LEMMA 22.15. *For any root $\mathbf{b}$, $\sigma_{w\mathbf{a}} = w\sigma_{\mathbf{a}}w^{-1}$ for each $w \in \mathcal{W}$.*

Proof. $w\sigma_{\mathbf{a}}w^{-1}$ sends $w\mathbf{a} \mapsto -w\mathbf{a}$, so we need to show that it fixes $(w\mathbf{a})^\perp$. If $\langle v, w\mathbf{a} \rangle = 0$, then $\langle w^{-1}v, \mathbf{a} \rangle = 0$, so $w^{-1}v$ is fixed by $\sigma_{\mathbf{a}}$; hence

$$w\sigma_{\mathbf{a}}w^{-1}(v) = ww^{-1}v = v,$$

as desired. □

Remark 22.16. Here is our procedure for constructing a root system from a reflection group $\mathcal{W}$ acting on V. Pick some $\mathbf{a} \in V$ and let $\Phi = \mathcal{W}\mathbf{a} = \{w\mathbf{a} : w \in \mathcal{W}\}$. Replace V by the subspace spanned by Φ. Condition (R1) of Definition 21.92 now holds automatically, as does (R3). To check (R2), note that if $w\mathbf{a} = \alpha\mathbf{a}$ for $\alpha \in \mathbb{R}$, then $|\alpha| = 1$ since w is an isometry, and consequently $\alpha = \pm 1$.

(Note that we modified V in the course of Remark 22.16.) Having defined the root system Φ for the reflection group $\mathcal{W}$, we can apply the proof of Theorem 21.97 and the discussion preceding it to get a simple root system that is a base of V. Exercise 14 shows how to regard the classical Lie algebras from this point of view.

It remains for us to build diagrams from reflection groups. One key ingredient could be Condition (R4) of Definition 21.92; in this case, we also say that the group $\mathcal{W}$ is **crystallographic**. Instead, let us turn now to an approach that does not rely on root systems.

Coxeter groups and Coxeter graphs.

Definition 22.17. A **Coxeter group** is a (not necessarily finite) group $\mathcal{W}$ generated by a set $S = \{s_i : 1 \le i \le n\}$ of elements each having order 2, together with relations of the form $(s_i s_j)^{m_{ij}} = 1$; thus, m_{ij} to be the order of the element $s_i s_j$. (Formally, $m_{ij} = \infty$ if $s_i s_j$ has infinite order.) S is called the set of **Coxeter generators**, and $(\mathcal{W}, S)$ is called a **Coxeter system**.

For example, the dihedral group and symmetric groups are Coxeter groups; cf. Example 17A.2. Note that $m_{ii} = 1$ for all i, but $m_{ij} > 1$ whenever $i \neq j$. It turns out that any finite reflection group is Coxeter (cf. Exercise 19), and the remainder of this chapter deals with Coxeter systems. Although we introduce Coxeter groups as an outgrowth of Lie theory, this class of groups has become an object of intense interest in its own right, with varied applications in geometry and combinatorics.

Note. These m_{ij} are not the m_{ij} of Definition 21.113. In fact, this definition implies $m_{ij} = m_{ji}$, since in any group, $(ab)^m = 1$ iff $(ba)^m = 1$. See Exercise 23 for a comparison of these numbers.

Definition 22.18. The **Coxeter graph** of a Coxeter system $(\mathcal{W}, S)$ has vertices corresponding to the Coxeter generators $\{s_1, \dots, s_n\}$, with vertices s_i, s_j connected by an edge whenever $m_{ij} \ge 3$. (In particular, $i \neq j$ since $m_{ii} = 1$.) One writes m_{ij} over an edge when $m_{ij} > 3$. Thus, an unlabelled edge between s_i and s_j indicates $m_{ij} = 3$ (and the absence of an edge indicates $m_{ij} = 2$).

Certain Coxeter graphs may fail to be attached to semisimple Lie algebras; cf. Exercise 24. Nevertheless, we can recapture most of the theory of Dynkin diagrams (i.e., Coxeter-Dynkin graphs) and obtain a full classification of Coxeter graphs corresponding to finite Coxeter groups. First, we note that it is enough to consider connected graphs.

Remark 22.19. If $s_i \neq s_j \in S$ are not joined by an edge in the Coxeter graph, then $s_i s_j s_i s_j = (s_i s_j)^2 = 1$, implying that $s_j s_i = s_i s_j$; i.e., s_i and s_j commute. Thus, Coxeter generators from different components commute; we conclude that a Coxeter group $\mathcal{W}$ is a direct product of Coxeter groups whose graphs are the connected components of the Coxeter graph of $\mathcal{W}$.

Definition 22.20. Notation as in Definition 22.18, put $V = \mathbb{R}^{(n)}$ with standard base $e_1, \dots, e_n$. Define the **Coxeter bilinear form** on V by

$$\langle e_i, e_j \rangle = -\cos \frac{\pi}{m_{ij}}. \tag{22.6}$$

(Thus, $\langle e_i, e_j \rangle = -1$ iff $m_{ij} = \infty$.)

(Conversely, we can use (22.6) to recover any Coxeter graph from its Coxeter bilinear form.) We can turn the circle of ideas, as follows:

Definition 22.21. Given a Coxeter system $(\mathcal{W}, S)$ and its Coxeter bilinear form, define the **reflection** σ_i for each $1 \le i \le n$ by

$$\sigma_i(v) = v - 2\langle e_i, v \rangle e_i.$$

THEOREM 22.22.

(i) *For any Coxeter group $\mathcal{W}$, the map $\rho\colon \mathcal{W} \to \mathrm{GL}(n, \mathbb{R})$ given by $\rho(s_i) = \sigma_i$ is a group representation that respects the Coxeter bilinear form.*

(ii) *$\circ(\sigma_i \sigma_j) = m_{ij}$ for all $1 \le i, j \le n$.*

(iii) *For $i \neq j$, the Coxeter bilinear form restricts to a positive semidefinite form on the two-dimensional subspace $Fe_i + Fe_j$, which is positive definite iff $m_{ij} < \infty$.*

Proof. The proof goes backwards. Let $m = m_{ij}$ and $V_{ij} = Fe_i + Fe_j$, clearly invariant under σ_i and σ_j. If $v = \alpha e_i + \beta e_j \in V_{ij}$, then

$$\langle v, v \rangle = \alpha^2 + \beta^2 - 2\alpha\beta \cos \frac{\pi}{m} = \left(\alpha - \beta \cos \frac{\pi}{m}\right)^2 + \beta^2 \sin^2 \frac{\pi}{m},$$

so V_{ij} is positive definite iff $\sin \frac{\pi}{m} \neq 0$, i.e., $m < \infty$, proving (iii). Moreover, $\mathbb{R}^{(n)} = V_{ij} \perp V_{ij}^{\perp}$, the latter component fixed by σ_i and σ_j, so $\circ(\sigma_i \sigma_j)$ can

be computed on the plane V_{ij}, But $\sigma_i\sigma_j$ restricted to V_{ij} is just the rotation of angle $\frac{2\pi}{m}$, which has order m. This proves (ii).

(i) follows formally from (ii), for the canonical homomorphism from the free group (generated on S) to $\mathrm{GL}(n, F)$ factors through the Coxeter relations. □

Theorem 22.22 readily implies that the Coxeter bilinear form of a finite Coxeter group is positive definite; cf. Exercise 26. Furthermore, the representation ρ is faithful (Exercise 35). Hence, every finite Coxeter group can be viewed naturally as a reflection group. Using properties of topological groups and introducing "fundamental domains," Humphreys [**Hum3**] closes the circle by showing that any Coxeter group having a positive definite bilinear form is a discrete subgroup of the orthogonal group (which is compact), and thus is finite.

It remains to describe the Coxeter systems in terms of Coxeter graphs. For which Coxeter graphs is the Coxeter form positive definite? By Exercise 28, the Coxeter graph must be a tree. One can exclude certain configurations; cf. Exercises 29 and 30. When the smoke has cleared, the only remaining connected Coxeter graphs that have multiple edges are B_n and F_4 (now labelled by 4 instead of 2), the Coxeter graphs •$\overset{n}{—}$•(denoted $I_2(n)$) of the dihedral group of order n for each $n > 3$, and two new exceptional graphs:

H_3 : •$\overset{5}{—}$•—• $\qquad$ H_4 : •$\overset{5}{—}$•—•—• ,

which have interesting interpretations as noncrystallographic finite reflection groups. This leaves us to determine the Coxeter graphs without multiple edges, which, as we are about to see, are A_n, D_n, E_6, E_7, and E_8.

A categorical interpretation of abstract Coxeter graphs

We complete the list of Coxeter graphs by means of a lovely formalism utilizing quadratic forms.

Definition 22.23. An **abstract Coxeter graph** Γ is a finite, undirected, loopless graph with multiplicity ≥ 1 assigned to each edge.

Remark 22.24. Any abstract Coxeter graph has a natural associated symmetric bilinear form. Namely, writing the vertices as $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, we define m'_{ij} to be the multiplicity of the edge connecting $\mathbf{e}_i$ and $\mathbf{e}_j$, formally putting $m'_{ij} = 0$ if there is no edge connecting $\mathbf{e}_i$ and $\mathbf{e}_j$. (Note that m'_{ij} is two less than the corresponding m_{ij} of Definition 22.17.) Then we define

the quadratic form

$$Q(x_1, \ldots, x_n) = \sum x_i^2 - \sum_{i \neq j} m'_{ij} x_i x_j. \tag{22.7}$$

The symmetric bilinear form of the quadratic form Q is by far the most straightforward of the three bilinear forms that we have encountered in this chapter.

Proposition 22.25. *The only abstract Coxeter graphs for which the quadratic form Q is positive definite are A_n, D_n, E_6, E_7, and E_8.*

Proof. As usual, the idea is to limit the graphs by excluding isotropic vectors. First of all, there are no labelled edges, for otherwise taking $m'_{i,j} \geq 2$ and $v = \mathbf{e}_i + \mathbf{e}_j$, we have $Q(v) = 1+1-m'_{ij} \leq 0$. There cannot be any cycles, for then $Q(v) \leq 0$ where v is 1 on each vertex of the cycle and 0 elsewhere. No vertex can have degree ≥ 4, since otherwise taking coefficient 2 for that vertex and 1 for four of its neighbors would yield 0. Likewise, there cannot be two branch points, since taking coefficient 1 for the extreme vertices and 2 elsewhere would yield 0.

Thus, the Coxeter graph must be a tree with at most one branch point. Take $p \leq q \leq r$ to be the respective lengths of the branches of the graph. Define the new quadratic form

$$\begin{aligned} &Q_m(x_0, x_1, \ldots, x_t) \\ &\quad = -x_0x_1 - x_1x_2 - x_2x_3 - \cdots - x_{t-1}x_t + x_1^2 + \cdots + x_t^2 + \frac{t}{2(t+1)}x_0^2 \end{aligned}$$

(where $t = p, q, r$, respectively). Then Q_m can be rewritten as

$$\sum_{i=0}^{t-1} \frac{t-i+1}{2(t-i)} \left(x_{i+1} - \frac{t-i}{t-i+1} x_i \right)^2,$$

and thus is positive nondefinite. Moreover, relabelling the x_i so that x_0 is the coefficient of the branch vertex, we proceed up the jth branch with $x_{j1}, x_{j2}, \ldots$, and rewrite

$$\begin{aligned} Q &= Q_p(x_0, x_{11}, \ldots, x_{1p}) + Q_q(x_0, x_{21}, \ldots, x_{2q}) \\ &\quad + Q_r(x_0, x_{31}, \ldots, x_{3r}) + \left(1 - \frac{p}{2(p+1)} - \frac{q}{2(q+1)} - \frac{r}{2(r+1)} \right) x_0^2. \end{aligned}$$

Thus, the quadratic form Q is positive definite iff

$$\frac{p}{2(p+1)} + \frac{q}{2(q+1)} + \frac{r}{2(r+1)} < 1.$$

The only solutions (for natural numbers $p \le q \le r$) are for $p = 0$ with $q \le r$ arbitrary (yielding the Coxeter graph A_n), or $p = 1$ and $q = 1$ with r arbitrary ≥ 1 (yielding the Coxeter graph D_n), or $p = 1$, $q = 2$, and $r \in \{2, 3, 4\}$ (yielding E_6, E_7, and E_8). □

For Coxeter graphs without multiple edges, this bilinear form is the same as the one we obtained in Definition 22.20 for the same graph with unlabelled edges. Indeed, $\cos\frac{\pi}{3} = \frac{1}{2}$, and thus, for any edge connecting i, j, we have

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = -\frac{1}{2} = \frac{1-1-1}{2} = \frac{Q(\mathbf{e}_i + \mathbf{e}_j) - Q(\mathbf{e}_i) - Q(\mathbf{e}_j)}{2},$$

which matches Equation (0A.2) of Volume 1. Thus we have a more conceptual alternative for the last part of the proof of Theorem 22.11. Together with the discussion following Theorem 22.22, this also completes the classification of finite Coxeter groups, as those groups having Coxeter graphs A_n, B_n, D_n, E_6, E_7, E_8, F_4, H_3, H_4, or $I_2(n)$.

An application to representation theory.

Bernstein, Gel′fand and Ponomarev [**BernGP**] took this reasoning one step further to interpret graph-theoretically the property that the quadratic form Q is positive definite. Consider any finite (directed) graph Γ, denoting its set of vertices as $\Gamma_0 = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and its set of edges as Γ_1. We also define functions $\nu_0, \nu_1 \colon \Gamma_1 \to \{1, \dots, n\}$, where for any edge ℓ from $\mathbf{e}_i$ to $\mathbf{e}_j$, $\nu_0(\ell) = i$ and $\nu_1(\ell) = j$. We write $(\Gamma; \nu)$ for this oriented graph.

Definition 22.26. The **category $\mathcal{C}(\Gamma; \nu)$ of representations of** $(\Gamma; \nu)$ is defined as follows: An object (V, f) of $\mathcal{C}(\Gamma; \nu)$ is a set of f.d. vector spaces $\{V_i : 1 \le i \le n\}$ (over $\mathbb{R}$) together with linear transformations denoted as $f_\ell \colon V_{\nu_0(\ell)} \to V_{\nu_1(\ell)}$ for each $\ell \in \Gamma_1$. A morphism $\Phi \colon (V, f) \to (W, g)$ is a collection of linear transformations $\phi_\gamma \colon V_\gamma \to W_\gamma$ such that for each edge ℓ the following diagram is commutative:

$$\begin{array}{ccc} V_{\nu_0(\ell)} & \xrightarrow{f_\ell} & V_{\nu_1(\ell)} \\ {\scriptstyle \phi_{\nu_0(\ell)}}\downarrow & & \downarrow{\scriptstyle \phi_{\nu_1(\ell)}} \\ W_{\nu_0(\ell)} & \xrightarrow{g_\ell} & W_{\nu_1(\ell)}. \end{array}$$

We call any such (V, f) a **representation** of the oriented graph $(\Gamma; \nu)$. We always assume that Γ is loopless. Define the **degree** of a representation (V, f) to be the n-tuple $(\dim V_i)$. Likewise, we list the edges as $\Gamma_1 = \{\ell_1, \dots, \ell_t\}$ and write $\nu_u(k)$ for $\nu_u(\ell_k)$ for $u = 1, 2$, and $1 \le k \le t$.

Remark 22.27. Definition 22.26 could be understood in the language of category theory. Let Γ be a graph, viewed as a small category whose objects are its vertices and whose morphisms are the edges, i.e., $\mathrm{Hom}(\mathbf{e}_i, \mathbf{e}_j)$ is just the set of edges from $\mathbf{e}_i$ to $\mathbf{e}_j$. Let $\mathcal{D}$ be the category of f.d. vector spaces whose morphisms are the linear transformations. A functor F from Γ to $\mathcal{D}$ then assigns a f.d. vector space to each vertex and a linear transformation to each edge; in other words, F is a representation of $(\Gamma; \nu)$. In the terminology of Definition 1A.17 of Volume 1, $\mathcal{C}(\Gamma; \nu)$ is the category $\mathbf{Fun}(\Gamma, \mathcal{D})$.

We define the **direct sum** of representations (U, f) and (V, g) to be the representation (W, h), where $W_i = U_i \oplus V_i$ and $h_i = f_i \oplus g_i$. A representation of $(\Gamma; \nu)$ is called **indecomposable** if it cannot be written as a direct sum of nonzero representations. Any representation is a finite direct sum of indecomposable representations, so this leads us to try to classify the indecomposable representations. Gabriel showed that the property of having finitely many indecomposable representations is equivalent to Γ being one of the Dynkin diagrams A_n, D_n, E_6, E_7, and E_8. We give half of the theorem here, following Tits' proof given in Bernstein, Gel′fand, and Ponomarev [**BernGP**].

THEOREM 22.28. *If an abstract Coxeter graph $(\Gamma; \nu)$ has only finitely many nonisomorphic indecomposable representations, then its quadratic form Q (Remark 22.24) is positive definite.*

Proof (Tits). Consider an arbitrary representation (V, f) of degree (d_i) with $d_i = \dim V_i$. We view f as an n-tuple of matrices $(A_1, \dots, A_n)$ (where $n = |\Gamma_1|$); here A_j corresponds to the transformation $f_j \colon V_{\nu_0(j)} \to V_{\nu_1(j)}$. Let $\mathcal{A}$ be the variety comprised of n-tuples of $d_{\nu_0(j)} \times d_{\nu_1(j)}$ matrices, $1 \le j \le n$. Thus,

$$\dim \mathcal{A} = \sum_{k=1}^{n} d_{\nu_0(j)} d_{\nu_1(j)} = \sum_{i,j=1}^{n} m_{ij} d_i d_j, \tag{22.8}$$

since we count $d_i d_j$ once for each edge from i to j. Now consider the algebraic group $G = \mathrm{GL}(V_1) \times \cdots \times \mathrm{GL}(V_n)$, which has dimension $\sum_{j=1}^{n} d_j^2$. Clearly, conjugation by G acts as change of base for $V_1 \times \cdots \times V_n$ and also acts on the A_j accordingly. Explicitly, writing an element of G as $(g_1, \dots, g_n)$, A_j is transformed to $g_{\nu_0(j)}^{-1} A_j g_{\nu_1(j)}$. In this way, $\mathcal{A}$ decomposes into orbits of G, where isomorphic representations belong to the same orbit. Hence, by hypothesis, $\mathcal{A}$ decomposes into only finitely many orbits; i.e., there are only finitely many isomorphism classes of representations.

Now identify $\mathbb{R}^\times$ with $\{(\alpha, \dots, \alpha) : \alpha \in \mathbb{R}^\times\} \subset G$. The action of $\mathbb{R}^\times$ on $\mathcal{A}$ is trivial, for each A_j goes to $\alpha^{-1} A_j \alpha = A_j$. Thus, $\mathcal{A}$ breaks up into a

finite number of orbits of $G/\mathbb{R}^\times$, implying that

$$\sum_{i,j=1}^{n} m_{ij} d_i d_j = \dim \mathcal{A} < \dim G = \sum_{j=1}^{n} d_j^2. \tag{22.9}$$

Since (22.9) holds for any (d_i), the quadratic form (22.7) is positive definite on $\mathbb{Q}$, and thus on all of $\mathbb{R}$. □

Bernstein, Gel′fand and Ponomarev [**BernGP**] went on to prove the converse also, constructing examples of indecomposable representations for each of these diagrams. We return to this application in Appendix 25C.

Exercises – Part V

Chapter 19

1. Show that the only degree 1 representations of S_n are **1** and **sgn**. (Hint: The only nontrivial normal subgroup is A_n, except for $n = 4$.)
2. $C_3 = \langle a \rangle$ has the faithful representation $a \mapsto \left(\begin{smallmatrix} 0 & 1 \\ -1 & -1 \end{smallmatrix}\right)$ of degree 2. Show that this representation is reducible, iff the field F contains a primitive cube root of 1.
3. Prove that a group G has a unique degree 1 complex representation (namely, **1**) iff $G' = G$.
4. The **contragredient representation** ρ^* of a representation ρ of degree n is defined by taking $\rho^*(g)$ to be the transpose of the matrix $\rho(g^{-1})$. Prove that ρ^* also is a representation of degree n; ρ^* is irreducible iff ρ is irreducible.
5. Suppose M is the G-space corresponding to a representation ρ. Show that the dual space $M^* = \mathrm{Hom}_F(M, F)$ is also a G-space, where gf is taken to be the map $v \mapsto f(g^{-1}v)$; show when $\deg \rho < \infty$ that M^* corresponds to the contragredient representation ρ^*. (Hint: Take the dual base $\hat{e}_1, \ldots, \hat{e}_n$ of M^*. Taking $ge_i = \sum \alpha_{ij} e_j$, show that $g^{-1}\hat{e}_i(e_j) = \alpha_{ji}$.)
6. (Schur's Lemma, representation-theoretic formulation.) Assume that F is a splitting field for G. Letting L_i denote the module corresponding to ρ_i, show that $\mathrm{End}_{F[G]}(L_i) \cong F$ and $\mathrm{Hom}_{F[G]}(L_i, L_j) = 0$ for all $i \neq j$. (Hint: The first assertion is as in Proposition 14.28. The second assertion is Proposition 13.34.)
7. An irreducible representation $\rho : G \to \mathrm{GL}(n, F)$ is called **absolutely irreducible** if it remains irreducible over any field extension of F. Prove that every irreducible representation over a splitting field for G is absolutely irreducible.
8. Although the group algebra $R = F[C_2 \times C_2]$ is commutative and $[R : F] = 4$, show for any infinite field F of characteristic 2 that R has infinitely many nonisomorphic indecomposable modules. (Hint: Write the group as $\langle g, h : g^2 = h^2 = 1 \rangle$ and let $a = g + 1$ and $b = h + 1$. Then a and b are nilpotent and $Fab = \mathrm{soc}(R)$. Let $M_\alpha = F(\alpha a + b) + Fab \triangleleft R$. Each M_α is indecomposable, and the M_α are nonisomorphic.)
9. For $\mathrm{char}(F) = p > 0$ and any element $g \in G$ of order $m = p^i q$ with q prime to p, show that $g = g_1 g_2$ where g_1, g_2 are powers of g such that, for any representation $\rho : G \to \mathrm{GL}(n, F)$, $\rho(g_1)$ is separable (with eigenvalues that are q-th roots of 1) and $\rho(g_2)$ is unipotent. (Hint: Remark 0.0 of Volume 1.)

The invariant bilinear form of a representation

A bilinear form $\langle\ ,\ \rangle$ on a G-space V is called G-**invariant** if $\langle gv, gw \rangle = \langle v, w \rangle$, $\forall g \in G$, $\forall v, w \in V$.

10. Suppose $\rho : G \to \mathrm{GL}(n, \mathbb{R})$ is a real representation of a finite group G. Show that $V = \mathbb{R}^{(n)}$ has a positive definite G-invariant bilinear form with respect to which each element of G is unitary. (Hint: Apply the averaging process to the usual positive definite bilinear form; i.e., take $\sum_{g \in G} \langle \rho(g)v, \rho(g)w \rangle$.)
11. Suppose $\rho : G \to \mathrm{GL}(n, \mathbb{C})$ is a complex representation of a finite group G. Show that $V = \mathbb{C}^{(n)}$ has a Hermitian G-invariant bilinear form with respect to which each element of G is unitary. (Hint: Apply the averaging process to the usual Hermitian form.)
12. Show that the forms in Exercises 10 and 11 are unique up to scalar multiple.
13. Show that a representation ρ of finite degree is completely reducible whenever the corresponding G-space V has a G-invariant Hermitian form. (Hint: If V has a G-subspace W, then $V = W \oplus W^\perp$.) Use this and the preceding exercises to reprove Maschke's Theorem when the base field is $\mathbb{C}$ or $\mathbb{R}$.
14. Show that any real representations of finite degree that have equivalent complexifications are equivalent. (Hint: One needs $A \in \mathrm{GL}(n, \mathbb{R})$ satisfying Equation (19.3). Some $A + Bi \in \mathrm{GL}(n, \mathbb{C})$ satisfies (19.3), and so does $A + \alpha B$ for each $\alpha \in \mathbb{R}$; choose α such that $A + \alpha B$ is invertible.)
15. Whenever a group G acts on a set X, show that the free module $\mathbb{Z}X$ is a G-module under the natural action $g(\sum m_x x) = \sum m_x (gx)$; this is called a **permutation module**. If X is a disjoint union of sets X_i

invariant under the G-action, then $\mathbb{Z}X = \oplus \mathbb{Z}X_i$ as G-modules. Taking a set $\{x_i : i \in I\}$ of representatives of the partition of X under G, conclude that $\mathbb{Z}X \cong \bigoplus_{i\in I} \mathbb{Z}[G/G_{x_i}]$. Show that permutation modules correspond to permutation representations.

The structure of group algebras

16. What is the kernel of the homomorphism $F[G] \to F[G/G']$ when G is finite and F is a splitting field of G? (Hint: Simple components must be preserved or go to 0.)
17. Determine the ring-theoretic structure of the group algebra $\mathbb{C}[G]$ for each group G of order ≤ 12.
18. Compute all the idempotents of $\mathbb{C}[C_n]$.
19. Show that $\mathbb{Q}[C_4] \cong \mathbb{Q}[i]\times\mathbb{Q}\times\mathbb{Q}$, whereas $\mathbb{Q}[C_2\times C_2] \cong \mathbb{Q}\times\mathbb{Q}\times\mathbb{Q}\times\mathbb{Q}$. (Hint: Factor $\lambda^4 - 1$.) More generally, show that $\mathbb{Q}[A_1] \cong \mathbb{Q}[A_2]$ for finite *Abelian* groups A_1 and A_2 iff A_1 and A_2 are isomorphic.
20. Prove that a field F is a splitting field for a finite Abelian group of exponent m iff F contains a primitive m-th root of 1.
21. When $|F| = 7$, show that $F[C_4]$ is a direct product of three fields, and thus is split, but F is not a splitting field of C_4 in the sense of Definition 19.30. What finite field extension of F is a splitting field of C_4?
22. Given any finite group G, define $z = \sum_{g\in G} g \in \mathrm{Cent}(F[G])$. Show that $gz = z$ for each $g \in G$. Conclude that $z^2 = |G|z$.
23. Using Exercise 22, whenever $\mathrm{char}\, F$ divides $|G|$, show that z is a central nilpotent element, and thus $F[G]$ is not semiprime.
24. The **group algebra trace map** $\mathrm{tr}: F[G] \to F$ is defined by $\mathrm{tr}(\sum \alpha_g g) = \alpha_1$. Show that $\mathrm{tr}(ab) = \mathrm{tr}(ba)$, and thus tr gives rise to the symmetric bilinear form $\langle a, b\rangle = \mathrm{tr}(ab)$. Furthermore, this form is **associative** in the sense that $\langle ab, c\rangle = \langle a, bc\rangle$. This characteristic-free property of group algebras is significant — an associative algebra with a symmetric associative bilinear form is called a **symmetric algebra**. An algebra R satisfying $R \cong \mathrm{Hom}(R, F)$ as R-modules is called **Frobenius**; show that any f.d. symmetric algebra is Frobenius. There is an extensive theory for Frobenius algebras.

The augmentation ideal

25. Show that the augmentation ideal of $F[G]$ equals $\sum_{g\in G} F[G](g - 1)$.
26. Show that the sum in Exercise 25 can be taken over any set of generators of G. (Hint: $g_1g_2 - 1 = g_1(g_2-1) + g_1-1$.)
27. For $\mathrm{char}\, F = p$ and $|G| = p^k$, show that the augmentation ideal is nilpotent; conclude that $F[G]$ is a local ring. (Hint: First do this for G Abelian.)

Involutions of the group algebra, cf. Definition 14.32

28. Show that $C[G]$ has an involution given by $\left(\sum_g c_g g\right)^* = \sum_g c_g g^{-1}$.
29. Suppose F is a subfield of $\mathbb{C}$ that is closed under complex conjugation. Show for any group G that the group algebra $F[G]$ has an involution given by $\left(\sum_g \alpha_g g\right)^* = \sum_g \bar{\alpha}_g g^{-1}$.

Generalizations of Maschke's Theorem for infinite groups

30. Under the hypothesis of Exercise 29, show that $F[G]$ has no nonzero nil left or right ideals. (Hint: Compute

$$\mathrm{tr}\left(\left(\sum_g \alpha_g g\right)\left(\sum_g \alpha_g g\right)^*\right)$$

and apply Exercise 15.17.)
31. Prove that $C[G]$ is semiprime, for any group G and any integral domain C of characteristic 0. (Hint: Take an algebraic closure $\bar{F}$ of C and write $\bar{F} = K[i]$ for a real-closed field K; cf. Exercise 4C.11 of Volume 1. As in Exercise 30, $\bar{F}[G]$ is semiprime, so the same is true of $C[G]$.)
32. Reprove Maschke's Theorem over any field F of characteristic 0, using Exercise 31.
33. (Amitsur) Prove that $\mathrm{Jac}(F[G])$ is a nil ideal, for any uncountable field F. (Hint: Otherwise take non-nilpotent $a = \sum \alpha_g g \in \mathrm{Jac}(F[G])$. Let H be the subgroup generated by those g in $\mathrm{supp}(a)$. Then $a \in \mathrm{Jac}(F[H])$, but Exercise 15A.10 shows that $\mathrm{Jac}(F[H])$ is a nil ideal of $F[H]$.)
34. (Herstein; Amitsur) Prove that $\mathrm{Jac}(F[G]) = 0$ for any uncountable field F of characteristic 0. (Hint: The argument of Exercise 31 shows that $F[G]$ has no nonzero nil ideals, so apply Exercise 33.)

Skew group algebras

35. Given a group of automorphisms G on an algebra R, define the **skew group algebra** $R\#G$ to have the same vector space structure as the group algebra $R[G]$, but with multiplication "skewed" by the action of G; i.e.,

$$(a_\sigma\sigma)(b_\tau\tau) = a_\sigma\sigma(b_\tau)\sigma\tau.$$

Verify that the skew group algebra is an associative algebra, but in opposition to the unskewed case, find an example when $F\#G$ is simple for F a field of characteristic 0 and $|G|$ finite.

The skew group algebra construction is useful in studying group actions and is put in a much more general context in Exercises 26.34ff.

Young diagrams

36. What is the maximal degree of an irreducible representation of S_n?
37. Define an equivalence class on tableaux by saying that $T \sim T'$ if there is a row permutation π such that $\pi T = T'$; in other words, for each i, the entries of the i row of T are the same (up to permutation) as the entries of the i row of T'. Write $[T]$ for the equivalence class of T. Let V denote the vector space over F whose base is the set of equivalence classes of tableaux. V is an S_n-space. Prove the following variant of Lemma 19.54: If $H \leq S_n$ has a transposition $\tau = (\ell\, \ell')$ such that ℓ and ℓ' lie in the same row of T, then $s_H[T] = 0$.
38. Given a descending partition $\lambda = (m_1, m_2, \dots)$ of n, take the smallest possible tableau T arising from λ; i.e., the first row is $(1\ 2\ \dots\ m_1)$ for some m_1, the second row is $(m_1{+}1\ m_1{+}2\ \dots\ m_2)$, and so forth. Define $S^\lambda \subset S_n$ to be the direct product of the permutation groups on the rows of T_λ. Show that the action of S_n induces a 1:1 correspondence between the cosets of S^λ and the equivalence classes of tableaux arising from λ.

The Double Centralizer Theorem for group representations

39. Suppose $\rho\colon G \to \mathrm{GL}(n, F)$ is a completely reducible representation. Let $R = M_n(F)$. Taking the corresponding algebra representation $\hat{\rho}\colon F[G] \to R$, let $A = \hat{\rho}(F[G]) \subseteq R$. Prove that $C_R(C_R(A)) = A$. (Hint: Write $M_n(F) = \mathrm{End}_F V$ and note that V is a semisimple A-module; the assertion follows from Jacobson's Density Theorem.) Compare with Theorem 24.32(iii).
40. Suppose A is a f.d. semisimple algebra over an algebraically closed field F of characteristic 0; i.e., $A = \prod_{i=1}^t A_i$, where $A_i \cong M_{n_i}(F)$. Let M be an A-module. Viewing $A \subseteq \mathrm{End}_F M$, let $B = \mathrm{End}_A M$. Using results from Chapter 15, prove the following:

    (i) (Isotypic decomposition) Let J_i be a minimal left ideal of A_i; thus, $[J_i : F] = n_i$ and $\mathrm{End}_{A_i} J_i = F$. Let $M_i = A_i M \cong J_i^{(m_i)}$ for suitable m_i. Then $M = \oplus M_i$ as A-modules.

    (ii) $B = \prod_{i=1}^t B_i$, where $B_i = \mathrm{End}_{A_i} M_i \cong M_{m_i}(F)$.

    (iii) $M_i \cong L_i^{(n_i)}$ as B-modules, where L_i is a minimal left ideal of B_i, so $[L_i : F] = m_i$.
41. In Exercise 40, show that A_i and B_i are the centralizers of each other in $\mathrm{End}_F M_i$.
42. (Schur's Double Centralizer Theorem.) Suppose V is any f.d. vector space over a field of characteristic 0, and define two types of actions on $V^{\otimes n} = V \otimes \cdots \otimes V$:

    (i) The diagonal action of $\mathrm{GL}(V)$, given by $T(v_1 \otimes \cdots \otimes v_n) = T(v_1) \otimes \cdots \otimes T(v_n)$.

    (ii) The action of the symmetric group S_n, by $\sigma(v_1 \otimes \cdots \otimes v_n) = v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(n)}$.

    Show that these two actions centralize each other, and provide respective representations $\rho\colon \mathrm{GL}(V) \to \mathrm{GL}(V^{\otimes n})$ and $\tau\colon S_n \to \mathrm{GL}(V^{\otimes n})$, and thus algebra homomorphisms $\hat{\rho}\colon F[\mathrm{GL}(V)] \to \mathrm{End}_F V^{\otimes n}$ and $\hat{\tau}\colon F[S_n] \to \mathrm{End}_F V^{\otimes n}$. Let A_1, A_2 be the respective images of $\hat{\rho}, \hat{\tau}$. Prove that A_1 and A_2 are the centralizers of each other in $\mathrm{End}_F V^{\otimes n}$.

Appendix 19A

1. If $V = F^{(n)}$ is semisimple as a G-space, prove that V is also semisimple as an H-space whenever $[G : H] = m$ is prime to $\mathrm{char}(F)$. (Hint: Use the averaging procedure of Lemma 19.25.)

Representations into $\mathrm{GL}(n, F)$

2. (Decomposing $M_n(F)$ with respect to Ad; cf. Example 19A.5′.) Let $s\ell(n, F)$ denote the matrices of trace 0, clearly invariant under conjugates and thus a $\mathrm{GL}(n, F)$-subspace under the action of Ad. In characteristic 0, show that $M_n(F) = F \oplus s\ell(n, F)$ as $\mathrm{GL}(n, F)$-subspaces are each simple; thus Ad is a completely reducible representation. (Hint: It is enough to show that $s\ell(n, F)$ is a simple $\mathrm{GL}(n, F)$-space; i.e., if $\mathrm{tr}(a) = 0$ for $a \neq 0$, then the $\mathrm{GL}(n, F)$-subspace M generated by a is all of $s\ell(n, F)$. To see this, note for $E_{ij} = 1 + e_{ij}$ that $E_{ij} a E_{ij}^{-1} - a$ has nonzero entries only in the i row or j column.)
3. (The **representation in bilinear forms** on $M_n(F)$.) Here, one views a matrix $a = (\alpha_{ij})$ as a bilinear form $\langle v, w\rangle = v^t a w$, cf. Equation (0.5) of Volume 1, and $\rho(g)$ is the bilinear form

$$\langle v, w\rangle = (g^{-1}v)^t\, a g^{-1} w;$$

    i.e., we apply g^{-1} to both sides before taking the inner product. Show that $M_n(F)$ has two natural invariant simple $\mathrm{GL}(n, F)$-subspaces $M_n(F)^+$ and $M_n(F)^-$, respectively the subspaces of **symmetric** and **skew-symmetric** matrices, which are irreducible under this action.
4. Embed $\mathrm{PGL}(n, F)$ into $\mathrm{GL}(n^2, F)$. (Hint: $\mathrm{PGL}(n, F)$ acts faithfully on $\mathrm{GL}(n, F)$ via the adjoint representation of Example 19A.5′; if $gag^{-1} = a$ for all $a \in M_n(F)$, then $g \in F^\times$.)
5. (Mal′cev) Show that a group has a faithful representation of degree n iff each of its finitely generated subgroups has a faithful representation of degree n. (This is really an exercise about ultraproducts; cf. Exercise 0.17 of Volume 1.)

Burnside's Problem for linear groups

6. Prove for any field F that any f.g. periodic subgroup G of $GL(n,F)$ is finite. (Hint: For $\mathrm{char}(F)=0$ one could use Theorem 19A.9(ii), so assume that the characteristic subfield F_0 is finite.

   Taking a composition series of G-subspaces of $F^{(n)}$, one may assume that $F^{(n)}$ is a simple G-space; hence G contains an F-base $B=\{b_1=1,\dots,b_{n^2}\}$ of $M_{n^2}(F)$. Expand $\{b_ub_v : 1\le u,v\le n^2\}$ to a set $\{g_1,\dots,g_m\}$ that generates G as a monoid. By assumption, the eigenvalues of each g_i are roots of 1. Write $g_i=\sum_{j=1}^{n^2}\gamma_{ij}b_j$ for $\gamma_{ij}\in F$. As in the proof of Theorem 19A.9(ii), one can solve for the γ_{ij} in terms of $\mathrm{tr}(g_ib_k)$, $1\le k\le n^2$, and $\delta=\mathrm{disc}\,B$. Thus, each γ_{ik} and δ are generated by roots of 1, so $F_1=F_0(\delta,\mathrm{tr}(g_ib_k):1\le i\le m,\ 1\le k\le n^2)$ is a finite algebraic extension of F_0, and thus is a finite field. Let $R_1=\sum F_1b_j$, a finite ring that contains all g_i, and thus contains G.)
7. Using Remark 15B.2, write down a periodic linear infinite group of bounded period.
8. (Schur) Prove that each periodic subgroup G of $\mathrm{GL}(n,\mathbb{C})$ consists of unitary matrices with respect to some positive definite Hermitian form. (Hint: In view of Exercise A1, one may assume that G is irreducible. Any set of n^2 $\mathbb{C}$-linearly independent elements of G generates an irreducible subgroup H; then H is thus finite and has a Hermitian form as in Exercise 19.11. By Exercise 19.12, the Hermitian form for the group generated by H and arbitrary $g\in G$ is the same.)

Unitary matrices and Jordan's theorem

The following exercises sketch the development in [**CuR**].

9. Show that any commuting unitary matrices can be simultaneously diagonalized by a unitary transformation. (Hint: Diagonalize the bilinear form.)
10. For a matrix $r=(r_{ij})\in M_n(\mathbb{C})$, define a topological norm $||r||=\mathrm{tr}(r\,\bar{r}^t)=\sum|r_{ij}|^2$. Suppose $a,b\in\mathrm{GL}(n,\mathbb{C})$ are unitary matrices, and denote their group commutator as (a,b). Note that $||a||=n$ and $||ar||=||ra||=||r||$ for every $r\in M_n(\mathbb{C})$. Prove the following assertions:

    (i) If a and (a,b) commute, and if $||1-b||<4$, then $ab=ba$. (Hint: a and bab^{-1} commute, so assume that they are diagonal. Hence, $bab^{-1}=\pi a\pi^{-1}$ for a suitable permutation matrix π, so $\pi^{-1}b$ commutes with a. If $\pi a\pi^{-1}\ne a$, compute $||1-b||=||\pi^{-1}-\pi^{-1}b||\ge 4$, a contradiction.)

    (ii) $||1-(a,b)||\le 2||1-a||\,||1-b||$. (Hint: Assume a is diagonal.)

    (iii) If the group H generated by a and b is finite and $||1-a||<\frac12$ and $||1-b||<4$, then $ab=ba$. (Hint: Define inductively $b_0=b$ and $b_i=(a,b_{i-1})$. By (ii), the determinants decrease. Since H is finite, some $b_i=1$; now apply (i) with reverse induction.)
11. (Jordan) Prove that if $G\subseteq\mathrm{GL}(n,\mathbb{C})$ is unitary, then G has a normal Abelian subgroup of index bounded by $(\sqrt{8n}+1)^{2n^2}-(\sqrt{8n}-1)^{2n^2}$. (Hint: By Exercise A8, assume that G is unitary. Let H be the subgroup generated by $\{a\in G : ||1-a||<\frac12\}$, a normal Abelian subgroup, by Exercise A10. Take a transversal $\{g_1,\dots,g_m\}$ of H in G. Since each $||g_i-g_j||\ge\frac12$, define a distance function between matrices a and b as $\sqrt{\|a-b\|}$. The spheres of radius $\frac{1}{\sqrt8}$ are non-overlapping. Since each sphere drawn around some g_i lies in the shell bounded from $\sqrt{n}-\frac{1}{\sqrt8}$ to $\sqrt{n}+\frac{1}{\sqrt8}$, comparing volumes yields

$$[G:H]\left(\frac{1}{\sqrt8}\right)^{2n^2}\le\left(\sqrt{n}+\frac{1}{\sqrt8}\right)^{2n^2}-\left(\sqrt{n}-\frac{1}{\sqrt8}\right)^{2n^2}.)$$

Topological groups

12. Recall that the **convex hull** of a compact set $S\subset\mathbb{R}^{(n)}$ is $\{\sum_{i=1}^m t_is_i : m\in\mathbb{N},\ s_i\in S,\ t_i\ge0,\ \sum_{i=1}^m t_i=1\}$. Show that the convex hull of a set equals the union of the convex hulls of its finite subsets.
13. For any $a\in\mathbb{R}^{(n)}$, and any convex set $S\subset\mathbb{R}^{(n)}$, show that the convex hull of $S\cup\{a\}$ is the union of all lines joining a to a point of S.
14. Prove that the convex hull of a set S equals the union of the convex hulls of all subsets $\{s_1,\dots,s_d : s_i\in S\}$, where d is the dimension of the smallest hyperplane of $\mathbb{R}^{(n)}$ containing S. (Hint: In view of Exercise A12, one may assume that S is finite; i.e., $S=\{s_1,\dots,s_m\}$. Apply induction on m, starting with Exercise A13.)
15. Show that the convex hull of a compact subset of $\mathbb{R}^{(n)}$ is compact.
16. (Generalizing Exercise 19.11.) For any continuous complex representation $\rho: G\to\mathrm{GL}(n,\mathbb{C})$ of a compact topological group G, prove that $\mathbb{C}^{(n)}$ has a positive definite G-invariant Hermitian form. (Hint: Apply the averaging process of Lemma 19A.15 to the usual Hermitian form $\langle\ ,\ \rangle$, i.e., taking $\int_{g\in G}\langle gv,\overline{gw}\rangle dg$.

    Here is another method, following Vinberg [**Vinb**], that does not rely on the Haar measure. Let P be the set of all positive definite Hermitian functions, which is an open convex subset of the space of all symmetric bilinear functions on $\mathbb{R}^{(n)}$. Take any compact subset S_0 of P having positive measure $\mu(S_0)$, where μ denotes the Euclidean measure, and let $S=\bigcup_{g\in G}\ \rho(g)S_0$, a compact subset of P since any sequence has a converging subsequence. The **center of mass** $\mu(S)^{-1}\int_{x\in S}x\mu(dx)$ lies in P since P is convex and is G-invariant.)

17. In Exercise A16, show that $\mathbb{R}^{(n)}$ has a positive definite G-invariant symmetric form when ρ is real.
18. Show that any continuous f.d. representation of G having a positive definite G-invariant Hermitian form is completely reducible. (Hint: As in Exercise 19.13.)
19. Show that the orthogonal group $\mathrm{O}(n,\mathbb{R})$ is not connected; in fact, the connected component of the identity is $\mathrm{SO}(n,\mathbb{R})$. (Hint: The inverse of an orthonormal matrix a is its transpose, so $\det a = \pm 1$.)

Lie groups

20. Verify that $\mathrm{SL}(n,\mathbb{C})$ and $\mathrm{SL}(n,\mathbb{R})$ are Lie groups. (Hint: Write the determinant det as a function of the entries $\{x_{ij} : 1 \le i,j \le n\}$ of a matrix. Then $\frac{\partial \det}{\partial x_{ii}}$ is the determinant of the adjoint of x_{ii}, which evaluated at the identity is 1.)
21. Display the orthogonal group $\mathrm{O}(n,\mathbb{R})$ as a Lie subgroup of $\mathrm{GL}(n,\mathbb{R})$ of dimension $\frac{1}{2}n(n-1)$. (Hint: The equations defining $aa^t = 1$ are $\frac{1}{2}n(n+1)$ in number, and their Jacobian has a minor of size $\frac{1}{2}n(n+1)$ which does not vanish when evaluated at the identity element.)

    $\mathrm{SO}(n,\mathbb{R})$, being the connected component of $\mathrm{O}(n,\mathbb{R})$, is also a Lie subgroup of the same dimension.
22. Viewing $\mathrm{GL}(n,\mathbb{C})$ as a $2n^2$-dimensional manifold over $\mathbb{R}$, show as in Exercise A21 that the unitary group is a Lie group of dimension n^2.
23. Define the product $[\xi_1\xi_2]$ for ξ_1,ξ_2 in D_eG by

$$[\xi_1\xi_2] = \frac{\partial^2}{\partial t_1 \partial t_2}(f_1(t_1), f_2(t_2))\,\big|_{t_1=t_2=0};$$

    here $f_i\colon \mathfrak{I} \to G$ (where $\mathfrak{I}$ denotes the unit interval $[0,1]$) with $f_i(0) = e$ and $f_i'(0) = \xi_i$. Prove that $\mathrm{ad}([ab]) = [\mathrm{ad}(a), \mathrm{ad}(b)]$, $\forall a,b \in D_eG$. Hence, this definition corresponds to the one given in the text.

Braid groups

The **braid group** B_n is generated by elements $\sigma_1,\dots,\sigma_{n-1}$, subject to the relations $\sigma_i\sigma_j = \sigma_j\sigma_i$ whenever $|i-j| > 1$, as well as the **braid relation** $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$. Bigelow and Krammer succeeded in showing that the braid group is linear. This result is out of range of these exercises, but we do present several interesting properties of the braid group. The above notation is used throughout Exercise A34.

24. Verify the group surjection $\psi\colon B_n \to S_n$ given by $\sigma_i \mapsto (i\ i+1)$. Its kernel $\ker\psi$ is denoted as P_n, the subgroup of **pure braids**.
25. (Another definition of the braid group.) Define a **path** to be a continuous function $f\colon [0,1] \to \mathbb{C}$ with respect to the Euclidean topology, such that $f(0), f(1) \in \mathbb{Z}$, and a **braided path** to be an n-tuple of nonintersecting paths $(f_1,\dots,f_n)$ such that $\{f_1(0),\dots,f_n(0)\} = \{1,2,\dots,n\} = \{f_1(1),\dots,f_n(1)\}$. (In other words, the ordered n-tuple $(f_1(1),\dots,f_n(1))$ is a permutation of $(f_1(0),\dots,f_n(0)$.) Paths are multiplied by juxtaposing and then renormalizing the domain of definition to $\mathcal{I}$. Two braids are called **equivalent** if one can be deformed continuously to the other. Identify the braid group with the equivalence classes of braided paths, where the crossovers correspond to the braid relations; under this identification, the pure braids satisfy $f_i(1) = f_i(0)$ for all i.

26. Display the pure braid group P_n as the fundamental group of $\mathbb{C}^{(n)}\backslash\Delta$, where $\Delta = \{(z_1,\dots,z_n) : z_i = z_j \text{ for some } i<j\}$.
27. Find a natural injection $B_{n-1} \to B_n$ under which P_{n-1} can be viewed as a subgroup of P_n.
28. Verify the injection of B_n into the automorphism group of the free group of rank n on generators $x_1,\dots,x_n$, given by identifying σ_i with the automorphism sending $x_i \mapsto x_ix_{i+1}x_i^{-1}$, $x_{i+1} \mapsto x_i$, and fixing all other x_j.
29. Let $\theta_i = \sigma_1\cdots\sigma_{i-1}$, and $a_i = \theta_i\sigma_i^2\theta_i^{-1}$. Show that $a_1,\dots,a_{n-1}$ generate a subgroup A of B_n which is free, seen by applying the injection of Exercise A28.
30. (The Artin combing procedure.) Define a map $f\colon P_n \to P_{n-1}$ obtained by cutting the n-th strand. Show that $\ker f$ is then the fundamental group of the complement of $n-1$ points, and thus is the free group of rank $n-1$; cf. Remark 17A.4. (This is also seen algebraically, using Exercise A29.) If b is a braid, then $f(b)$ can be viewed again in P_n, implying $f(b)b^{-1} \in \ker f$. Thus, P_n is embedded in the semidirect product of P_{n-1} and $\ker f$, given by $b \mapsto (f(b), f(b)b^{-1})$. Continuing this process, embed P_n into an iterated semidirect product of free groups of ranks $n-1,\ \dots,\ 1$.
31. Use the Artin combing procedure to obtain an algorithmic solution to the word problem in P_n, and thus in B_n since P_n is of finite index in B_n.
32. Let $\theta = \theta_{n-1}$ of Exercise A29. Show that $\theta\sigma_i = \sigma_{i+1}\theta$; conclude that θ and any σ_i generate B_n. For $n \ge 3$, show that the center of B_n is infinite cyclic, generated by θ^n. (Hint: Induction and Exercise A29.)
33. Show that $B_n/B_n' \cong \mathbb{Z}$.
34. Prove that $B_n' = (B_n', B_n)$. (Hint: It suffices to write the group commutator (σ_i,σ_{i+1}) in (B_n', B_n). The braid relation implies that $\sigma_i\sigma_{i+1}^{-1} \in B_n'$. But $(\sigma_i,\sigma_{i+1}) = (\sigma_i\sigma_{i+1}^{-1},\sigma_{i+1})$.)

Appendix 19B

1. Prove that every algebraic group is nonsingular as a variety. (Hint: By Exercise 10A.7 of Volume 1, there exists a nonsingular point, which can be translated anywhere.)
2. Show that the linear algebraic groups $\mathrm{SL}(n,F)$, $\mathrm{D}(n,F)$, $\mathrm{T}(n,F)$, $\mathrm{SO}(n,F)$, and $\mathrm{U}(n,F)$ are connected.
3. Suppose $(R,*)$ is any ring with involution, and $a^* = -a \in R$ with $1-a$ invertible. Show that the element $b = \frac{1+a}{1-a}$ satisfies $b^* = b^{-1}$. Conversely, if $b^* = b^{-1}$ and $b \neq -1$, then $a = \frac{b-1}{b+1}$ is skew-symmetric.
4. Show that $\mathrm{SO}(n,F)$ is $\mathrm{O}(n,F)_e$. (Compare with Exercise 19A.19; the usual topology on $\mathbb{R}$ is finer than the Zariski topology.)

The Tits alternative

5. For any infinite algebraic extension F of a finite field, show that $\mathrm{GL}(n,F)$ is a group that is not virtually solvable but does not contain any free group.
6. Verify Step V of the proof of Theorem 19B.21. (Hint: Passing to $\rho(G)$, assume that $G \subseteq \mathrm{GL}(n,F)$. Assume that F is algebraically closed. Using the Jordan decomposition, assume that a is in upper triangular form with its eigenvalues $\alpha_1, \dots, \alpha_n$ on the diagonal, and $|\alpha_1| \geq |\alpha_2| \geq \cdots \geq |\alpha_n|$. By assumption, $|\alpha_1| > |\alpha_i|$ for each $i > 1$, so other than the 1,1 entry of a, all the entries of the first row and first column of a may be taken to be 0.

   Let u be the largest size of a block in the Jordan form of a belonging to an eigenvalue $\alpha = \alpha_j$ such that $|\alpha| = |\alpha_n|$ (the minimal absolute value of an eigenvalue). Explicitly, the block has the form

$$B = \begin{pmatrix} \alpha & 1 & 0 & \dots & 0 & 0 \\ 0 & \alpha & 1 & \dots & 0 & 0 \\ 0 & 0 & \alpha & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \alpha & 1 \\ 0 & 0 & 0 & \dots & 0 & \alpha \end{pmatrix}.$$

   The matrices $k^{1-u}\alpha_n^k a^{-k}$, $k \in \mathbb{N}$ have a subsequence converging to some matrix c.

   Let $\bar{G}_e$ denote the connected component of e in $\bar{G}$. Let $S = \{(g_1,g_2) \in \bar{G}_e \times \bar{G}_e : g_1 c g_2 \text{ has 1,1 entry equal to } 0\}$, a proper Zariski closed subset of $\bar{G}_e \times \bar{G}_e$, as is S^{-1}. Take $(g_1,g_2) \in \bar{G}_e \times \bar{G}_e \setminus (S \cup S^{-1})$. In other words, $g_1 c g_2$ and $g_1^{-1} c g_2^{-1}$ both have nonzero 1,1 entry. Consider the powers $\tilde{a}_k = a^k g_1 a^{-k} g_2$ and $g_2 \tilde{a}_k^{-1} g_2^{-1} = a^k g_1^{-1} a^{-k} g_2^{-1}$; as $k \to \infty$, the rows other than the first converge to 0, so both matrices have unique dominant eigenvalues. Hence, $\tilde{a}_k$ is the desired element, for large enough k.)
7. Prove the Tits alternative over a field of characteristic > 0. (Hint: The same series of steps as in Remark 19B.23 (for characteristic 0) shows that there is an element $a \in G$ having a characteristic value α that is not a root of 1; one can choose an absolute value on F with $|\alpha| \neq 1$ and thus assume that $|\alpha| > 1$. In view of Proposition 00.6 (of the Prerequisites), one may take some high enough power of a and assume that a is diagonal. Modifying the argument of Step IV by means of composition series, one may assume that α has the largest absolute value of all characteristic values of a. Now, in the notation of Exercise B6, take $u = 1$ and finish with the same argument.)

Products of closed subgroups of an algebraic group

A **locally closed** set is the intersection of an open set and a closed set. A set is called **constructible** if it is a finite union of locally closed sets (in the Zariski topology). Clearly the constructible sets are closed under finite intersections, unions, and set complements.

8. Prove that the image of a constructible set is constructible, for any morphism $\Phi\colon X \to Y$ of varieties. (Hint: Replace Y by the closure of $\Phi(X)$. By Theorem 10.36 of Volume 1, $\phi(X)$ contains an open subset U of Y, and $\dim Y \backslash U < \dim Y$. By induction, the image of each irreducible component of $\phi^{-1}(Y \setminus U)$ is constructible, and their union together with U comprise a constructible set, which is $\phi(X)$.)
9. Prove that the product of constructible subsets $Y_1, \dots, Y_k$ of G is constructible. (Hint: $Y_1 \cdots Y_k$ is the image of $Y_1 \times \cdots \times Y_k$ under the product morphism $G \times \cdots \times G \to G$.)
10. Suppose A and B are closed subgroups of an algebraic group G. Show that if $H = AB$ is a subgroup of G, then H is closed. (Hint: H is constructible, and thus contains an open subset of $\overline{H}$.)
11. Suppose G is a linear algebraic group, $\phi_i\colon X_i \to G$ are morphisms of varieties, and $e \in Y_i = \phi_i(X_i)$ for each i. Show that the smallest closed subgroup H of G containing all Y_i is connected and is a finite product of the Y_i. (Hint: If necessary, also throw in the morphisms $x_i \mapsto \phi_i(x_i)^{-1}$. Any finite product $Y = Y_{i_1} \cdots Y_{i_k}$ of the Y_i is constructible, by Exercise B9. Furthermore, given any two finite products Y, Z of the Y_i, one has $\overline{Y}\,\overline{Z} = \overline{YZ}$ by the following argument analogous to Exercise 10:

    Take a finite product Y of the Y_i such that $\overline{Y}$ is maximal possible. $\overline{Y}$ is unique, and in particular $\overline{Y}$ contains $\overline{Y}\,\overline{Y}$ and $\overline{Y^{-1}}$. Thus $\overline{Y}$ is a closed subgroup of G containing each Y_i, and thus equals H.

But Y, being constructible, contains an open subset of H, implying $H = YY$.)

12. Prove that any subgroup of a linear algebraic group generated by a family of closed connected subgroups is closed and connected, and is generated by a finite number of them.
13. Prove that for any closed subgroups A and B of an algebraic group G, with A connected, the commutator group (A, B) is closed and connected. (Hint: Define $\varphi_a\colon B \to G \times G$ by $\varphi_a(b) = (b^a, b^{-1})$. Then $\varphi_a(B)$ is closed and connected; apply Exercise B12.)
14. Using Exercise B13, show that the commutator of two closed subgroups of an algebraic group G is closed. In particular, all the derived subgroups of G are closed, and all subgroups in its upper central series are closed.

Solvable algebraic groups

Kolchin's Theorem has an analog for connected solvable groups that ties in later (Chapter 21) to Engel's Theorem.

15. For any solvable (resp. nilpotent, resp. unipotent) subgroup H of an algebraic group G, show that its closure $\overline{H}$ is also solvable (resp. nilpotent, resp. unipotent). (Hint: $H^{(t)} = \{e\}$ is a closed condition in the Zariski topology. Likewise for nilpotent or unipotent.)
16. Prove for F algebraically closed that any connected solvable algebraic subgroup G of $\mathrm{GL}(n, F)$ is conjugate to a subgroup of $\mathrm{T}(n, F)$. (Hint: Induction on n. $V = F^{(n)}$ must be simple as a G-space, since otherwise one could apply induction to a G-subspace W and V/W.

    It suffices to show that G has a common nonzero eigenvector $v \in V$. This would prove the theorem, since then Fv is a G-subspace of V, implying $V = Fv$ and G is Abelian.

    The assertion is obvious for solvability degree $t = 1$, so assume that $t \geq 2$. By induction on t, G' has a common nonzero eigenvector. The span of all such eigenvectors is a G-subspace, so V has a base consisting of eigenvectors of G'. With respect to this base, G' is Abelian.

    Now fix $a \in G'$ and consider the morphism $\phi\colon G \to G'$ given by $g \mapsto gag^{-1}$. Since gag^{-1} is diagonal with the same entries as a, $\phi(G)$ is finite, as well as connected, and thus is the single element a. Hence, $G' \subseteq Z(G)$, so $G' = F$ by Schur's Lemma. On the other hand, any element of G' has determinant 1, so G' consists of roots of 1, and being connected, must be $\{e\}$.)
17. For any connected solvable algebraic group G, show that the group G' is nilpotent. (Hint: Reduce the question to $T(n, F)$.)

Chapter 20

Throughout these exercises, G denotes a finite group and ζ_n denotes a primitive n-th root of unity.

1. Show that the complex conjugate $\bar{\chi}$ of a character χ is also a character; $\bar{\chi}$ is irreducible iff χ is irreducible. (Hint: Use the contragredient representation.)
2. Using Exercise 1, show that if t_1 is the number of real irreducible characters of G, then $t \equiv t_1 \pmod 2$.

Schur's orthogonality relations

3. Writing the matrix $\rho_i(g)$ as $\left(g^{(i)}_{uv}\right)$ and $\rho_i(g^{-1})$ as $\left(\tilde{g}^{(i)}_{uv}\right)$, show that

$$\sum_{g \in G} g^{(i)}_{ks} \tilde{g}^{(j)}_{u\ell} = \delta_{i,j}\delta_{u,s}\delta_{k,\ell}\frac{|G|}{n_i}.$$

    (Hint: Notation as in Exercise 19.6, identify L_i and L_j respectively as $F^{(n_i)}$ and $F^{(n_j)}$ with bases $e_1, e_2, \ldots, e_{n_i}$ and $e'_1, e'_2, \ldots, e'_{n_j}$. For each k and ℓ, define the F-linear map $\psi_{k,\ell}\colon L_i \to L_j$ by $\sum_s \alpha_s e_s \mapsto \alpha_k e'_\ell$. As in Lemma 19.25, $\overline{\psi}_{k,\ell}(e_s) = \sum_{u=1}^{n_j} \sum_{g \in G} g^{(i)}_{ks} \tilde{g}^{(j)}_{u\ell} e'_u$; confront this with Exercise 19.6 and match entries in the matrices. Note that $\sum_{s=1}^{n_i} g^{(i)}_{ks} \tilde{g}^{(i)}_{s\ell}$ is the k, ℓ position of the identity matrix.)
4. Reprove Theorem 20.5 by taking traces in Exercise 3.
5. Prove the following converse to Proposition 20.4(v): If $\chi_i(g)$ are real for all i, then g is conjugate to g^{-1}. (Hint: Schur II.)

Character tables

6. Finish Example 20.16 without using Schur's relations. (Hint: Let ρ denote the irreducible representation of degree 2. If $\ker \rho$ were nontrivial, then $G/\ker \rho$ would be Abelian; hence, ρ is faithful. In particular, $\rho(c) = -I$ and thus $\chi_\rho(c) = -2$. For every element $g \in G \setminus Z(G)$, the two eigenvalues of $\rho(g)$ must be distinct. But $\rho(g)^2 \in \langle \rho(c) \rangle$. Hence, $\mathrm{tr}\, \rho(g) = 0$.)
7. For any splitting field F of G, show that the values of the character table of G are elements of F integral over $\mathbb{Z}$.
8. Verify the following counterexample to the converse of Exercise 7: $\mathbb{R}$ is not a splitting field of the quaternion group Q, even though the values of the character table are integers.
9. Prove that any two nonisomorphic finite Abelian groups have different character tables.
10. Compute the character table of $G = S_4$. (Hint: The Klein group $K = \{(1), (12)(34), (13)(24), (14)(23)\}$ is a normal subgroup. Furthermore, identifying S_3 with the subgroup of permutations fixing the

number 4, one sees that $S_3 \cap K = (1)$, so three characters arise from $G/K \cong S_3$. Since the images of (12) and (1234) are the same, start with the first three rows and finish with Schur.)

11. Compute the character table of any nonabelian group G of order 21. (Hint: Let a be an element of order 7. The subgroup $\langle a\rangle$ is normal, so $G/\langle a\rangle$ has order 3. On the other hand, no subgroup of order 3 is normal, so the only characters of degree 1 are χ_1, χ_2, and χ_3. Hence, $t = 5$ and $n_4 = n_5 = 3$; also, $m_4 = m_5 = 3$. The six elements $a, \dots, a^6$ are in the kernel of χ_4 and thus are divided among the last two conjugacy classes. Assume that $a \in \mathcal{C}_4$ and $a^{-1} \in \mathcal{C}_5$, yielding

	Conj.		Reps.		
	1	b	b^2	a	a^{-1}
χ_1	1	1	1	1	1
χ_2	1	ζ	ζ^2	1	1
χ_3	1	ζ^2	ζ^4	1	1
χ_4	3	?	?	?	?
χ_5	3	?	?	?	?

,

where $\zeta = \zeta_3$. Using Schur II on each column shows that $\chi_{42} = \chi_{52} = \chi_{43} = \chi_{53} = 0$. Let $\alpha = \chi_{44}$. $-2\alpha(\alpha+1) = 4$, implying $\alpha = -\frac{1}{2} \pm i\frac{\sqrt{7}}{2}$.)

12. Show for any G with only two conjugacy classes, that $G \cong C_2$.
13. The dihedral group D_n has the following representations of degree 2, for $1 \le j \le n$:

$$a \mapsto \begin{pmatrix} \omega_n^j & 0 \\ 0 & -\omega_n^j \end{pmatrix}, \qquad b \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

For which j are these irreducible? Use this information to build the character table of D_n.

14. Build the character table of A_4, noting how it varies from that of S_4.
15. Show that the number of conjugacy classes of a group G of order 18 must be 6, 9, or 18. (Hint: G has a normal subgroup of order 9, so $[G : G']$ is 2, 6, or 18.)
16. Work out the character tables of all groups of each order < 24 other than 16. The case of order 16 is cumbersome to calculate since there are ten nonabelian groups of order 16; however, even this case can be attacked by noting that there is a nontrivial central element z of order 2 and $|G/\langle z\rangle| = 8$.

Degrees of characters

17. Suppose χ is an integer-valued character of a finite group G and $g \in G$. For any prime number p dividing $\circ(g)$, write $\circ(g) = p^t m$ where

p does not divide m, and let $h = g^{p^t}$ (so $\circ(h)$ is prime to p). Show that $\chi(h) \equiv \chi(g) \pmod p$. (Hint: One may assume that G is cyclic, and thus $\chi = \sum n_i\chi_i$ where the χ_i are of degree 1. But $h^q = g^q$ for a suitably large power q of p, so Fermat's Little Theorem shows that $\chi(g) \equiv \chi(g)^q \equiv \chi(h)^q \equiv \chi(h) \pmod p$.)

In Exercises 18–20 we use the notation of Remark 20.21.

18. Show that $Z_i/\ker\rho_i \cong Z(G/\ker\rho_i)$.
19. Prove that $n_i^2 \le \frac{|G|}{|Z_i|}$, equality holding iff $Z_i = \ker\rho_i$. (Hint: By Schur I, $|G| = \sum_{g\in G} |\chi_i(g)|^2 \ge \sum_{g\in Z_i} |\chi_i(g)|^2$.)
20. Prove that n_i divides $[G : Z_i]$ for each i. (Hint: Assume that $\ker\rho_i = 1$. Then $Z_i = Z(G)$ by Exercise 18. Take a subset $\{s_1, \dots, s_k\} \subset G$ of elements not conjugate modulo Z_i, maximal such that $\chi_i(s_j) \ne 0$ for each j. The s_j lie in distinct conjugacy classes, say $\mathcal{C}_1, \dots, \mathcal{C}_k$, and $\bigcup_j \mathcal{C}_j Z_i = \{g \in G : \chi_i(g) \ne 0\}$ with no duplication. Hence,

$$|G| = \sum_{\chi_i(g)\ne 0} \chi_i(g)\overline{\chi_i(g)} = \sum_{1\le i\le k} |Z_i|\chi_i(s)\overline{\chi_i(s)},$$

so conclude as in Theorem 20.18.)

In particular, n_i divides $[G : Z(G)]$, to be improved in Exercise 22.

21. Suppose ρ is an irreducible representation of G, and N is a normal subgroup of G. Show that either the restriction of χ_ρ to N has the form $m\chi_j$ for some $m \in \mathbb{N}$ and some irreducible character χ_j of N, or G has some proper subgroup $H \supseteq N$ such that ρ is induced from an irreducible representation of H. (Hint: Write the $F[G]$-module corresponding to ρ as $V = \oplus_{j=1}^k L_j^{(m_j)}$, where the L_j are simple $F[N]$-modules. For any $v \in L_j$, $gL_j = gNv = Ngv$, so G acts transitively as permutations on the L_j; either $k = 1$ or else H can be taken to be the stabilizer of L_1.)
22. Prove that the degree of each irreducible character of G divides $[G : A]$ for any Abelian normal subgroup A. (Hint: Induction on $|G|$, using Exercise 21. In the first case, replace G by $\rho(G)$, whose center contains $\rho(A)$. In the second case, replace G by H and induce.)
23. Prove that any nonabelian group G of order p^3 (for p prime) has precisely $p^2 + p - 1$ irreducible complex characters, p^2 of which are of degree 1, and $p-1$ of which are of degree p. (Hint: Each irreducible character has degree 1 or p.) Hence, G has p^2+p-1 conjugacy classes.

Frobenius reciprocity

24. For any representation ρ of H, show that $\chi_{\rho^G}(1) = [G : H]$.

25. Suppose $M = \oplus_{i=1}^{t} M_i^{(u_i)}$, where the M_i are nonisomorphic simple G-spaces, and assume that F is a splitting field for G. Then

$$\operatorname{Hom}_{F[G]}(M_i, M) \cong F^{(u_i)} \cong \operatorname{Hom}_{F[G]}(M, M_i)$$

as vector spaces over F. (Hint: Exercise 19.6.)

26. For any splitting field F of G, verify that the multiplicity of any irreducible representation ρ in a representation σ is equals to both $\dim_F \operatorname{Hom}_{F[G]}(M, N)$ and $\dim_F \operatorname{Hom}_{F[G]}(N, M)$, where M, N are the respective modules of ρ and σ. Thus, Frobenius Reciprocity follows from the "adjoint isomorphism" of Proposition 18.44.

27. For any representation ρ of finite degree of a subgroup $H \subseteq G$, show that the contragredient $(\rho^G)^*$ of the induced representation is equivalent to the induced representation $(\rho^*)^G$. (Hint: By Exercise 19.5, for any $F[H]$-module M, one needs to define a monic

$$\operatorname{Hom}_F(F[G] \otimes_{F[H]} M, F) \to F[G] \otimes_{F[H]} \operatorname{Hom}(M, F)$$

as $F[G]$-modules. Let (J) denote the involution on $F[H]$ of Exercise 19.28. Take a transversal $s_1, s_2, \dots$ of G over H. Given $(\sum s_i a_i, f)$ for $a_i \in F[H]$ and $f\colon M \to F$, define the map $F[G] \otimes_{F[H]} M \to F$ by $s_i a_i \otimes v \mapsto f(a_i^J v)$.)

Chapter 21

Lie algebras of low dimension over a field

1. Verify that any anticommutative algebra L of dimension 2 satisfies the Jacobi identity and thus is a Lie algebra; if L is nonabelian, then L is isomorphic to the Lie algebra of Example 21.5(ii). (Hint: Write $L' = Fa$ and extend a to a base $\{a, b\}$ of L.)
2. Suppose L is a Lie algebra of dimension 3 such that $\dim L' = 1$. Show that $Z(L) \neq 0$. Conclude that either L is isomorphic to the solvable Lie algebra of upper triangular 3×3 matrices, or L is the direct sum of Lie subalgebras of dimensions 1 and 2. (Hint: Take $a, b \in L$ with $c = [ab] \neq 0$. Assume $[bc] = 0$. If $[ac] = 0$, then identify a with e_{12} and b with e_{23}. If $[ac] = \alpha c \neq 0$, then $b - \alpha^{-1}c \in Z(L)$ and is not in $Fa + Fc$.)
3. Construct infinitely many nonisomorphic Lie algebras of dimension 3 over $\mathbb{C}$. (Hint: Take $\dim L' = 2$.)
4. Show that the Lie algebra L is solvable whenever $\dim L' \leq 2$.
5. Show that L is simple whenever $\dim L = 3$ with $L' = L$. (Hint: If $0 \neq I \lhd L$, then I and L/I are solvable.)

6. In Exercise 5, show that $L \cong sl(2, F)$ whenever there exists nonzero $a \in L$ for which ad_a has a nonzero eigenvalue in F; in particular, the Lie algebra of Exercise 5 is unique up to isomorphism when F is algebraically closed.
7. Construct a simple 3-dimensional real Lie algebra not isomorphic to $sl(2, \mathbb{R})$.

The classical Lie algebras

8. Take a vector space V having a bilinear form $\langle\ ,\ \rangle$. Let L be the set of skew transformations, i.e., those $a \in R$ satisfying $\langle av, w\rangle = -\langle v, aw\rangle$; cf. Exercise 14.11. Then L is a Lie subalgebra of $(\operatorname{End} V)^-$, called **orthogonal** (resp. **unitary**, **symplectic**) when $\langle\ ,\ \rangle$ is symmetric (resp. Hermitian, alternate). Show that B_n and D_n are orthogonal Lie algebras, whereas C_n is a symplectic Lie algebra. (Hint: $\left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$ corresponds to the alternate form defining C_n.)
9. Display B_n as the set of skew-symmetric matrices with respect to the quadratic form $x_0^2 + 2x_1x_2 + 2x_3x_4 + \cdots$, and display D_n as the set of skew-symmetric matrices with respect to the quadratic form $x_1x_2 + x_3x_4 + \cdots$.
10. In characteristic $\neq 2$, prove that the Lie algebra B_n is simple for each $n \geq 1$, and C_n and D_n are simple Lie algebras for all $n > 2$. (Hint: Any Lie ideal $I \neq 0$ of B_n contains some nonzero element $a = \sum_{i<j} \alpha_{ij}(e_{ij} - e_{ji})$. To ease notation, put $\alpha_{ji} = -\alpha_{ij}$. Assume that $\alpha_{12} \neq 0$. For $n \geq 2$,

$$[e_{23} - e_{32}, a] = \sum \big(\alpha_{3j}(e_{2j} - e_{j2}) - \alpha_{2j}(e_{3j} - e_{j3})\big).$$

Lie commutate this with $e_{12} - e_{21}$, and then again with $e_{23} - e_{32}$, to show that $e_{12} - e_{21} \in I$. The computations for C_n and D_n are similar.)
11. Prove that $gl(n, F)' = sl(n, F)$.
12. Show that $sl(2, F)$ is nilpotent if $\operatorname{char}(F) = 2$.
13. Show that the Lie algebra D_2 is not simple, since it has the Lie ideal consisting of all matrices of the form

$$\begin{pmatrix} 0 & \alpha & \beta & \gamma \\ -\alpha & 0 & -\gamma & -\beta \\ -\beta & \gamma & 0 & \alpha \\ -\gamma & \beta & -\alpha & 0 \end{pmatrix}.$$

Herstein's theorems on Lie structure

Verify the assertions in Exercises 14–28, which, following Herstein, tie the structure of an associative ring R with its Lie structure. *Assume throughout that R is a prime, associative, noncommutative ring, and* $2 \neq 0$. Z denotes $\operatorname{Cent}(R)$, and A' denotes $[A, A]$ for $A \subseteq R$.

14. If $B \lhd R$ and $[a, B] = 0$, then $a \in Z$. (Hint: $B[a, R] = 0$.)
15. If $a \in R$ satisfies $[a, [a, R]] = 0$, then $a \in Z$. (Hint: Applying Equation (21.2) to $[a, [a, bc]] = 0$ yields $2[a, b][a, c] = 0$. Taking rb instead of c and using (21.2) again yields $2[a, b]R[a, b] = 0$, so $[a, b] = 0$.)
16. If $[a, [a, B]] = 0$ for $0 \neq B \lhd R$, then $a \in Z$. (Hint: As in Exercise 15, show that $[a, B] = 0$.)
17. If $[c, R'] = 0$, then $c \in Z$. (Hint: c commutes with all $[a, b]$ and thus with $[a, ab] = a[a, b]$, so $[a, c][a, b] = 0$; conclude as in Exercise 15.)
18. Any Abelian Lie ideal I of R is central. (Hint: If $a \in I$, then $[a, [a, r]] \in [I, I] = 0$.)
19. If $L \neq 0$ is a left ideal of R and $0 \neq B \lhd R$, then $L + [L, B]$ contains the nonzero ideal LB of R. Likewise for right ideals.
20. Any Lie ideal I that is a subring of R is either central or contains a nonzero ideal of R. (Hint: Otherwise Exercise 18 implies there are $a, b \in I$ with $[a, b] \neq 0$. But $[a, bc] \in I$ for any $c \in R$, so (21.2) yields $[a, b]c \in I$. Hence, $[a, b]R \subseteq I$, so use Exercise 19.)
21. R satisfies $[a, bc] = [ab, c] + [ca, b]$ for all $a, b, c \in R$. Conclude that if I is a Lie ideal of R, then $\hat{I} = \{r \in R : [R, r] \subseteq I\}$ is a Lie ideal and subring containing I.
22. Any noncentral Lie ideal I of R contains $[R, B]$ for some $0 \neq B \lhd R$. In particular, for R simple, $R' \subseteq I$. (Hint: $\hat{I}$, defined in Exercise 21, contains an ideal $B \neq 0$ by Exercise 20; i.e., $[R, B] \subseteq I$.)
23. The left annihilator A of a nonzero Lie ideal I of R' is 0. (Hint: Otherwise Exercise 19 shows that $A + R'$ contains an ideal $B \neq 0$ of R. But $abI = a[b, I] = 0$ for $a \in A$ and $b \in [R, B]$, so $A[R, B] \subseteq A$. Now $A^2B \subseteq ABA + A[B, A]$ yields $A^2BI = 0$, which is impossible.)
24. Any Abelian Lie ideal I of R' is contained in Z. (Hint: One needs to show that $\mathrm{ad}_a = 0$ for any $a \in I$. Let $\delta = \mathrm{ad}_a$. By assumption, $\delta^2([r, s]) = 0$ for all $r, s \in R$. Equation (21.2) yields $\delta^2(R[r, I]) = 0$, taking $s \in RI$. Applying Leibniz' rule and taking $r \in R'$ yields $\delta^2(R)[I, R'] = 0$. By Exercise 23, either $\delta^2(R) = 0$ or $[I, R'] = 0$, implying $I \subseteq Z$ by Exercise 17.)
25. If I is a Lie ideal of R' such that $I' \subseteq Z$, then $I \subseteq Z$. (Hint: Otherwise, by Exercise 24, there are a, b in I such that $0 \neq [a, b] \in Z$. Let $\delta = \mathrm{ad}_a$. Then $\delta(b^2) = 2[a, b]b$, implying $\delta^2(ab^2) = 2a[a, b]^2 \in I \setminus Z$. But manipulating $[a, \delta^2(ab^2)]$ and $[b, \delta^2(a^2b^2)]$ leads to $\delta^2(ab^2) \in Z$, a contradiction. Alternatively, for R primitive, one can utilize Jacobson's Density Theorem to reduce to the easy case for which $R = M_n(D)$.)
26. $[R', R']$ contains $[R, B]$ for some $0 \neq B \lhd R$. (Hint: $R' \not\subseteq Z$ by Exercise 17, so $[R', R'] \not\subseteq Z$ by Exercise 25. Conclude with Exercise 22.)
27. If R is simple (of characteristic $\neq 2$), then the only proper Lie ideals of R' are central. (Hint: Take a proper Lie ideal I of R' and define $\hat{I}$ as in Exercise 21. Note that $[R', R'] = R'$ by Exercise 26. Hence, $R' \not\subseteq \hat{I}$, implying $I + \hat{I} \neq R$. But $R\hat{I}' \subseteq \hat{I}$, so then $I + \hat{I}$ contains $R\hat{I}'$ and thus $R\hat{I}'R$, implying $\hat{I}' = 0$. Hence, $I'' = 0$, so $I' \subseteq Z$, by Exercise 24, implying $I \subseteq Z$ by Exercise 25.)

28. If T is an additive subgroup of a simple ring R of characteristic $\neq 2$ such that $[T, R'] \subseteq T$, then either $T \supseteq R'$ or $T \subseteq Z$. (Hint: $[T, R']$ is a Lie ideal of R', so conclude by Exercise 27 unless it is central. But then the Lie ideal $\{r \in R : [r, R'] \subseteq Z\}$ contains T and is central in view of previous exercises.)
29. (Lanski) Give a version of Exercise 28 for prime rings.

Restricted Lie algebras

30. A Lie algebra L of characteristic p is called **restricted** if it also has a unary operation, denoted $a \mapsto a^{[p]}$, satisfying the axioms $\mathrm{ad}_{a^{[p]}} = \mathrm{ad}_a^p$, $(\alpha a)^{[p]} = \alpha^p a^{[p]}$, and $(a + b)^{[p]} - a^{[p]} - b^{[p]} = \sum_{j=1}^{p-1} \Delta_j(a, b)$, where $j\Delta_j(a, b)$ is the coefficient of λ^{j-1} in $\mathrm{ad}^{p-1}_{\lambda a+b}(a)$. Show that this definition agrees with Definition 21.21′ for the case of linear Lie algebras. In characteristic p, show that R^- is a restricted Lie algebra for any associative algebra R, where $a^{[p]} = a^p$, and $\mathrm{Der}(R)$ also is a restricted Lie algebra. Describe the algebraic structure theory of restricted Lie algebras.

Jordan-Hölder theory for Lie modules

31. Define a chain of Lie modules to be a **composition series** if every factor is simple as a Lie module. For each f.d. Lie module, prove that the factors of all composition series are unique up to permutation and isomorphism. (Hint: Insert the word 'Lie' before 'module' at every instance in the proof of the Jordan-Hölder Theorem; cf. Theorem 3.11 of Volume 1.)

Nilpotent Lie ideals

32. Show that $(I_1 + I_2)^{m+n} \subseteq \sum\{I_1^i I_2^j : i \geq m \text{ or } j \geq n\}$ for any Lie ideals I_1 and I_2. Conclude that the sum of two nilpotent Lie ideals is nilpotent. Consequently, any f.d. Lie algebra has a unique largest nilpotent Lie ideal.
33. Show that the nonabelian 2-dimensional Lie algebra L has a unique maximal nilpotent ideal I of dimension 1, but the Lie algebra L/I is Abelian (and thus nilpotent). This example indicates the difficulty in developing a "nilpotent" radical in Lie theory.
34. Show that $I \cap Z(L) \neq 0$, for any nilpotent Lie algebra and any nonzero Lie ideal I. (Hint: Take $V = I$ in Lemma 21.28; $\mathrm{ad}\, L$ is nilpotent, so $[La] = 0$ for some $0 \neq a \in I$.)

Linear Lie algebras of characteristic 0

In Exercises 35–39, assume that L is a Lie subalgebra of $gl(n, F)$ with $\operatorname{char} F = 0$. Let $\bar{L}$ denote the associative subalgebra of $M_n(F)$ generated by L.

35. For any Lie subalgebras I and J of L with $I \subseteq [J, L]$ and $[I, J] = 0$, show that $I^n = 0$. If, furthermore, I is a Lie ideal of L and $\bar{L}$ is a semisimple ring, show that $I = 0$. (Hint: Take $a \in I$. Then $a^k = aa^{k-1} \in [J, L]a^{k-1} = [J, La^{k-1}]$ has trace 0 for each k, so by Newton's formulas [**Row3, p. 153**], each coefficient of the characteristic polynomial of a is 0; i.e., a is nilpotent. Hence, I is nilpotent. For the last assertion, $\bar{L}I$ is a nilpotent ideal of $\bar{L}$.)
36. When $\bar{L}$ is semisimple, prove that $\operatorname{rad} L = Z(L)$ and $L = Z(L) \oplus L_1$ for some semisimple Lie subalgebra L_1 of L containing L'. (Hint: Let $J = [L, \operatorname{rad} L]$. Then $J^{(k)} = 0$ for some $k \geq 1$. Hence, $J^{(k-1)}$ is an Abelian Lie ideal, so Exercise 35 implies $I = [J^{(k-1)}, L]$ is 0. Now $J^{(k-1)} \subseteq Z$, so Exercise 35 implies $J^{(k-1)} = 0$. By induction conclude that $J = 0$.

    Thus $\operatorname{rad} L \subseteq Z(L)$, so $\operatorname{rad} L = Z(L)$. Exercise 35 now shows that $L' \cap \operatorname{rad} L = 0$. Take a complementary space L_1 to $\operatorname{rad} L$ containing L'.)
37. When L is solvable and the ring $\bar{L}$ is semisimple, show that L is Abelian. (Hint: L_1 of Exercise 36 is solvable semisimple and thus 0.)

The radical of a f.d. Lie algebra

38. Prove that $[\operatorname{rad} L, L] \subseteq \operatorname{Jac}(\bar{L})$. (Hint: $\bar{L}/\operatorname{Jac}(\bar{L})$ is a semisimple associative algebra generated by the image of L, implying the image of $\operatorname{rad} L$ is central.)
39. Prove that $[\operatorname{rad} L, L]$ is Lie nilpotent for any f.d. Lie algebra L of characteristic 0. (Hint: Apply the adjoint representation to Exercise 38.) This result is improved in Exercise 50.
40. For any derivation δ of a f.d. Lie algebra L of characteristic 0, show that $\delta(\operatorname{rad} L)$ is contained in a nilpotent Lie ideal. (Hint: Assume that $L = \operatorname{rad} L$ is solvable. Define the Lie algebra $L_1 = L \oplus Fd$, where $\operatorname{ad}_d = \delta$. Then L_1 is solvable, so $L_1' = L' + \delta(L)$ is nilpotent.)
41. Show that the radical of a Lie algebra is contained in the radical of the trace bilinear form with respect to any representation. (Hint: As in the proof of Lemmas 21.45 and 21.46.)

The Casimir element and Whitehead's lemmas

42. Show that the Casimir element of an irreducible Lie representation is always invertible. (Hint: Schur's Lemma.)

43. For any Lie module V over L and any $v \in V$, show that the map $f\colon L \to V$ defined by $f(a) = av$ satisfies

$$f([ab]) = af(b) - bf(a), \quad \forall a, b \in L. \tag{E21.1}$$

44. (Whitehead's First Lemma.) Suppose the Lie module V is f.d. over F and $f\colon L \to V$ is an F-linear map satisfying (E21.1). Show that there is $v \in V$ such that $f(a) = av$, $\forall a \in L$. (Hint: First assume that V is simple, corresponding to the representation ρ. Let $\{e_1, \ldots, e_n\}$ be a base of L over F, and let $\{e_1^*, \ldots, e_n^*\}$ denote the dual base with respect to the bilinear form of ρ. Put $v = c_\rho^{-1} \sum e_i f(e_i^*)$. For V not simple, proceed by induction on dimension: Take a Lie submodule W of V, find $\bar{v} = v_1 + W$ for $L \to V/W$, and then apply induction on the map $L \to W$ given by $a \mapsto f(a) - av_1$.)
45. Prove Weyl's Theorem as a consequence of Whitehead's First Lemma.
46. Suppose L is a Lie algebra and $I \lhd L$ satisfies $I^2 = 0$. Let $\bar{L} = L/I$ and view I as an $\bar{L}$-module via $b\bar{a} = [ba]$. Write $L = T \oplus I$ as vector spaces. Identify $\bar{L}$ with T to get a vector space injection $h\colon \bar{L} \to L$. Define $f\colon \bar{L} \to I$ by

$$f(\bar{a}_1, \bar{a}_2) = [h(\bar{a}_1), h(\bar{a}_2)] - h([\bar{a}_1 \bar{a}_2]).$$

    Prove that $h(\bar{L})$ is a Lie subalgebra of L iff f is 0.
47. (Whitehead's Second Lemma.) Suppose L is a f.d. semisimple Lie algebra over a field F of characteristic 0, the Lie module V is f.d. over F, and the map $f\colon L \times L \to V$ satisfies $f(a, a) = 0$ and

$$\sum_{i=1}^{3} f(a_i, [a_{i+1}, a_{i+2}]) + a_i f(a_{i+1}, a_{i+2}) = 0,$$

    subscripts modulo 3. Prove that there is a map $g\colon L \to V$ satisfying

$$f(a_1, a_2) = a_1 g(a_2) - a_2 g(a_1) - g([a_1 a_2]). \tag{E21.2}$$

    (Hint: As in Exercise 44, take dual bases of L with respect to the trace form of the representation ρ corresponding to V; substitute $a_1 \mapsto e_i$, multiply on the left by e_i^*, and use the Casimir element c_ρ in a way reminiscent of Whitehead's First Lemma. Using the Fitting decomposition with respect to c_ρ, cf. Remark 2.67 of Volume 1, reduce to the case when c_ρ is either invertible or nilpotent.)

    Whitehead's Second Lemma can be formulated in terms of Lie cohomology, saying that $H^2(L, V) = 0$; cf. Example 25.39.

Levi's Theorem and its consequences

48. (Levi's Theorem.) Prove that any f.d. Lie algebra L of characteristic 0 can be decomposed as vector spaces $L = S \oplus I$, where $I = \operatorname{rad} L$ and $S \cong L/I$ is a semisimple Lie algebra. (Hint: By induction on $\dim L$, one may assume that $I^2 = 0$; i.e., I is Abelian. Then $I \subseteq \ker \rho$, where ρ is the representation corresponding to L as a Lie module. Write $\bar{L} = L/I$. Notation as in Exercise 46, let

$$f(\bar{a}_1, \bar{a}_2) = [h(\bar{a}_1), h(\bar{a}_2)] - h([\bar{a}_1, \bar{a}_2]). \tag{E21.3}$$

Multiply Equation (E21.3) on the left by $\bar{a}_0$; also substitute $\bar{a}_0\bar{a}_1$ for $\bar{a}_1$ in (E21.3). Combining these two equations attains the hypotheses of Whitehead's Second Lemma; then take $\tilde{h} = h - g$. Exercise 46 shows that $\tilde{h}(\bar{L})$ is a Lie subalgebra of L.)
49. In characteristic 0, show that $[L, \operatorname{rad} L] = L' \cap \operatorname{rad}(L)$. (Hint: Exercise 48 implies $L' = S' \oplus [L, \operatorname{rad} L]$; intersect with $\operatorname{rad}(L)$.)
50. (Strengthening Corollary 21.34.) Prove that $L' \cap \operatorname{rad}(L)$ is Lie nilpotent, for any f.d. Lie algebra L of characteristic 0. (Hint: Exercises 39 and 49.)

Cartan subalgebras over arbitrary fields of characteristic 0

51. Let $L = L_0 \oplus L_1$ denote the Fitting decomposition of a Lie algebra L under ad_a; cf. Remark 2.67 of Volume 1. When ad_a is nilpotent, show that L_0 is a Lie subalgebra and $[L_0 L_1] \subseteq L_1$. (Hint: Iterate the Jacobi identity, noting that $L_1 = \operatorname{ad}_a^n L$.)
52. Show for any nilpotent Lie subalgebra N of a f.d. Lie algebra L (over an arbitrary field), that the intersection of the null components of L with respect to every $a \in N$ is a Lie subalgebra of L.
53. Apply Exercise 52 to the proof of Theorem 21.71 to establish the existence of Cartan subalgebras in any f.d. Lie algebra of characteristic 0.
54. A **toral subalgebra** of a f.d. semisimple Lie algebra is a Lie subalgebra consisting only of semisimple elements. Prove that any toral subalgebra T is Abelian. (Hint: For each $a \in T$, the action of ad_a on T has no nonzero eigenvalues; indeed, if $[a, b] = \alpha b$, then write a as a linear combination of eigenvectors of ad_b, and apply ad_b.) Humphreys [**Hum1, pp. 36, 37**] shows that any maximal toral algebra is Cartan, thereby providing an alternate approach to the theory.

Weight spaces

The next few exercises generalize the root space decomposition of a Lie algebra L with respect to a nilpotent Lie subalgebra. A **weight** of a Lie module V is some $\mathbf{w} \in L^* = \operatorname{Hom}_F(L, F)$ such that, $\forall a \in L$, $(a - \mathbf{w}(a)I)^k v = 0$ for some $k = k(a)$ and suitable $0 \neq v \in V$; the set of such v is called the **weight space** for $\mathbf{w}$. A weight space that is also a Lie submodule of V is called a **weight module**.

55. Verify that each weight space M is a subspace of V; show that M is a weight module when L is Lie nilpotent.
56. For any Lie modules V_1 and V_2, show that $V_1 \otimes V_2$ is a Lie module with respect to multiplication $a(v_1 \otimes v_2) = av_1 \otimes v_2 + v_1 \otimes av_2$.
57. If M_1 and M_2 are weight spaces for weights $\mathbf{w}_1$ and $\mathbf{w}_2$ in Lie modules V_1 and V_2, respectively, show that $M_1 \otimes M_2$ is the weight module for $\mathbf{w}_1 + \mathbf{w}_2$ in $V_1 \otimes V_2$. (Hint:

$$(a - (\alpha + \beta)I)\,(v_1 \otimes v_2) = (a - \alpha)v_1 \otimes v_2 + v_1 \otimes (a - \beta)v_2;$$

iterate and check that one eventually gets 0.)
58. Suppose $L = \oplus L_{\mathbf{a}}$ is the root space decomposition of L with respect to a nilpotent Lie subalgebra N, and $V = \oplus V_{\mathbf{w}}$ is the corresponding decomposition of V into weight spaces. Show that $L_{\mathbf{a}} V_{\mathbf{w}} \subseteq V_{\mathbf{a}+\mathbf{w}}$ if $\mathbf{a}+\mathbf{w}$ is a weight, and $L_{\mathbf{a}} V_{\mathbf{w}} = 0$ otherwise.

Root systems of f.d. semisimple Lie algebras

59. Show that the bilinear form (21.18) satisfies $\langle f, g\rangle = \sum_{\mathbf{b}} \langle f, \mathbf{b}\rangle\langle \mathbf{b}, g\rangle$, summed over all roots $\mathbf{b}$ (Hint: Remark 21.63 and Summary 21.80).
60. Show that $\langle \mathbf{a}, \mathbf{a}\rangle = \sum_{\mathbf{b}} \langle \mathbf{a}, \mathbf{b}\rangle^2$ for any root $\mathbf{a}$. (Hint: Exercise 59.)
61. Verify that the angle between any two roots $\mathbf{a} \neq \pm\mathbf{b}$ is 30°, 45°, 60°, 90°, 120°, 135°, or 150°.

The multiplication table of a semisimple Lie algebra

As usual, assume throughout that the base field is algebraically closed field of characteristic 0.

62. Show that any relation in the semisimple Lie algebra L is a consequence of the relations (21.27) and the relations in $\mathfrak{n}_+$ and $\mathfrak{n}_-$.
63. In Corollary 21.110, prove the formulas

$$[e_{j_1} e_{j_2} \cdots e_{j_\ell} h_i] = -\sum_{u=1}^{\ell} m_{ij_u} [e_{j_1} e_{j_2} \cdots e_{j_\ell}]$$

and $[f_{j_1} f_{j_2} \cdots f_{j_\ell} h_i] = \sum_{u=1}^{\ell} m_{ij_u} [f_{j_1} f_{j_2} \cdots f_{j_\ell}]$.
64. Show that $f(\Phi) = \Phi'$ whenever V and V' are vector spaces with respective root systems Φ and Φ' and $f: V \to V'$ is an isometry that sends the simple roots of Φ onto the simple roots of Φ'. (Hint: Induction on height.)
65. Prove that if two semisimple Lie algebras have identical Cartan matrices (with respect to their respective Cartan subalgebras and simple root systems), then they are isomorphic. (Hint: Match their generators.)

66. Prove that the decomposition of a root system into indecomposable root systems is unique.
67. Call a root **a maximal** if $\mathbf{a}+\mathbf{a}_i$ is not a root for all positive roots $\mathbf{a}_1,\dots,\mathbf{a}_n$. Prove that every root system of a simple Lie algebra L has a unique maximal root $\mathbf{a}$. Explicitly, if $\mathbf{a}=\sum k_i\mathbf{a}_i$, then, for any root $\sum j_i\mathbf{a}_i$, each $|j_i|\le k_i$. In particular, each $k_i\ge 1$. (Hint: Take a maximal root $\mathbf{a}=\sum k_i\mathbf{a}_i$. Each $k_i\ge 0$. If some $k_i=0$, then put $S_1=\{\mathbf{a}_i : k_i=0\}$ and $S_2=S\setminus S_1$. For $\mathbf{a}_i\in S_1$, maximality of $\mathbf{a}$ confronted with Remark 21.87 implies $\langle\mathbf{a},\mathbf{a}_i\rangle=0$. Let Φ_u be the set of roots spanned by S_u for $u=1,2$. Then $\Phi=\Phi_1\cup\Phi_2$, since otherwise one would have a root $\mathbf{b}+\mathbf{a}_i$ where $\mathbf{a}_i\in\Phi_2$ and $\mathbf{b}\in\Phi_1$; then the root $\sigma_i(\mathbf{b}+\mathbf{a}_i)=\mathbf{b}-\mathbf{a}_i$ would be neither positive nor negative, impossible. But L is simple, so $S_1=\emptyset$. If $\mathbf{b}$ were another maximal root then, as in the proof of Theorem 21.91, $\langle\mathbf{a},\mathbf{b}\rangle>0$, implying $\mathbf{a}-\mathbf{b}$ were a root, again impossible).
68. Notation as in Theorem 21.108, show that L has the **Chevalley involution** (i.e., Lie anti-automorphism of order 2) given by $e_i^*=-f_i$ and $h^*=-h$ for $h\in\mathfrak{h}$.

Geometry of positive root systems: Weyl chambers

Suppose Φ is a root system of $V=F^{(n)}$. For any $\mathbf{a}\in\Phi$, define the **hyperplane** $\mathbf{a}^\perp=\{v\in V : \langle\mathbf{a},v\rangle=0\}$. The **Cartan-Stieffel diagram** $D'=D'(\Phi)$ is defined as $\bigcup_{\mathbf{a}\in\Phi}\mathbf{a}^\perp$. $D'\ne V$ by the exclusion principle (Exercise 0.3 of Volume 1). The connected components of $V\setminus D'$ are open cones; their closures are called the **Weyl chambers** of Φ.

69. Show that the diagram D' is invariant under the action of the Weyl group $\mathcal{W}$, so the Weyl chambers are permuted by $\mathcal{W}$. Furthermore, the action of $\mathcal{W}$ on the Weyl chambers is transitive.
70. For a simple root system $\{\mathbf{a}_1,\dots,\mathbf{a}_n\}$, show that the set $\{v\in V : \langle\mathbf{a}_i,v\rangle\le 0 \text{ for all } 1\le i\le n\}$ is a Weyl chamber, called the **fundamental Weyl chamber** of the system. Conclude that the Weyl group acts transitively on each simple root system.

Affine Lie algebras – The Kac-Moody-Kantor approach

Generalizing Proposition 21.60, we say that a Lie module L over a Lie algebra $\mathfrak{h}$ has a **root space decomposition** if $L=\mathfrak{h}\oplus(\oplus L_{\mathbf{a}})$, where each $L_{\mathbf{a}}$ is an eigenspace of dimension 1.

71. Show that any generalized $n\times n$ Cartan matrix $A=(m_{ij})$ of rank k corresponds "uniquely" (up to base change) to a vector space $\mathfrak{h}$ of dimension $2n-k$ together with linearly independent elements $\mathbf{a}_1,\dots,\mathbf{a}_k\in\mathfrak{h}^*$ and $h_1,\dots,h_k\in\mathfrak{h}$ satisfying $\mathbf{a}_i(h_j)=m_{ij}$. (Hint: The $\mathbf{a}_i$ are the rows of A; the rest is linear algebra.)

72. Use Exercise 71 to define a Lie algebra $\mathcal{L}$ containing an Abelian Lie subalgebra $\mathfrak{h}$ and generators e_i,f_i, $1\le i\le k$, via the rules $[e_i,f_j]=\delta_{ij}h_i$, $[e_i,h]=-\mathbf{a}_i(h)e_i$, $[f_i,h]=\mathbf{a}_i(h)f_i$. Obtain a root space decomposition for $\mathcal{L}$, and write $\mathcal{L}=\mathfrak{n}_+\oplus\mathfrak{h}\oplus\mathfrak{n}_-$, where $\mathfrak{n}_+$ (resp. $\mathfrak{n}_-$) are free Lie algebras generated by the e_i (resp. f_i). (Hint: Verify a Lie structure inductively on the homogeneous components of the tensor algebra $T(\mathfrak{h})$; cf. Example 18.38.)
73. Show that the Lie algebra $\mathcal{L}$ of Exercise 72 has a unique maximal Lie ideal $\mathcal{I}$ intersecting $\mathfrak{h}$ trivially; define the **Kac-Moody** Lie algebra $\mathcal{L}/\mathcal{I}$ and obtain its root space decomposition. Also, generalize the Chevalley involution (Exercise 68) to Kac-Moody algebras.
74. Generalize Zassenhaus' theorem to Kac-Moody Lie algebras.
75. Show that the following assertions are equivalent for an indecomposable, symmetric generalized Cartan matrix A:

    (i) A has finite type.

    (ii) The bilinear form of Corollary 21.116 is positive definite.

    (iii) The Weyl group is finite.

    (iv) The number of roots is finite.

    (v) The Lie algebra of A is simple finite-dimensional.

    (vi) There exists a maximal root $\mathbf{a}$, in the sense of Exercise 67.

    (Hint: ((ii) $\Rightarrow$ (iii)) This requires some theory of algebraic groups; cf. Appendix 19B. The Weyl group is a closed subgroup of the orthogonal group and thus is compact as well as discrete; hence, it is finite.)
76. Verify that the loop algebra $\mathcal{L}(L)$ is indeed a Lie algebra, and the injection $L\to\mathcal{L}(L)\otimes L$ given by $a\mapsto 1\otimes a$ is a Lie homomorphism.
77. Notation as in Example 21.119, define a derivation δ_k on $\tilde{\mathcal{L}}$ by $\epsilon\mapsto 0$ and $\delta_k|_{\mathcal{L}(L)}=-\lambda^{k+1}\frac{\partial}{\partial\lambda}$. Define the **centerless Virasoro**, or **Witt**, Lie algebra to be the Lie subalgebra $\mathfrak{d}=\oplus_{k\in\mathbb{Z}}F\delta_k$ of $\operatorname{Der}\tilde{\mathcal{L}}$. Show that $\mathfrak{d}$ is defined by the relations $[\delta_i\delta_j]=(i-j)\delta_{i+j}$. Also verify the ascending chain condition on Lie ideals for the Lie subalgebra $\mathfrak{d}^+=\oplus_{k\in\mathbb{N}^+}F\delta_k$, although the descending chain condition on Lie ideals fails for $\mathfrak{d}^+$.
78. (The Virasoro algebra.) Notation as in Exercise 77, define the **Virasoro** Lie algebra to be the extension $\mathfrak{d}\oplus Fc$ of $\mathfrak{d}$ by a one-dimensional central subspace Fc, satisfying $[\delta_j\delta_{-j}]=2j\delta_0+\frac{j^3-j}{12}c$ and $[\delta_i\delta_j]=(i-j)\delta_{i+j}$ for $i\ne -j$. Verify that this is an affine Lie algebra; also show that any central extension of $\mathfrak{d}$ is a homomorphic image of

the Virasoro algebra. (Hint: Write $[\delta_i\delta_j] = (j-i)\delta_{i+j} + c(i,j)z$ and apply anticommutativity and the Jacobi identity.)

Duality in Cartan matrices

Define the following notation on vectors $\mathbf{a} = (\alpha_1, \ldots, \alpha_\ell)$ and $\mathbf{b} = (\beta_1, \ldots, \beta_\ell)$ in $\mathbb{R}^{(\ell)}$: $\mathbf{a} > \mathbf{b}$ if $\alpha_i > \beta_i$ for each $1 \le i \le \ell$; likewise, $\mathbf{a} \ge \mathbf{b}$ if $\alpha_i \ge \beta_i$ for each $1 \le i \le \ell$.

79. (Farkas' Theorem [**DorSS**].) Suppose $A = (\alpha_{ij})$ is an arbitrary $k \times \ell$ matrix over the field F. Write $\mathbf{a}_i = (\alpha_{i1}, \ldots, \alpha_{i\ell})$ for $1 \le i \le k$. Prove that the system $\sum_j \alpha_{ij}\lambda_j > 0$ of linear inequalities for $1 \le i \le k$ has a simultaneous solution over $\mathbb{R}$ iff every non-negative, nontrivial, linear combination of the $\mathbf{a}_i$ is nonzero. (Hint: ($\Rightarrow$) Suppose there is a solution $\mathbf{c} = (\gamma_1, \ldots, \gamma_\ell) \in \mathbb{R}^{(\ell)}$. For any $\beta_i \ge 0$, not all 0,

$$\Big(\sum_i \beta_i \mathbf{a}_i\Big) \cdot \mathbf{c} = \sum_{ij} \beta_i \alpha_{ij} \gamma_j = \sum_i \beta_i \Big(\sum_j \alpha_{ij}\gamma_j\Big) > 0.$$

($\Leftarrow$) An argument based on the dot product metric. Consider the cone $\mathcal{C}$ of vectors of the form $\mathbf{v}A$ where $\mathbf{v} \ge \mathbf{0}$. $\mathcal{C}$ is convex. One may assume that $\mathbf{0} \notin \mathcal{C}$, so by compactness, $\mathcal{C}$ has a point $\mathbf{p}$ at a minimum distance from $\mathbf{0}$.

Consider any $\mathbf{q} \ne \mathbf{p}$ on $\mathcal{C}$. By hypothesis on $\mathbf{p}$, the cosine of the angle $\mathbf{0pq}$ is not positive, i.e.,

$$0 \ge (\mathbf{q}-\mathbf{p}) \cdot (\mathbf{0}-\mathbf{p}) = \mathbf{p}\cdot\mathbf{p} - \mathbf{q}\cdot\mathbf{p}.$$

Hence, $\mathbf{q}\cdot\mathbf{p} \ge \mathbf{p}\cdot\mathbf{p} > 0$ for each $\mathbf{q} \in \mathcal{C}$, so take $\mathbf{q} = \mathbf{a}_i = e_iA$ for each i.)

80. (The Fundamental Theorem of Game Theory!) Prove that if there does not exist $\mathbf{x} > 0$ in $\mathbb{R}^{(\ell)}$ (written as a column) with $A\mathbf{x} < 0$, then there exists $\mathbf{w} \ge 0$ (written as a row) in $\mathbb{R}^{(k)}$ with $\mathbf{w}A \ge 0$. (Hint: Define $\mathbf{a}_i = (a_{i1}, \ldots, a_{i\ell})$. By hypothesis, there is no simultaneous solution to the inequalities $-\mathbf{a}_i \cdot \mathbf{x} > 0$, $\mathbf{x} > 0$. In other words, defining $\tilde{A}$ to be the $(k+\ell) \times \ell$ matrix $\begin{pmatrix} -A \\ I \end{pmatrix}$, there is no solution to $\tilde{A} \cdot \mathbf{x} > 0$. Hence, by Exercise 79, the rows of $\tilde{A}$ are dependent with non-negative coefficients.)

81. Show that the generalized Cartan matrix A^t has the same type as A. (Hint: Since A^t and A have the same rank, it suffices to show that if A has indefinite type, then A^t also has indefinite type; this follows by Exercise 80.)

Poisson algebras

A **Poisson algebra** is an associative algebra together with a bilinear operation $\{\ ,\ \}: A \times A \to A$, called a **Poisson bracket**, satisfying the **Leibniz identities**

$$\{ab, c\} = a\{b,c\} + \{a,c\}b, \quad \{a, bc\} = \{a,b\}c + b\{a,c\}, \quad \forall a,b,c \in A.$$

82. For any Poisson algebra A and any $a \in A$, show that there is a derivation δ_a of A given by $b \mapsto \{a,b\}$.

Verify that Exercises 82–86 are examples of Poisson algebras, all of which are commutative.

83. If L is a f.d. Lie algebra with base $a_1, \ldots, a_n$, then, viewing the a_i as commuting indeterminates in the commutative polynomial algebra $R = F[a_1, \ldots, a_n]$, introduce a Poisson structure on R by defining $\{a_i, a_j\}$ to be the Lie product in L and extending the Poisson bracket via the Leibniz identities. (The same idea works more generally for an arbitrary Lie algebra L, where one takes its symmetric algebra.)

84. Suppose V is a f.d. vector space with an alternating bilinear form. Take a base $\{x_1, \ldots, x_n\}$ of V. The polynomial algebra $F[x_1, \ldots, x_n]$ becomes a Poisson algebra, where one defines $\{x_i, x_j\}$ to be $\langle x_i, x_j\rangle$.

85. If R is an associative algebra filtered by $\mathbb{Z}$ such that the associated graded algebra $\mathrm{gr}(R)$ is commutative, then $\mathrm{gr}(R)$ has a Poisson bracket defined as follows: For $a + R_{i-1} \in \mathrm{gr}(R)_i$ and $b + R_{j-1} \in \mathrm{gr}(R)_j$, define $\{a,b\}$ to be $[a,b] + R_{i+j-2} \in \mathrm{gr}(R)_{i+j-1}$. Thus, the Poisson bracket has a (-1)-grading.

86. Suppose R is an algebra with a regular central element z such that $\bar{R} = R/Rz$ is commutative. Then $\bar{R}$ has a Poisson structure, obtained by defining $\{\bar{a}, \bar{b}\} = \bar{r}$, where r is taken such that $[a,b] = rz$. As pointed out by Goodearl, excellent examples are the quantized coordinate algebras of Appendix 16A, where $z = q - 1$.

87. Show that any associative algebra is a Poisson algebra with respect to the Lie bracket $\{a,b\} = ab - ba$.

The next few exercises lead to a theorem of D. Farkas and G. Letzter, stating that, up to multiplication by an element of the extended centroid (cf. Exercise 16.24), Exercise 87 provides the only possible Poisson structure on a prime, associative, noncommutative algebra.

88. Suppose R is an associative ring with a Poisson bracket $\{\ ,\ \}$. Writing $[a,b] = ab - ba$, verify the identity

$$[a,c]\{b,d\} = \{a,c\}[b,d], \qquad \forall a,b,c,d \in R.$$

(Hint: Compute $\{ab, cd\}$ twice, playing off the two equations in Exercise 82.)

89. Notation as in Exercise 88, prove that $[a,c]r\{b,d\} = \{a,c\}r[b,d]$ for all a,b,c,d,r in R. (Hint: Compute $[a,c]\{rb,d\}$ twice, using Exercise 88.)
90. (Farkas-Letzter) For any prime ring R with a Poisson bracket, show that every $a,b \in R$ satisfy $[a,b] = c\{a,b\}$ for some c in the extended centroid of R. (Hint: $[a,b]r\{a,b\} = \{a,b\}r[a,b]$ for all r in R; apply Exercise 16.25.)
91. In Exercise 90, c is independent of the choice of a and b. Conclude that the Poisson bracket reduces to the ring commutator when R is prime associative and noncommutative.

Appendix 21A

The Lie algebra of an algebraic group

Suppose G is an algebraic group with coordinate algebra $\mathcal{A}$. For any $g \in G$, let $\mathfrak{m}_g \subset \mathcal{A}$ denote the maximal ideal of polynomials vanishing on g. Let F_g denote the copy $\mathcal{A}/\mathfrak{m}_g$ of F. $T(G)_e$ denotes the tangent space at e as defined in Definition 10A.3 of Volume 1; in our notation, $T(G)_e = \mathrm{Der}(\mathcal{A}, F_e)$. We use the notation of Remark 19B.13ff. Given $f \in \mathcal{A}$, write $\Delta f = \sum_i f_{i1} \otimes f_{i2} \in \mathcal{A} \otimes \mathcal{A}$.

1. Given $\mathbf{x} \in T(G)_e$ and $f \in \mathcal{A}$, define the **convolution** $f * \mathbf{x}$ to be the map $g \mapsto \mathbf{x}(\hat{\ell}_{g^{-1}} f)$, $\forall g \in G$. Show that $f * \mathbf{x} = \sum_i \mathbf{x}(f_{i2}) f_{i1} \in \mathcal{A}$. (Hint: $\hat{\ell}_{g^{-1}} f = \sum_i f_{i1}(g) f_{i2}$ for each $g \in G$, implying
$$(f * \mathbf{x})(g) = \mathbf{x}(\hat{\ell}_{g^{-1}} f) = \mathbf{x}\big(\sum f_{i1}(g) f_{i2}\big) = \mathbf{x}(f_{i2}) f_{i1}(g).)$$
2. Define $\Phi\colon \mathrm{Lie}(G) \to T(G)_e$ by taking $\Phi(\delta)$ to be the derivation $f \mapsto \delta(f)(e)$, viewed as an element in $T(G)_e$. Prove that $\Phi(\delta)$ is a vector space isomorphism whose inverse is given by $\mathbf{x} \mapsto \delta_{\mathbf{x}}$, where $\delta_{\mathbf{x}}\colon \mathcal{A} \to \mathcal{A}$ is defined as the map $f \to f * \mathbf{x}$. (Hint: $\delta_{\mathbf{x}}$ is indeed a derivation, being the composite of a derivation and a homomorphism. Next, $\hat{\ell}_a(\delta_{\mathbf{x}} f)(g) = (\delta_{\mathbf{x}} f)(a^{-1}g) = \mathbf{x}(\hat{\ell}_{g^{-1}} \hat{\ell}_a f) = \delta_{\mathbf{x}}(\hat{\ell}_a f)(g)$ for all $a, g \in G$.

   Thus, $\mathbf{x} \mapsto \delta_{\mathbf{x}}$ does define a map $T(G)_e \to \mathrm{Lie}(G)$, which is seen to be the inverse of Φ.)
3. Define $\mathbf{x}_1 \cdot \mathbf{x}_2$ by $(\mathbf{x}_1 \cdot \mathbf{x}_2)(f) = (\mathbf{x}_1 \otimes \mathbf{x}_2)(\Delta f) = \sum \mathbf{x}_1(f_{i1}) \mathbf{x}_2(f_{i2})$. Verify that $[\mathbf{x}_1, \mathbf{x}_2] = \mathbf{x}_1 \cdot \mathbf{x}_2 - \mathbf{x}_2 \cdot \mathbf{x}_1 = \Phi([\delta_{\mathbf{x}_1}, \delta_{\mathbf{x}_2}])$; conclude that the Lie product in $T(G)_e$ corresponds (via Φ) to the natural Lie product of derivations in $\mathrm{Lie}(G)$.
4. For any morphism $\varphi\colon G \to H$ of algebraic groups, prove that the morphism $d\varphi\colon T(G)_e \to T(H)_e$ preserves the Lie product. (Hint: Take h in the coordinate algebra of H and let $f = \varphi^*(h)$. For $\mathbf{x}_i \in T(G)_e$, and $\mathbf{y}_i = d\varphi(\mathbf{x}_i)$,
$$\begin{aligned}[\mathbf{x}_1, \mathbf{x}_2](h) &= \delta_{\mathbf{y}_1}\delta_{\mathbf{y}_2}(h)(e) - \delta_{\mathbf{y}_2}\delta_{\mathbf{y}_1}(h)(e) \\ &= \mathbf{x}_1(\varphi^*(\delta_{\mathbf{y}_2}(h))) - \mathbf{x}_2(\varphi^*(\delta_{\mathbf{y}_1}(h)))\end{aligned}$$
and $d\varphi([\mathbf{x}_1, \mathbf{x}_2])(h) = \mathbf{x}_1(\delta_{\mathbf{x}_2}(f)) - \mathbf{x}_2(\delta_{\mathbf{x}_1}(f))$, so it suffices to show that $\delta_{\mathbf{x}_i}(f) = \varphi^*(\delta_{\mathbf{y}_i}(h))$, or $\hat{\ell}_{a^{-1}} f = \varphi^*(\hat{\ell}_{\varphi(a)^{-1}} h)$.)

Computing the Lie algebra

5. By Example 6B.2 of Volume 1, any derivation $\delta\colon F[\lambda_1, \ldots, \lambda_n] \to F$ is determined by its restriction $\bar{\delta}$ to $V = \sum F\lambda_i$. $\bar{\delta}$ belongs to the dual space V^*, whose base can be identified with $\partial/\partial_1, \ldots, \partial/\partial_n$, since $\partial\lambda_j/\partial_i = \delta_{ij}$ (the Kronecker delta function).

   For $A = (x_{ij})$, show that the differential of the formula for $\det A$ is $\sum_{i,j=1}^n A_{ij} d(x_{ij})$, where A_{ij} is the i,j signed minor of A; moreover, specializing A to the identity matrix sends A_{ij} to 0 except for $i = j$. Conclude that any derivation δ of $\det(x_{ij})$ evaluated at the identity matrix is $\sum_{i=1}^n \delta(x_{ii})$, i.e., the trace of $\delta(A)$.
6. Using Exercise A5, compute the Lie algebras of the algebraic groups GL, SL, O, and Sp, thereby obtaining the classical Lie algebras. (Hint: View the Lie algebra of G as $T(G)_e$ and then differentiate the defining relations of G, as a variety, as follows:

   (i) $\mathrm{GL}(n,F)$ is viewed as an affine variety in $F^{(n+1)^2}$, whose indeterminates can be written as $\{\lambda_{ij} : 1 \le i,j \le n^2\}$. $\mathrm{GL}(n,F)$ is the variety of the polynomial $\lambda_{n+1,n+1} \det(\lambda_{ij}) - 1$. Differentiating and evaluating at the identity matrix yields $\sum_{i=1}^{n+1} \delta(\lambda_{ii}) = 0$. Thus, $T(G)_e$ is identified with $M_n(F)$, where the entries are indexed by the ∂/∂_{ij}, the partial derivatives with respect to the λ_{ij} for $1 \le i,j \le n$. Recalling that $\Delta(\lambda_{ij}) = \sum_{k=1}^n \lambda_{ik} \otimes \lambda_{kj}$, note that
$$\big((\alpha_{ij}) \cdot (\beta_{ij})\big)(\lambda_{ij}) = \sum_{k=1}^n (\alpha_{ij})(\lambda_{ik}) \cdot (\beta_{ij})(\lambda_{kj}) = \sum_{k=1}^n \alpha_{ik}\beta_{kj}(\lambda_{ij}),$$
   so $(\alpha_{ij}) \cdot (\beta_{ij})$ is precisely the matrix product! Conclude that the Lie algebra of $\mathrm{GL}(n,F)$ can be identified with $gl(n,F)$.

   (ii) $\mathrm{SL}(n,F)$ is the variety of the polynomial $\det(\lambda_{ij}) - 1$. The same line of reasoning (but using n instead of $n+1$) yields $\sum_{i=1}^n \delta(\lambda_{ii}) = 0$; i.e., $T(G)_e$ is identified with $sl(n,F)$, the Lie algebra A_{n-1}.

   (iii) The orthogonal group $\mathrm{O}(n,F) = \{a \in \mathrm{GL}(n,F) : aa^t = I\}$. Differentiating $(\lambda_{ij})(\lambda_{ji}) = 1$ yields $(\lambda_{ij})\delta(\lambda_{ij}) + \delta(\lambda_{ij})(\lambda_{ij}) = 0$, so specializing (λ_{ij}) to the identity matrix yields $\delta(\lambda_{ij}) + \delta(\lambda_{ji}) = 0$;

the Lie algebra of $\mathrm{O}(n,F)$ is identified as the set of skew-symmetric matrices, which is B_m for $n=2m+1$ and D_m for $n=2m$.

(iv) By the analogous argument to (iii), the Lie algebra of the symplectic group $\mathrm{Sp}(n,F)$ is the set of matrices that are skew-symmetric with respect to the symplectic involution, which is C_m for $n=2m$.)

Appendix 21B

Structure of nonassociative algebras

1. Show that $a0=0a=0$ for any element a of a nonassociative algebra A. For any homomorphism $f\colon A\to B$ of nonassociative algebras, $\ker f = f^{-1}(0)$ is an ideal of A. Conversely, for $I\lhd A$, define an algebra structure on A/I with I the kernel of the natural map $A\mapsto A/I$. Conclude that A is a simple algebra iff any homomorphism $f\colon A\to B$ is 1:1.
2. When C contains an infinite field, show for each commutative associative C-algebra H and C-algebra R that every identity of R is also an identity of $R\otimes_C H$. (Hint: Vandermonde argument.)
3. Suppose A is a f.d. algebra over a field K, and F is any subfield of K. Taking a base $b_1,\dots,b_n$ of A, write

$$b_ib_j=\sum_{i=1}^n \alpha_{ijk}b_k,$$

and let $F_1=F(\alpha_{ijk}: 1\le i,j,k\le n)$, a field of finite transcendence degree over F. Show that $A_1=\sum F_1b_i$ is an algebra with multiplication as in A, and A can be identified with $A_1\otimes_{F_1}K$. In this way one can cut the base field K down to a field F_1 of finite transcendence degree over F. This method of reduction via structure constants was introduced in Remark 5.15′ of Volume 1.

The nucleus

4. Prove that any algebra satisfies the identity

$$a[b,c,d]+[a,b,c]d=[ab,c,d]-[a,bc,d]+[a,b,cd].$$

5. Define the **nucleus** to be the set of elements $n\in A$ such that $[n,b,c]=[b,n,c]=[b,c,n]=0$, $\forall b,c\in A$. Show that the nucleus is an associative subalgebra of A.

Alternative algebras

6. Show that the tensor product of an alternative algebra with an associative algebra is alternative. In particular, if A is alternative, then the matrix algebra $M_n(A)$ is alternative.

7. Is the tensor product of two alternative algebras necessarily alternative?
8. If A is an algebra, define the new algebra $A_1=F\oplus A$ with respect to multiplication $(\alpha,a)(\beta,b)=(\alpha\beta,\alpha b+\beta a+ab)$. Verify that if A is alternative, then A_1 is alternative with unit element $(1,0)$.
9. Prove that any alternative algebra satisfies the **Moufang identities**

$$a\big(b(ac)\big)=(aba)c,\quad c\big(a(ba)\big)=c(aba),\quad (ab)(ca)=a(bc)a.$$

(Hint: $[a,ab,c]=[a^2,b,c]-a[a,b,c]$. Hence, $(aba)c-a\big(b(ac)\big)=[ab,a,c]+[a,b,ac]=-[a,ab,c]-[a,ac,b]=-[a^2,b,c]-[a^2,c,b]+a\big([a,b,c]+[a,c,b]\big)=0$. The next identity is by symmetry. Finally, $(ab)(ca)-a(bc)a=[a,b,ca]-a[b,c,a]=-[a,ca,b]-a[b,c,a]=a\big([c,a,b]-[b,c,a]\big)=0$. This requires the first Moufang identity.)
10. Verify the following identities in an alternative algebra:

$$[a,bc,d]+[c,ba,d]=-a[b,c,d]-c[b,a,d]; \tag{E21B.1}$$

$$(ab)(cd)+(db)(ca)=(a(bc))d+(d(bc))a. \tag{E21B.2}$$

(Hint: The first Moufang identity yields $[a,ba,d]=-a[b,a,d]$ which linearized yields (E21B.1); (E21B.2) follows from the third Moufang identity.)
11. (Artin) Prove that any alternative algebra generated by two elements is associative. (Hint: It suffices to check $[h_1,h_2,h_3]=0$ where $h_i=h_i(a_1,a_2)$ are nonassociative monomials in a_1 and a_2. Proceeding by induction on the total degree of the associator, one may write each h_i without parentheses, so in particular $h_i=g_ia_{u_i}$ for $u_i\in\{1,2\}$. Two of the u_i are the same, so one may assume that $a_{u_1}=a_{u_2}$, i.e., $h_i=g_ia$ for $i=1,2$, where $a\in\{a_1,a_2\}$. Applying induction to (E21B.1) shows that $[g_1a,g_2a,h_3]=-[a,g_2g_1a,h_3]=-\big(a(g_2g_1)a\big)h_3+a\big((g_2g_1a)h_3\big)$, which by the first Moufang identity is $a[g_2g_1,a,h_3]$ and thus 0.)

Composition algebras

12. Verify the following identities for all a,b,c in a composition algebra:

(i) $Q(c)\langle a,b\rangle=Q(ca+cb)-Q(ca)-Q(cb)=\langle ca,cb\rangle$.

(ii) $\langle ac,bc\rangle=Q(c)\langle a,b\rangle$.

(iii) $\langle a,b\rangle\langle c,d\rangle=\langle ac,bd\rangle+\langle ad,bc\rangle$.

(Hint: (iii) Substitute $\langle c,d\rangle=Q(c+d)-Q(c)-Q(d)$ on the left side and reduce.)

13. Show that any composition algebra $(A,*)$ is alternative when $\frac12\in F$. (Hint: It suffices to prove that $0=[a,a^*,b]=Q(a)b-a(a^*b)$. Taking $b=1$ in Exercise B12(iii) yields

$$(a+a^*)\langle c,d\rangle=\langle a^*,1\rangle\langle c,d\rangle=\langle a^*c,d\rangle+\langle a^*d,c\rangle.$$

Taking $d = a^*b$ yields $\langle a(a^*b), c\rangle = \langle (a+a^*)a^*b, c\rangle - \langle a^*(a^*b), c\rangle = \langle a^*c, a^*b\rangle = Q(a)\langle c, b\rangle = \langle Q(a)b, c\rangle$.)

14. Show that any composition algebra $(A, *)$ is $(*)$-simple in the sense that it has no proper nonzero $(*)$-ideals. (Hint: For any ideal I and any $b \in I$, $b + b^* \in I \cap F = 0$. But for all $a \in A$, note that $a^*b \in I$, implying $\langle a, b\rangle = a^*b + (a^*b)^* = 0$; conclude that $I = 0$.)
15. Show that any composition algebra that is not simple has the form $F \oplus F$ with the exchange involution. (Hint: Start as in Exercise 14.20.)
16. Prove that if K is some ν-double of a field F, then the μ-double of K is the generalized quaternion F-algebra (μ, ν) of Exercise 14.9.
17. Suppose $(\mathcal{A}, *)$ is a composition algebra and $(A, *)$ is a subalgebra under which the bilinear form arising from the restriction of Q is nondegenerate. Prove that $(\mathcal{A}, *)$ contains some ν-double of $(A, *)$. (Hint: Take the orthogonal complement $A^\perp$ of A and $c \in A^\perp$ such that $Q(c) \neq 0$. Then $0 = \langle 1, c\rangle = c + c^*$, so $c^* = -c$. Also $\langle a, c\rangle = 0$, $\forall a \in A$, implying $a^*c = ca$. It follows that $A \cap Ac = 0$, so $A \oplus Ac$ is the ν-double of A, where $\nu = -Q(c) = c^2$.)
18. Verify the following assertions, where $(\mathcal{A}, *)$ is the ν-double of the composition algebra $(A, *)$:

    (i) $\mathcal{A}$ is commutative associative iff A is commutative associative with $(*) = 1_A$.

    (ii) $\mathcal{A}$ is associative iff A is commutative associative.

    (iii) $\mathcal{A}$ is alternative iff A is associative.

Octonian algebras

19. Define a **generalized octonion algebra** to be the double of a generalized quaternion algebra (Exercise 14.9). Prove that any composition F-algebra must be either F itself, the direct product of two copies of F (with the exchange involution), a quadratic field extension of F, a generalized quaternion algebra, or a generalized octonion algebra. (Hint: $F \subseteq A$. If $F \neq A$, then some double is in A. Continue; the doubling process must stop after 3 steps; cf. Exercise B18.)
20. (Hurwitz' Theorem) Prove that if $\mathbb{C}$ satisfies an identity

$$\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} z_i^2,$$

    where z_i are forms of degree 2 in the x_i and y_j, then $n = 1, 2, 4$, or 8. (Hint: Translate this to a quadratic form on a composition algebra, so reduce to Exercise B19.) Compute the actual identities corresponding to the composition algebras $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$. For example the identity for $\mathbb{C}$ is $(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 - x_2y_2)^2 + (x_1y_2 + x_2y_1)^2$.

21. Show that every nonassociative, f.d. alternative division algebra A is a generalized octonion algebra. (Hint: Any two noncommuting elements generate an associative division algebra D. Let $\bar{F}$ denote the algebraic closure of F. Then $D \otimes_F \bar{F} \cong M_t(\bar{F})$ for some t. Then $t \leq 2$, for otherwise $M_3(\bar{F})$ is contained in the simple alternative algebra $A \otimes \bar{F}$, which must thus be associative, a contradiction. This proves that every element of D is quadratic, and it follows that D is a composition algebra. Conclude that A is a composition algebra.)
22. (Zorn) Prove that every f.d. simple nonassociative, alternative algebra A is a generalized octonion algebra. (Hint: Tensor by $F(\lambda_1, \lambda_2, \dots)$ where the λ_i are infinitely many indeterminates over F. Then two "generic" elements generate an associative division subalgebra, which by Exercise B21 must be a generalized quaternion algebra. Conclude that A is a composition algebra.)
23. Define the **generic** octonion algebra to be the λ_3-double of the symbol (λ_1, λ_2) over the field $F(\lambda_1, \lambda_2, \lambda_3)$. Show that this is an alternative division algebra.
24. Show that over an algebraically closed field, the only nonassociative simple alternative algebra up to isomorphism is the split octonion algebra.

Idempotents in alternative algebras

25. Define idempotents as in the associative theory. For any orthogonal idempotents e_i, e_j of an alternative algebra A, show that $[e_i, e_j, a] = 0$ for all $a \in A$. (Hint: $[e_i, e_j, a] = -e_i(e_ja) = -e_i^2(e_ja) = e_i[e_j, e_i, a] = 0$ using (E21B.1).)
26. (Peirce decomposition.) Suppose $e_1, \dots, e_t$ are pairwise orthogonal idempotents of an alternative algebra A. Then defining

$$A_{ij} = \{a \in A : e_k a = \delta_{ik} a \quad \text{and} \quad ae_k = \delta_{jk} a, \quad 1 \leq k \leq t\},$$

    prove that $A = \bigoplus_{i,j} A_{ij}$, using Exercise B25.
27. Prove that the Peirce decomposition satisfies $A_{ij}A_{jk} \subseteq A_{ik}$ and $A_{ij}A_{ij} \subseteq A_{ji}$.
28. For any simple alternative algebra A with nontrivial orthogonal idempotents e_1 and e_2 satisfying $e_1 + e_2 = 1$, show that $e_1Ae_1 = A_{11} = A_{12}A_{21}$. (Hint: $A_{12}A_{21} + A_{12} + A_{21} + A_{21}A_{12}$ is an ideal and thus is all of A; match components.)
29. Prove that any simple alternative algebra A containing three pairwise orthogonal idempotents e_1, e_2, and e_3 is associative. (Hint: Show that $(e_1 + e_2)A(e_1 + e_2)$ is associative, and in particular $A_{12}^2 = 0$. By symmetry, all A_{ii} are associative and all $A_{ij}^2 = 0$.)

Jordan algebras

30. Show that A^+ is a Jordan algebra for any alternative algebra A. (Hint: One need only check the identities on algebras generated by two elements; cf. Exercise B11.)
31. In any special Jordan algebra, show that one could define the quadratic operator $U_a\colon J \to J$ given by $U_a(r) = ara$ and replace $\circ$ by the operators U_a for all $a \in J$. (Hint: Equation (21B.3).)
32. Given any Jordan algebra J and $a \in J$, let $\ell_a \in \mathrm{End}_F J$ denote the map $x \mapsto a \circ x$. Show that $\ell_{a^2}\ell_a = \ell_a\ell_{a^2}$, and
$$[[\ell_a\ell_b]\ell_c] = \ell_{a\circ(b\circ c)-b\circ(a\circ c)}.$$
Conclude that $[\ell_{a\circ b}\ell_c] + [\ell_{b\circ c}\ell_a] + [\ell_{c\circ a}\ell_b] = 0$.
33. Continuing Exercise B32, define $U_a = 2\ell_a^2 - \ell_{a^2}$. Verify the following operator identities:
(Q1) $U_1 = 1_J$.
(Q2) $U_{U_a b} = U_a U_b U_a$.
(Q3) If $V_{a,b}$ is defined as $V_{a,b}(r) = (U_{a+r} - U_a - U_r)(b)$ for $a, b, r \in J$, then $U_a V_{b,a} = V_{a,b} U_a$.
Conversely, if $\frac{1}{2} \in F$ and J satisfies (Q1), (Q2), and (Q3), then J becomes a Jordan algebra by defining $\ell_a = \frac{1}{2}U_{a,1}$ and $a \circ b = \ell_a(b)$.
34. Define a **quadratic Jordan algebra** over F to be a vector space J endowed with the operations $U_a\colon J \to J$ such that every tensor extension $J \otimes_F K$ (K a field extension of F) satisfies the identities (Q1), (Q2), and (Q3) of Exercise B33. This concept does not require any assumption on F, but in the case when $\frac{1}{2} \in F$, show that it coincides with the linear definition given in the text.
35. Show that a subspace I of a Jordan algebra J is a Jordan ideal of J, iff $U_r a \in I$ and $U_a r \in I$ for any $a \in I$ and $r \in J$. (When J is special this is easy.) This exercise motivates dividing the definition of Jordan ideal into **inner ideal** and **outer ideal**, respectively, where $U_a r \in J$ for all $a \in I$, $r \in J$ (resp. all $a \in J$, $r \in I$). Inner Jordan ideals play an important role in the Jordan structure theory.
36. Define an automorphism σ in $J(V, Q)$ by $\sigma(\alpha + v) = \alpha - v$, and the **norm form** $N(a) = a\sigma(a) = \alpha^2 - Q(v)$, where $a = \alpha + v$. Show that N is a quadratic form on $J(V, Q)$ satisfying $N(ab) = N(a)N(b)$.
37. (Glennie) Write $\{xyz\}$ for $xyz + zyx$. Verify the identity
$$\{xyx\} \circ \{y\{zy^2z\}x\} - \{yzy\} \circ \{x\{xy^2z\}y\}$$
$$= \{x\{z\{x\{yzy\}y\}z\}x\} - \{y\{z\{y\{xzx\}x\}z\}y\}$$
in any special Jordan algebra, a straightforward but tedious computation. However, substituting $x = e_{12} + e_{21}$, $y = e_{23} + e_{32}$, and $z = ae_{21} + a^*e_{12} + be_{13} + b^*e_{31} + ce_{32} + c^*e_{23}$ in $M_3(A)$, one gets equality iff $(ab)c = a(bc)$. Hence, the Glennie identity does not hold for Albert algebras. Conclude that the Albert algebra is not special.

Herstein's theorems on Jordan structure

38. (Herstein) Verify the following assertions where $(R, *)$ is associative of characteristic $\neq 2$ with involution $(*)$, K is the Lie subalgebra of skew-symmetric elements, $J = \mathcal{S}(R, *)$ is the Jordan algebra of symmetric elements, I is a Jordan ideal of J, and $a \in I$:
(i) If $r, s \in J$, then $ras + sar \in I$.
(ii) If $r \in J$, $s \in K$, then $ra^2s - sa^2r \in I$.
(iii) For all $r, s \in J$, $ra^4s + sa^4r \in I$.
(iv) $ra^4s + s^*a^4r^* \in I$, $\forall r, s \in R$.
(Hint: (i) is obtained by linearizing Exercise B31. (ii) is a computation, which starts by noting that $[a, s] \in J$, so $[a^2, s] \in I$. For (iii), note by (ii) that $a^2s - sa^2 \in I$, so $(a^2s - sa^2)^2 \in I$. But $s^2, a^2, sa^2s \in J$, so this and (i) yields $sa^4s \in I$; linearize. To obtain (iv), recall that $R = J \oplus K$ and apply (i),(ii), and (iii).)
39. (Herstein) Prove that $J = \mathcal{S}(R, *)$ is Jordan simple for any simple associative algebra with involution $(R, *)$ of characteristic $\neq 2$. (Hint: The assertion follows from Exercise B38(iv) unless $a^4 = 0$ for all $a \in I$. Thus, $0 = (ra^2 + a^2r^*)^4$, implying $0 = ra^2(ra^2 + e^2r^*)^4 = (ra^2)^5$. Hence, $Ra^2 = 0$ by Exercise 15.14; i.e., $a^2 = 0$, $\forall a \in I$. Linearize and apply Exercise B38 to get $aJa = 0$. Hence, $arara = 0$ for all $r \in R$. Conclude that $a = 0$ by Exercise 15.14.)

The Lie algebra of a Jordan algebra

40. Show that the Lie algebra of skew elements of the Jordan algebra $J(V, Q)$ (cf. Example 21B.23) is precisely the set of linear transformations of $J(V, Q)$ that are skew with respect to the bilinear form of the norm form N of Exercise B36.
41. A **generalized Jordan triple system** over a field is a vector space J together with a product $J \times J \times J \to J$, denoted $(a, b, c) \mapsto \{abc\}$, satisfying the identity
$$\{x_1x_2\{x_3x_4x_5\}\}$$
$$= \{\{x_1x_2x_3\}x_4x_5\} - \{x_3\{x_2x_1x_4\}x_5\} + \{x_3x_4\{x_1x_2x_5\}\}.$$
This product is called a **Jordan triple product** if it also satisfies the identity $\{x_1x_2x_3\} = \{x_3x_2x_1\}$. Verify that the set of $m \times n$ matrices over a field F satisfies the Jordan triple product $\{abc\} = a\,b^t\,c + c\,b^t\,a$. In any linear Jordan algebra with multiplication $\circ$, one has the Jordan

triple product $\{abc\} = (a \circ b) \cdot c + (b \circ c) \cdot a - (a \circ c) \circ b$. Kantor [**Kan**] discovered a key functor taking Jordan triple systems to $\mathbb{Z}$-graded Lie algebras with involution, thereby providing an alternate description of the connection between Lie and Jordan algebras.

Growth in nonassociative algebras

42. Define the growth function of a finitely generated nonassociative algebra in analogy to Definition 17.32, but now taking into account placement of parentheses. How many words are there of length $\leq k$ in the free nonassociative algebra? What is its growth function? What is the growth rate of the Virasoro algebra of Exercise 21.78?

Appendix 21C

1. Prove the PBW theorem by considering $V = \oplus U_k$, where U_k formally has base (21C.1) as a Lie module. (Hint: One needs to define av for $a \in L$ and $v \in V$. Take a to be some base element b_j of L, and $v = b_{i_1} \cdots b_{i_k} \in U_k$; write $v = b_{i_1} v'$. Now define $b_j v \in U_{k+1}$ to be formally $b_j b_{i_1} \cdots b_{i_k}$ if $j < i_1$, and inductively $b_{i_1}(b_j v') + [b_j b_{i-1}]v'$ otherwise. Verify that this is a Lie module action and thus yields a faithful Lie representation $L \to (\operatorname{End} V)^-$.)

Ring-theoretic properties of enveloping algebras

2. For any Lie algebra L over a field F, show that its universal enveloping algebra $U(L)$ is a domain whose only invertible elements are in F, and the Jacobson radical of any subring of $U(L)$ is 0. Conclude for any Lie ideal I of L that $IU(L)$ is a prime ideal of $U(L)$. (Hint: Check the leading terms of the filtration. For the last assertion, note that $U(L)/U(IL) \cong U(L/IL)$.)
3. Prove that the enveloping algebra of any f.d. Lie algebra is Noetherian. (Hint: The associated graded algebra is commutative affine.)
4. Prove that $U(L)$ is an Ore domain for any Lie algebra L of subexponential growth.
5. Show that an algebra R is almost commutative in the sense of Exercise 17.33, iff it is the homomorphic image of a universal enveloping algebra of a Lie algebra. (Hint: Define a Lie algebra structure on R_1 by noting that if $a, b \in R_1$ then $[a, b] \in R_1$.)
6. (Suggested by L. Small.) Notation as in Exercise 21.77, show that $U(\mathfrak{d})^+$ has an infinite descending chain of prime ideals. Is $U(\mathfrak{d})$ Noetherian? Is $U(\mathfrak{d})^+$ Noetherian?
7. Prove that if I_1 and I_2 are ideals of finite codimension in $U = U(L)$ with L finite-dimensional, then $I_1 I_2$ is also of finite codimension.

8. For any restricted Lie algebra L (over a field F of characteristic p, cf. Exercise 21.30), show that $a^p - a^{[p]} \in \operatorname{Cent}(U(L))$ for every $a \in L$; these elements generate an ideal A such that $\dim_F U(L)/A = p^n$, where $n = \dim_F L$. Show that this algebra, called the **restricted universal enveloping algebra**, is the universal with respect to representations of L as a restricted Lie algebra.

   On the other hand, for any base $\{b_1, \ldots, b_n\}$ of L over F, show that $U(L)$ is a free module of rank p^n over its central subalgebra $F[b_1 - b_1^{[p]}, \ldots, b_n - b_n^{[p]}]$.

Faithful Lie representations

In these exercises, the enveloping algebra $U(L)$ is used to prove that every f.d. Lie algebra is linear.

9. Show that $\ker(\rho \oplus \tau) = \ker \rho \cap \ker \tau$ for any Lie representations ρ, τ.
10. Show that any Abelian Lie algebra L of dimension n has a faithful representation of degree $n + 1$, in which the image of every element of L is nilpotent.
11. Suppose L is a finite-dimensional solvable Lie algebra over a field of characteristic 0, and N is the largest nilpotent Lie ideal of L; cf. Exercise 21.33. Suppose I is an ideal of finite codimension in $U(L)$ for which the image of N in $U(L)/I$ is nil. Show that I contains an ideal J having the same properties, invariant under every derivation of N. (Hint: Let $\tilde{I}$ be the ideal generated by I and N. Then $\tilde{I}/I$ is nilpotent, so there is m such that $\tilde{I}^m \subseteq I$. Take $J = \tilde{I}^m$, which has finite codimension by Exercise C7. Conclude using Exercise 21.40.)
12. Suppose $L = S \oplus L_1$ is a f.d. Lie algebra of characteristic 0, where $S \lhd L$ is solvable. Let N be the largest nilpotent Lie ideal of L. If there is a representation $\rho\colon S \to gl(n, F)$ for which the image of N is nil, show that there is a representation τ on L whose restriction to S has the same kernel as ρ and such that $\tau(a + b)$ is nilpotent for every $a \in N$ and $b \in L_1$ for which $\operatorname{ad}_S(b)$ is nilpotent. (Hint: ρ induces an algebra homomorphism $U(S) \to M_n(F)$ with kernel I of codimension $\leq n^2$. Apply Exercise C11 to get J of finite codimension. $U(S)/J$ is a Lie module under the adjoint action of L; apply Engel's Theorem.)
13. (Ado's Theorem.) Prove that any f.d. Lie algebra L of characteristic 0 is linear. (Hint: Since the kernel of the adjoint representation is the center Z, it suffices by Exercise C9 to find a representation of L that is faithful on Z. Let N be the largest nilpotent ideal of L; clearly $Z \subseteq N$. Take a chain of ideals descending from N to Z, each of codimension 1 in the previous ideal. These can be described in terms of upper triangular matrices. By Exercise C10, Z has a faithful representation of finite degree, so N also has, by Exercise C12. Next take a chain

of ideals of codimension 1 from $\operatorname{rad} L$ to N. Finally, appeal to Levi's Theorem.)

Lie Representations and enveloping algebras

14. Show that a Lie representation of L of degree n is irreducible iff the corresponding map $U(L) \to gl(n, F)$ is onto.
15. For any Lie representation ρ of a semisimple Lie algebra L, one can view $L = \ker\rho \times \rho(L)$ and thus view $\rho(L) \subseteq L$. From this point of view, construct the Casimir element of $\rho(L)$ inside $U(L)$, and show that $\rho(L)$ lies in the augmentation ideal of $U(L)$.
16. Recall that a derivation of a Lie algebra L to a Lie module M is a linear map $\delta\colon L \to M$ satisfying $\delta([ab]) = a\delta(b) - b\delta(a)$, $\forall a, b \in L$. Taking $J = U(L)L$, the augmentation ideal of Remark 21C.2′, show that $\operatorname{Hom}(J, M) \cong \operatorname{Der}(L, M)$ as vector spaces. (Hint: Define the map $\Phi\colon \operatorname{Hom}(J, M) \to \operatorname{Der}(L, M)$ by restricting the domain of definition. For the other direction, given $\delta\colon L \to M$, define the map $\psi_\delta\colon U(L) \otimes_F L \to M$ by $u \otimes a \mapsto u\delta(a)$. Then $\ker\psi_\delta$ contains the kernel of the natural map $U(L) \otimes L \to U(L)L = J$, and thus ψ_δ yields a map from J to M; hence, one has the desired correspondence $\operatorname{Der}(L, M) \to \operatorname{Hom}(J, M)$.)

Quantized enveloping algebras

These exercises are culled mainly from [**BrokaG, Part I**]. In Exercises C17–C20, $F = K(q)$ where q is a commuting indeterminate.

17. Display $U_q(sl(2, F))$ as a skew polynomial ring and conclude that it is a Noetherian domain. (Hint: Define $C = F[q, q^{-1}, k, k^{-1}]$, which has an automorphism σ given by $\sigma(k) = q^{-2}k$. Then let $R = C[\lambda; \sigma]$, which has an automorphism τ fixing λ with $\tau(k) = q^2k$, and a derivation δ with $\delta(k) = 0$ and $\delta(\lambda) = \frac{\lambda^{-1} - \lambda}{q - q^{-1}}$. The isomorphism $U_q(sl(2, F)) \to R[\mu; \tau, \delta]$ is given by $e \mapsto \lambda$ and $f \mapsto \mu$.)
18. Define $[n] = \frac{q^n - q^{-n}}{q - q^{-1}}$. Show that $[-n] = -[n]$ and $[n] \in \mathbb{Z}[q, q^{-1}]$.
19. Define the q-**factorial** $n!_q = [n][n-1]\cdots[1]$ and the q-**binomial coefficient** $\left[\begin{smallmatrix} m \\ n \end{smallmatrix}\right]_q$ as $\frac{m!_q}{n!_q(m-n)!_q}$. Show that

$$\begin{bmatrix} m+1 \\ n \end{bmatrix}_q = q^{-n}\begin{bmatrix} m \\ n \end{bmatrix}_q + q^{m-n+1}\begin{bmatrix} m \\ n-1 \end{bmatrix}_q.$$

Conclude that $\left[\begin{smallmatrix} m \\ n \end{smallmatrix}\right]_q \in \mathbb{Z}[q, q^{-1}]$ for all integers $m \geq n \in \mathbb{N}$.

20. Define $[k; n] = \frac{kq^n - k^{-1}q^{-n}}{q - q^{-1}}$. Show in $U = U_q(sl(2, F))$ that $[k; n]e = e[k; n+2]$, $[k; n]f = f[k; n-2]$, and $f^n e = ef^n - [n]f^{n-1}[k; n-1]$. Use these calculations to reprove that U is a domain, and show that its center is generated over F by the element

$$ef + \frac{kq^{-1} + k^{-1}q}{(q - q^{-1})^2}.$$

21. For F algebraically closed, suppose the semisimple Lie algebra L has a Cartan subalgebra $\mathfrak{h}$, a simple root system $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, and Cartan matrix (a_{ij}). Let $q_i = q^{\frac{1}{2}\langle \mathbf{a}_i, \mathbf{a}_i\rangle}$, $1 \leq i \leq n$. The **quantized enveloping algebra** $U_q(L)$ is defined to be the F-algebra with generators e_i, f_i, k_i, k_i^{-1}, $1 \leq i \leq n$, satisfying the following relations:

$$k_i e_j k_i^{-1} = q_i^{a_{ij}} e_j, \quad k_i f_j k_i^{-1} = q_i^{-a_{ij}} f_j, \quad [k_i, k_j] = 0,$$

$$[e_i, f_j] = \delta_{ij}\frac{k_i - k_i^{-1}}{q_i - q_i^{-1}}, \quad \sum_{\ell=0}^{1-a_{ij}} (-1)^\ell \begin{bmatrix} 1-a_{ij} \\ \ell \end{bmatrix}_{q^i} e_i^{1-a_{ij}-\ell} e_j e_i^\ell = 0,$$

$$\sum_{\ell=0}^{1-a_{ij}} (-1)^\ell \begin{bmatrix} 1-a_{ij} \\ \ell \end{bmatrix}_{q^i} f_i^{1-a_{ij}-\ell} f_j f_i^\ell = 0$$

for all $i \neq j$. Obtain a PBW-type basis for $U_q(L)$ and use this to build a filtration whereby the associated graded algebra is Noetherian. Conclude that $U_q(L)$ is a Noetherian domain.

22. Show that the enveloping algebra of a composition algebra is a Clifford algebra.

Chapter 22

Dynkin diagrams – Existence results

Verify Exercises 1–4, which provide the root systems and Weyl groups for the Lie algebras A_n, $B_n (n \geq 2)$, $C_n (n > 2)$, and $D_n (n > 2)$ respectively; cf. Example 21.6. (Define the **signed permutation group** acting on $\{\pm\hat{e}_1, \ldots, \pm\hat{e}_n\}$ to be the set of permutations that can also change the sign; thus, its order is $2^n n!$)

1. A_n: Here $k = n + 1$. $sl(n + 1, F)$ has a base consisting of

$$\{h_i = e_{ii} - e_{i+1,i+1} : 1 \leq i \leq n\} \cup \{e_{ij} : i \neq j\}.$$

Take $\mathfrak{h}$ to be the set of diagonal matrices

$$\Big\{ \operatorname{diag}(\alpha_1, \alpha_2, \ldots, \alpha_{n+1}) : \sum \alpha_i = 0 \Big\} = \ker(\hat{e}_1 + \cdots + \hat{e}_n),$$

where $\hat{e}_i\colon F^{(n+1)} \to F$ is the linear functional sending a vector to its i-th component. This is a Cartan subalgebra since it is Abelian and

is its own normalizer, and has base $\{h_1, \ldots, h_n\}$. The matrix units e_{ij} $(i \neq j)$ are simultaneous $\mathfrak{h}$-eigenvectors with roots $\hat{e}_i - \hat{e}_j$; defining the root to be positive when $i < j$, show that $\{\hat{e}_i - \hat{e}_{i+1} : 1 \leq i \leq n\}$ is a simple root system. The Weyl group is S_{n+1} since it contains all transpositions $(i\ i{+}1)$.

2. B_n $(n \geq 2)$: It is convenient to rewrite the symmetric bilinear form (with respect to which the elements of B_n are antisymmetric) in terms of the matrix $\left(\begin{smallmatrix} 0 & I_n & 0 \\ I_n & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$. Now the diagonal matrices

$$\{\mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n, -\alpha_1, -\alpha_2, \ldots, -\alpha_n, 0) : \alpha_i \in F\}$$

comprise a Cartan subalgebra; the other base elements are of the form $e_{i,j} - e_{j+n,i+n}$ $(i{\neq}j)$, $e_{i,j+n} - e_{j,i+n}$ $(i{<}j)$, $e_{i+n,j} - e_{j+n,i}$ $(i{<}j)$, $e_{i,n+1} - e_{n+1,i+\ell}$, and $e_{i+\ell,n+1} - e_{n+1,i}$, whose respective roots are $\hat{e}_i - \hat{e}_j$, $-\hat{e}_i - \hat{e}_j$, $\hat{e}_i + \hat{e}_j$, $-\hat{e}_i$, and $\hat{e}_i$. One can take the positive roots to be $\hat{e}_i$ and $\hat{e}_i \pm \hat{e}_j$ for $i < j$, a simple root system being

$$\{\hat{e}_1 - \hat{e}_2,\ \ldots,\ \hat{e}_{n-1} - \hat{e}_n,\ \hat{e}_n\}.$$

The Weyl group is the signed permutation group.

3. C_n $(n \geq 3)$: Write the alternating bilinear form (with respect to which the elements of C_n are antisymmetric) in terms of the matrix $\left(\begin{smallmatrix} 0 & I_n \\ -I_n & 0 \end{smallmatrix}\right)$. Again the diagonal matrices

$$\{\mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n, -\alpha_1, -\alpha_2, \ldots, -\alpha_n, 0) : \alpha_i \in F\}$$

comprise a Cartan subalgebra; the other base elements are of the form $e_{i,j} - e_{j+n,i+n}$ $(i{\neq}j)$, $e_{i,j+n} - e_{j,i+n}$ $(i{\leq}j)$, and $e_{i+n,j} + e_{j+n,i}$ $(i{\leq}j)$, whose respective roots are $\hat{e}_i - \hat{e}_j$, $-\hat{e}_i - \hat{e}_j$, $\hat{e}_i + \hat{e}_j$. One can take the positive roots to be $\hat{e}_i$ and $\hat{e}_i \pm \hat{e}_j$ for $i \leq j$, a simple root system being

$$\{\hat{e}_1 - \hat{e}_2,\ \ldots,\ \hat{e}_{n-1} - \hat{e}_n,\ 2\hat{e}_n\}.$$

The Weyl group is again the signed permutation group.

4. D_n $(n \geq 4)$: Write the symmetric bilinear form (with respect to which the elements of D_n are antisymmetric) in terms of the matrix $\left(\begin{smallmatrix} 0 & I_n \\ I_n & 0 \end{smallmatrix}\right)$. The diagonal matrices $\mathrm{diag}\{a_1, -a_1, a_2, -a_2, \ldots, a_n, -a_n\}$ comprise a Cartan subalgebra. The roots are $\pm\hat{e}_i \pm \hat{e}_j$. One can take the positive roots to be $\hat{e}_i \pm \hat{e}_j$ for $i < j$, a simple root system being $\{\hat{e}_1 - \hat{e}_2, \ldots, \hat{e}_{n-1} - \hat{e}_n\}$. The Weyl group is the subgroup of index 2 of the signed permutation group consisting of those signed permutations that have an even number of sign changes.

5. Given a root system Φ, define the **coroot** of $\mathbf{a}$ to be $\frac{2\mathbf{a}}{\langle \mathbf{a}, \mathbf{a} \rangle}$. The set of coroots also comprises a root system, called the **dual root system** to Φ, and it has the same Weyl group as Φ. Show that the root systems of B_n and C_n are dual.

Constructions of exceptional Lie algebras

6. The multiplication algebra of an alternative algebra satisfies the identity $[\ell_a r_b] = [r_a \ell_b]$. For any a, b, show that $\delta_{a,b} = [\ell_a \ell_b] + [\ell_a r_b] + [r_a r_b]$ is a derivation, called the **inner derivation** with respect to a and b.
7. In a composition algebra $\mathcal{A}$, show that the (trace) bilinear form $\langle a, b \rangle = T(ab) = ab + b^* a^*$ is associative. Let $\mathcal{A}_0 = 1^\perp$, the elements of trace 0. Define a binary operation $\cdot$ on $\mathcal{A}_0$ by $a \cdot b = ab - \frac{1}{2}T(ab)$. Show that $\langle a \cdot b, c \rangle = \langle b \cdot c, a \rangle$ and $a \cdot b = -b \cdot a$, $\forall a, b, c \in \mathcal{A}_0$. Also $\delta_{a,b} = -\delta_{b,a}$, cf. Exercise 6, and

$$\delta_{a \cdot b, c} + \delta_{b \cdot c, a} + \delta_{c \cdot a, b} = 0.$$

Furthermore, $\delta(\mathcal{A}_0) \subseteq \mathcal{A}_0$ and $\langle \delta a, b \rangle = -\langle a, \delta b \rangle$ for any $\delta \in \mathrm{Der}\,\mathcal{A}$.

8. Let J be a Jordan algebra of degree 3, and again using the trace map, define $a \cdot b = a \circ b - \frac{1}{3}\,\mathrm{tr}(ab)$.
9. Complete the definition of Lie multiplication of L following Example 22.2 by putting

$$(a \otimes v)(b \otimes w) = \frac{1}{12}T(vw)\delta_{a,b} + (a \cdot b) \otimes (v \cdot w) + \frac{1}{2}T(ab)[\ell_v \ell_w] \in L$$

for $a, b \in \mathcal{A}$, $v, w \in J$. Prove that L is a Lie algebra by piecing together Exercises 7 and 8 to verify the Jacobi identity. (This example has recent, far-reaching generalizations.)

Generalized Cartan matrices revisited

10. For any generalized Cartan matrix A of affine type, show that any proper subdiagram of its Dynkin diagram is the disjoint union of Dynkin diagrams of simple f.d. Lie algebras.

Abstract root systems of a reflection group (following Humphreys [Hum3])

In these exercises, $\mathcal{W}$ denotes a reflection group.

11. Show that an element $\sigma \in O(V)$ is a hyperplane reflection iff $\mathrm{rank}(1-\sigma) = 1$, $\sigma^2 = 1$, and $V = \ker(1-\sigma) \perp (1-\sigma)(V)$. (This leads one to define a **reflection** of an arbitrary vector space V to be an element of $\sigma \in GL(V)$ satisfying $\mathrm{rank}(1-\sigma) = 1$ and $\sigma^2 = 1$, but we do not pursue this further generality.)
12. Show that the only possible root system of rank 1 is $\{\pm\mathbf{a}\}$, depicted geometrically as $\underset{-\mathbf{a}}{\longleftarrow} \cdot \underset{\mathbf{a}}{\longrightarrow}$. For rank 2, there are four possible root systems, up to a suitable transformation by an element of $\mathcal{W}$.
13. For any $w \in \mathcal{W}$ show that $\det w = (-1)^m$, where m is the number of positive roots sent to negative roots by w,

14. (Compare with Exercises 1–4.) Let $V = \mathbb{R}^n$ with the standard base $e_1, \dots, e_n$.

    (i) S_n acts as the group of permutations of the base. The transposition (ij) corresponds to $\sigma_{e_i - e_j}$. Since the $n-1$ transpositions $\{(i\,i+1), 1 \le i \le n-1\}$ generate S_n (and also in view of Exercise 1), S_n is called A_{n-1} as a Coxeter group (identified with the Weyl group of the Lie algebra A_{n-1}).

    (ii) The signed permutation group (the Weyl group of B_n and C_n) is generated by $\{\sigma_{e_i},\ 1 \le i \le n\} \cup \{\sigma_{e_i - e_j} : 1 \le i < j \le n\}$.

    (iii) The even signed permutation group (the Weyl group of D_n) is generated by the $\sigma_{e_i + e_j}$ and the $\sigma_{e_i - e_j}$.
15. Show that $\sigma_{w\mathbf{a}} \in \mathcal{W}$ for any $w \in \mathcal{W}$ and $\sigma_{\mathbf{a}} \in \mathcal{W}$. (Hint: Lemma 22.15.)
16. Prove that $\mathcal{W}$ is generated by the reflections with respect to the simple roots. (Hint: Let S denote the simple roots, and let $\mathcal{W}_0$ denote the subgroup generated by the reflections $\{\sigma_{\mathbf{a}} : \mathbf{a} \in S\}$. It suffices to show that any root $\mathbf{b}$ is in $\mathcal{W}_0 S$, since then writing $\mathbf{b} = w\mathbf{a}$ one has $\sigma_{\mathbf{b}} = w\sigma_{\mathbf{a}} w^{-1} \in \mathcal{W}_0$. If $\mathbf{b} \in P$, take an element $\mathbf{a} \in \mathcal{W}_0\mathbf{b} \cap P$ of smallest height; as in Corollary 21.100, this is simple, and $\mathbf{b} \in \mathcal{W}_0\mathbf{a}$, as desired. If $\mathbf{b} \in -P$, one gets instead $-\mathbf{b} \in \mathcal{W}_0\mathbf{a}$, so $\mathbf{b} \in \mathcal{W}_0\sigma_{\mathbf{a}}\mathbf{a}$.)

    To improve Exercise 16, one needs the following technical result. Define $n(w) = |wP \cap -P|$, the number of positive roots sent to negative roots by $w \in \mathcal{W}$.
17. Suppose w is a product $\sigma_1 \cdots \sigma_t$ of reflections $\sigma_i = \sigma_{\mathbf{a}_i}$ for suitable simple roots $\mathbf{a}_i$, repetition permitted, with $n(w) < t$. Show that there are indices $i < j$ satisfying $\mathbf{a}_i = \sigma_{i+1} \cdots \sigma_{j-1}\mathbf{a}_j$, $\sigma_{i+1} \cdots \sigma_j = \sigma_i \cdots \sigma_{j-1}$, and
$$w = \sigma_1 \cdots \sigma_{i-1}\sigma_{i+1} \cdots \sigma_{j-1}\sigma_{j+1} \cdots \sigma_t.$$
    (Hint: By Exercise 15, $n(w\sigma_{\mathbf{a}}) = n(w) + 1$ whenever $w\mathbf{a} > 0$, so some $\sigma_1 \cdots \sigma_{j-1}\mathbf{a}_j < 0$. Then some σ_i switches the sign of $\sigma_{i+1} \cdots \sigma_{j-1}\mathbf{a}_j$, but the only new negative root can come from $\mathbf{a}_i$. This yields the first assertion, and the rest follows from applying Lemma 22.15.)
18. Show that $n(w)$ is the minimal number of reflections t in Exercise 17.
19. Prove that any finite reflection group is Coxeter. More precisely, defining $m(\mathbf{a}, \mathbf{b})$ to be the order of $\sigma_{\mathbf{a}}\sigma_{\mathbf{b}}$, prove that $\mathcal{W}$ is isomorphic to the group generated by the $\{\sigma_{\mathbf{a}} : \mathbf{a} \in S\}$ together with the relations $(\sigma_{\mathbf{a}}\sigma_{\mathbf{b}})^{m(\mathbf{a},\mathbf{b})}$; i.e., all other relations in $\mathcal{W}$ are superfluous. (Hint: One wants to reduce any relation $\sigma_1 \cdots \sigma_t = 1$ to one of the given relations. Note by taking determinants that $t = 2u$ for some u. Then $\sigma_1 \cdots \sigma_u = \sigma_{u+1} \cdots \sigma_t$, since each $\sigma_i^2 = 1$. Exercise 17 can be applied to the LHS, yielding some $\sigma_{i+1} \cdots \sigma_j = \sigma_i \cdots \sigma_{j-1}$, so $1 = \sigma_i \cdots \sigma_{j-1}\sigma_j \cdots \sigma_{i+1}$. If this relation has length $< t$, then apply induction and substitute back. So assume that $i = 1$, $j = u + 1$, and $\sigma_1 \cdots \sigma_u = \sigma_2 \cdots \sigma_{u+1}$. Hence, $\sigma_2\sigma_1 \cdots \sigma_u\sigma_{u+1} = \sigma_3 \cdots \sigma_u$, and applying the same sort of argument shows that $\sigma_2\sigma_1\sigma_2 \cdots \sigma_{u-1} = \sigma_1 \cdots \sigma_u$.

    But any relation can be permuted cyclically, so $\sigma_2 \cdots \sigma_t\sigma_1 = 1$, and the same argument shows that $\sigma_3\sigma_2\sigma_3 \cdots \sigma_u = \sigma_2 \cdots \sigma_{u+1}$. Conclude that $\sigma_1 = \sigma_3$, and thus $\sigma_1 = \sigma_3 = \sigma_5 = \cdots$ and $\sigma_2 = \sigma_4 = \cdots$, so the original relation is $(\sigma_1\sigma_2)^u = 1$.)

20. Show that any two positive systems Φ_1 and Φ_2 are conjugate under some element of the Weyl group, which also conjugates their simple systems. (Hint: Induction on $t = |\Phi_1 \cap -\Phi_2|$. Take a simple root $\mathbf{a} \in \Phi_1 \cap -\Phi_2$, which exists since otherwise $\Phi_1 \subseteq \Phi_2$, yielding equality. Then $|\sigma_{\mathbf{a}}(\Phi_1) \cap -\Phi_2| = t - 1$.)
21. For simple roots $\mathbf{a}_i \ne \mathbf{a}_j$, show that $\sigma_i\sigma_j$ acts on the plane spanned by $\mathbf{a}_i$ and $\mathbf{a}_j$, as rotation through the angle $\theta(\mathbf{a}_i, \mathbf{a}_j) = \arccos \frac{\pi}{m(\mathbf{a}_i, \mathbf{a}_j)}$.
22. Prove that a root system of V is crystallographic iff its Weyl group stabilizes a sublattice of V, i.e., a $\mathbb{Z}$-submodule spanned by a suitable base. (Hint: The base is the positive root system.)

    In the case when the condition of Exercise 22 holds, we also say that $\mathcal{W}$ is a **crystallographic** group.
23. For any crystallographic group $\mathcal{W}$, show that each $m_{i,j} \in \{2, 3, 4, 6\}$. (Hint: As in Proposition 21.101.) In this case, in Exercise 21, $4\cos^2\theta(\mathbf{a}_i, \mathbf{a}_j) = 0, 1, 2, 3$, respectively, which is $m_{i,j} - 2$ in the first three cases. In this way, one sees that the labels on the edges often are 2 more in the Coxeter graph than in the analogous Dynkin diagram.
24. The Coxeter graph of the dihedral group of order $2m$ has the form $\bullet \overset{m}{\text{---}} \bullet$. For $m > 4$, this graph, denoted $I_2(m)$, does not appear as the Dynkin diagram of a semisimple Lie algebra.

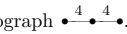
Coxeter groups

25. When the Coxeter graph $\mathcal{G}$ of a Coxeter group $\mathcal{W}$ is connected, show that, in the notation of Definition 22.20, any proper $\mathcal{W}$-submodule V_1 of V is contained in $\operatorname{rad} V$. (Hint: If some $e_i \in V_1$, then applying reflections by the neighboring vertices of $\mathcal{G}$ shows that every $e_i \in V_1$, contrary to $V \ne V_1$. Thus, for each i, $e_i \notin V_1$, so the only eigenvalue of σ_i on V_1 is 1, implying σ_i fixes V_1. In other words, $\langle e_i, V_1 \rangle = 0$ for each i, implying $V_1 \subseteq \operatorname{rad} V$.)
26. Prove that the bilinear form of any finite Coxeter group $\mathcal{W}$ is positive definite. More explicitly, $V = \mathbb{R}^{(n)}$ viewed as a $\mathcal{W}$-module via Theorem 22.22(i) is simple, with $\operatorname{End}_{\mathbb{R}[\mathcal{W}]} V = \mathbb{R}$. (Hint: One may assume that the Coxeter graph $\mathcal{G}$ is connected. V is completely reducible as

a $\mathcal{W}$-module by Maschke's Theorem, implying that $\operatorname{rad} V = 0$. Now Exercise 21 shows that V is a simple $\mathcal{W}$-module.

Any $f \in \operatorname{End}_{\mathbb{R}[\mathcal{W}]} V$ leaves any $\mathbb{R}e_i$ invariant, so $f(e_i) = \alpha e_i$ for some $\alpha \in \mathbb{R}$, implying $\ker(f - \alpha)$ is a nonzero $\mathcal{W}$-submodule of V, and thus must be V, i.e., $f = \alpha I$. This proves that $\operatorname{End}_{\mathbb{R}[\mathcal{W}]} V = \mathbb{R}$.)

Classification of Coxeter graphs arising from Coxeter groups with positive definite forms

27. A **subgraph** of a Coxeter graph is a Coxeter graph obtained by erasing vertices and edges. Show that if the bilinear form of a Coxeter graph is positive definite, then the bilinear form of any subgraph must remain positive definite.
28. Show that when the Coxeter graph contains a cycle, then the bilinear form has an isotropic vector. (Hint: The matrix is singular since the sum of the rows is 0.)
29. Prove that the Coxeter graph of a reflection group cannot contain the subgraph  . Furthermore, if the graph contains the subgraph  , then no other vertex can have degree > 2.
30. Show that the following Coxeter graphs have positive nondefinite bilinear forms:

31. Given a Coxeter system $(\mathcal{W}, S)$, define $\ell(w)$ to be the minimal length of w as a product of elements of S. Show that there is a homomorphism $\mathcal{W} \to \{\pm 1\}$ given by $w \mapsto (-1)^{\ell(w)}$. In particular, one cannot write w as two products of different parity. (Hint: Check this for the elements of S, and show that all the relations are in the kernel.)
32. Given a Coxeter system $(\mathcal{W}, S)$ and a subset $T \subset S$, let $\mathcal{W}_T$ denote the subgroup of $\mathcal{W}$ generated by T; if $w \in \mathcal{W}_T$, define $\ell_T(w)$ to be the length of w in $\mathcal{W}_T$. Show that $\ell(w) \leq \ell_T(w)$.

Root systems for Coxeter groups

Using the notation of Definition 22.20, define the **roots** $\Phi = \{we_i : w \in \mathcal{W},\ 1 \leq i \leq n\} \subset V$. A root $\mathbf{a}$ is **positive** if each coefficient of $\mathbf{a}$ (written as a linear combination of e_i) is non-negative.

33. For $w \in \mathcal{W}$ and $s_i \in S$, show that $we_i > 0$ when $\ell(ws_i) > \ell(w)$; likewise for $<$. (Hint: Pick $s_j \in S$ such that $\ell(ws_j) = \ell(w) - 1$, and take $T = \{s_i, s_j\}$. $\mathcal{W}_T$ is dihedral.

Let $A = \{w' \in \mathcal{W} : w = w'w_T$ with $w_T \in \mathcal{W}_T$ and $\ell(w) = \ell(w') + \ell(w_T)\}$. Pick $w' \in A$ of minimal possible length. Note that $ws_j \in A$, so $\ell(w') \leq \ell(w) - 1$. If $\ell(w's_i) < \ell(w')$, then $w's_i \in A$, contrary to the choice of w'. Hence, $\ell(w's_i) > \ell(w')$, implying that $w'e_i > 0$ by induction. Likewise $\ell(w's_j) > \ell(w')$, so $w'e_j > 0$. But w_Te_i is a non-negative linear combination of e_i and e_j.)

34. Prove that every root is either positive or negative.
35. Show that the representation ρ of Theorem 22.22 is faithful. (Hint: Otherwise, take $w \neq 1$ in $\ker \rho$ and s_i with $\ell(ws_i) < \ell(w)$. Then $e_i = we_i < 0$, absurd.)
36. Prove that every finite Coxeter group is a reflection group. (Hint: By Exercise 35, it is embedded in $\operatorname{End} V$, generated by reflections, and the associated bilinear form is positive definite by Exercise 26.)
37. Show that $(\mathcal{W}_T, T)$ is a Coxeter system whose relations are those relations of $(\mathcal{W}, S)$ that involve only elements of T. (Hint: Form the desired Coxeter group $\tilde{\mathcal{W}}_T = (\mathcal{W}_T, T)$, taking only those relations $(s_is_j)^{m(s_i,s_j)}$ for $s_i, s_j \in T$. Then letting $n' = |T|$, we have the standard representation $\rho_T \colon \tilde{\mathcal{W}}_t \to \mathrm{GL}(n', F)$ that factors through $\mathcal{W}_T$. But ρ_T is an injection, by Exercise 35.)

Part VI

Representable Algebras

Introduction to Representable Algebras

Part V dealt with the theory of representations of various algebraic structures into matrix algebras $M_n(\bar{F})$ over an algebraically closed field. This leads us to consider subrings of matrix algebras (over a larger field); such rings are called *representable*. Since such rings are often algebras over a subfield F of $\bar{F}$ that is not algebraically closed, we encounter some of the intricacies involved in dealing with non-algebraically closed fields.

In Chapter 23 we view the general question of representability of a ring in terms of a simple and natural condition called a *polynomial identity*, which also provides a beautiful structure theory.

The most important class of representable rings is that of f.d. division algebras, which have arisen already in the Wedderburn-Artin Theorem; these are studied in Chapter 24.

Viewing representations of algebraic structures from a categorical point of view, our target algebra, $M_n(\bar{F})$, the endomorphism algebra of a f.g. free module, should be replaced by a structure defined in categorical language, leading us in Chapter 25 to the notion of **projective** (and **injective**) module, which provides the setting for the powerful theories of homology and cohomology that underlie many of our earlier theorems. In Appendix 25A we study equivalences of categories of modules via Morita theory.

In Appendix 25B we introduce **Azumaya algebras**, the natural generalization of central simple algebras; there also is a description of Azumaya algebras in terms of polynomial identities.

Representation theory can be translated to the structure of modules, leading us in Appendix 25B to a remarkable connection with the classification of f.d. Lie algebras.

Various associative algebras (group algebras, enveloping algebras) arise in representation theory, and these are unified in Chapter 26 by means of the theory of Hopf algebras.

Chapter 23

Polynomial Identities and Representable Algebras

We have considered an assortment of algebraic structures that can be best studied as represented in the matrix algebra $M_n(K)$ for an appropriate field K. This leads us to the following definition.

Definition 23.0. A ring R is **representable** if there is an injection of R into $M_n(K)$ for some field K.

In the literature, one sometimes only requires K to be commutative Noetherian. But when R is an affine algebra (cf. Definition 17.1), these definitions are equivalent; cf. Exercise 4.

In the definition, we can always replace K by a field $\hat{K}$ containing K, since

$$M_n(K) \hookrightarrow M_n(\hat{K});$$

in particular, taking $\hat{K}$ to be the algebraic closure of K, we may assume that K is algebraically closed.

We have already encountered this concept in terms of representations of rings (Definition 13.38). By Example 13.49, any f.d. algebra over a field F is representable. This example is almost too easy. Also, any integral domain can be represented into its field of fractions.

It is rather easy to develop a rich theory of representable affine algebras, with hardly any background other than the commutative theory. In Exercise 6, we see that the Jacobson radical of a representable affine algebra is

nilpotent. This draws us to the class of all representable algebras, say over a given field F.

Of course, any subalgebra of a representable algebra is representable. Unfortunately, as we shall see, a homomorphic image of a representable algebra need not be representable. This leads us to search for an easily defined class of algebras encompassing all representable algebras, but which is closed with respect to taking homomorphic images as well as subalgebras. (Although f.d. algebras over a field are closed under homomorphic images and subalgebras, they miss certain very important examples of representable rings, such as $\mathbb{Z}$ and $F[\lambda]$.)

To help us in our search, we turn to some ideas from mathematical logic. An $\forall$-**sentence** is an elementary sentence whose only quantifiers are $\forall$. (We do not permit $\neg\forall$.) For example, a domain is a ring satisfying the following $\forall$-sentence:

$$\forall x, y \in R,\ x = 0 \ \vee\ y = 0 \ \vee\ xy \neq 0.$$

Clearly, any $\forall$-sentence holding in an algebra also holds in all of its subalgebras; in particular, each representable algebra R satisfies every $\forall$ sentence holding in $M_n(K)$. Unfortunately, some of these sentences are rather exotic and difficult to analyze.

Fortunately, one particular sort of $\forall$-sentence, called a **polynomial identity** (PI for short), is described easily and provides most of the structure of representable algebras. So in this chapter we study algebras satisfying a PI, focusing on (associative) representable algebras. In Appendix 23A, we discuss Kemer's Theorem, one of the highlights of associative PI theory.

We have already seen nonassociative versions of polynomial identities, while studying identities in Lie algebras, alternative algebras, and Jordan algebras. In Appendix 23B, we describe the general theory of identities and see how this leads to Zelmanov's striking solution of the Restricted Burnside Problem.

All representable algebras are PI, as we see presently. On the other hand, affine PI-algebras need not be representable. Indeed, it is not difficult to construct countable affine PI-algebras over $\mathbb{Q}$ with uncountably many nonisomorphic homomorphic images. Only countably many could be embedded in matrices over a f.g. field extension of $\mathbb{Q}$. But any representable affine $\mathbb{Q}$-algebra clearly can be embedded in matrices over the f.g. field extension of $\mathbb{Q}$ obtained by adjoining only the entries of the generators (viewed as matrices). Hence, uncountably many of these nonisomorphic affine PI-algebras are nonrepresentable. (This argument is due to Lewin; see Exercise 8 for the precise formulation.)

Nevertheless, the structure theory of PI-algebras turns out to be almost as rich as that of representable algebras. In this way, PI-theory becomes a very useful tool in representation theory. One pleasant side effect of PI-theory has been to provide a classic application for the lovely structure theory of rings that had been developed in the 1940's and 1950's (although we take a few shortcuts in order to keep the presentation concise).

Our main intuition comes from the observation that the commutativity law could be expressed by saying that the formal expression $x_1x_2 - x_2x_1$ vanishes identically on any commutative algebra. Viewing $x_1x_2 - x_2x_1$ as a noncommutative polynomial, we may wonder what can be said when some other polynomial vanishes for all substitutions in an algebra. For example, Fermat's Little Theorem translates to the fact that for any field F comprised of m elements, $a^m = a$ for all $a \in F$; i.e., $x_1^m - x_1$ vanishes identically on F.

The "correct" place to view noncommutative polynomials for associative algebras is the free (associative) algebra $C\{X\}$ in countably many noncommuting indeterminates $x_1, x_2, \dots$ (cf. Definition 17.5). Remark 17.6 describes the substitution (also called **specialization**) of x_i into given elements of a C-algebra R in terms of a homomorphism $C\{X\} \to R$. Given $f \in C\{X\}$, we write $f(x_1, \dots, x_m)$ to denote that $x_1, \dots, x_m$ are the only indeterminates occurring in f. The **evaluation** $f(r_1, \dots, r_m)$ denotes the image of f under the specialization $x_i \mapsto r_i$ for each $1 \le i \le m$. We write $f(R)$ for the set of evaluations of f in R. More generally, if $S \subseteq R$, we write $f(S)$ for $\{f(s_1, \dots, s_m) : s_i \in S\}$.

Definition 23.1. A **(polynomial) identity** of a C-algebra R is a (noncommutative) polynomial $f(x_1, \dots, x_m) \in C\{X\}$ for which $f(R) = 0$; i.e., $f(r_1, \dots, r_m) = 0$, $\forall r_i \in R$. We write $\mathrm{id}(R)$ for the set of identities of R.

Remark 23.1′. Since every homomorphism $C\{X\} \to R$ is described in terms of specializations of the x_i, we see that f is an identity of R iff $f \in \ker \varphi$ for every homomorphism $\varphi : C\{X\} \to R$.

Remark 23.2. If $f \in \mathrm{id}(R)$, then f is an identity of any homomorphic image of R and also of any subalgebra of R. Furthermore, if $f \in \mathrm{id}(R_i)$ for each $i \in I$, then $f \in \mathrm{id}(\prod_{i\in I} R_i)$. (The last assertion holds since

$$f\big((x_{1i}), (x_{2i}), \dots, (x_{mi})\big) = \big(f(x_{1i}, \dots, x_{mi})\big) = (0) = 0.)$$

Remark 23.2 could be used as an alternate approach to identities; cf. Remark 23.42 and Exercise 29.

PI-theory has two main strands: the structure theory of a given PI-algebra R and the combinatorics of the set $\mathrm{id}(R)$. But we shall see that these two threads are intertwined. We start by pinpointing certain kinds of identities in order to obtain large classes of PI-algebras (most notably representable algebras) and then utilize these identities to develop the structure theory.

We recall from Definitions 17.5 and 17.8 that any (noncommutative) polynomial $f \in C\{X\}$ is written uniquely as

$$f(x_1, \dots, x_m) = \sum c_j h_j, \quad (c_j \in C),$$

where the $h_j = h_j(x_1, \dots, x_m)$ run through the words in $x_1, \dots, x_m$. We call these $c_j h_j$ the **monomials** of f. We say that a word h is **linear** in x_i if x_i appears exactly once in h.

Definition 23.3. The polynomial $f(x_1, \dots, x_m)$ is t-**linear** (for $t \le m$) if each of its monomials is linear in x_i for each $1 \le i \le t$. If f is m-linear, we say that f is **multilinear**.

For example, the polynomial $[x_1, x_2] = x_1x_2 - x_2x_1$ is multilinear, but the 2-linear polynomial $x_1x_2x_3 - x_2x_1$ is not multilinear since x_3 does not appear in its second monomial. We are most interested in a certain kind of multilinear identity.

Remark 23.3′. Any multilinear polynomial $f(x_1, \dots, x_d)$ can be written as $\sum_{\pi\in S_d} c_\pi x_{\pi 1} \cdots x_{\pi d}$, summed over all permutations $\pi \in S_d$, where $c_\pi \in C$. (We write the subscript πi as short for $\pi(i)$.)

We say the polynomial f is **admissible** if $f \ne 0$ and all $c_\pi \in \{0, \pm 1\}$.

Definition 23.4. A ring R is a **PI-ring** if R satisfies an admissible multilinear identity f; then f is called a **PI** for R.

Of course, if f is a PI of R, then rearranging the indeterminates, we may (and shall) assume in the notation of Remark 23.3′ that $c_1 = 1$; i.e.,

$$f = x_1 \cdots x_d + \sum_{1 \ne \pi \in S_d} c_\pi x_{\pi 1} \cdots x_{\pi d}.$$

Definition 23.4 excludes the free algebra R over a field of characteristic p; obviously such R satisfies the identity px_1, but its structure theory is too broad to be useful. Other, seemingly more general, definitions exist in the literature, but they can be seen to be equivalent to Definition 23.4.

Example 23.5. Any product of n strictly upper triangular $n \times n$ matrices is 0. Since $[a, b] = ab - ba$ is strictly upper triangular, for any upper triangular matrices a, b, we conclude that the algebra of upper triangular $n \times n$

matrices satisfies the PI

$$[x_1, x_2][x_3, x_4] \cdots [x_{2n-1}, x_{2n}].$$

The reason we focus on multilinear identities is the following obvious but crucial observation:

Remark 23.6. If f is linear in x_i, then

$$f(r_1, \ldots, cr_i, \ldots, r_d) = cf(r_1, \ldots, r_i, \ldots, r_d)$$

for all $c \in C$, $r_i \in R$.

LEMMA 23.7. *Suppose R is spanned by a set B over a central subring C. Then a multilinear polynomial $f \in \mathrm{id}(R)$ iff f vanishes on all substitutions in B.*

Proof. $f(\sum c_{i1}b_{i1}, \ldots, \sum c_{im}b_{im}) = \sum c_{i1} \cdots c_{im} f(b_{i1}, \ldots, b_{im})$, as seen by applying Remark 23.6 repeatedly. □

Let us focus on an important special case.

PROPOSITION 23.8.

(i) *If f is a multilinear identity of W, and R is a central extension of W (cf. Definition 16.3), then $f \in \mathrm{id}(R)$.*

(ii) *Every multilinear identity of R is an identity of $R \otimes_C H$ for every commutative C-algebra H.*

Proof. (i) W spans R over $\mathrm{Cent}(R)$.

(ii) Special case of (i); cf. Remark 18.27. □

Identities of finite-dimensional algebras

Lemma 23.7 also provides a method for obtaining identities of arbitrary finite-dimensional algebras. It is useful to consider substitutions from $C\{X\}$ to itself; we define the specialization $x_j \mapsto x_k$ to be the homomorphism $C\{X\} \to C\{X\}$ sending x_j to x_k and leaving x_i fixed for all $i \neq j$. For example, the polynomial $[x_1, x_2]$ becomes 0 under the specialization $x_1 \mapsto x_2$.

We say that a polynomial $f(x_1, \ldots, x_m)$ is t-**alternating** (for $t \leq m$) if f becomes 0 under each specialization $x_j \mapsto x_k$ for every $1 \leq j < k \leq t$; i.e., if $f(x_1, \ldots, x_k, \ldots, x_k, \ldots, x_m) = 0$ for each $k \leq t$. This property is

reminiscent of the determinant and gives a determinant-like flavor to the theory, to be made explicit in Proposition 23.15.

For example, $[x_1, x_2]$ is 2-alternating, as is $x_1x_3x_2x_4 - x_2x_3x_1x_4$. Let us generalize these two polynomials. Since the alternating variables are so important, sometimes we write the non-alternating variables as $y_1, y_2, \ldots$. Also, we may write y as shorthand for all the y_j; thus, a t-alternating polynomial f often is written as $f(x_1, \ldots, x_t; y)$.

Here is an easy construction of t-alternating polynomials.

Remark 23.9. Given a t-linear polynomial $h(x_1, \ldots, x_t; y)$, define the **alternator**

$$h_{\mathrm{alt}}(x_1, \ldots, x_t; y) = \sum_{\pi \in S_t} \mathrm{sgn}(\pi) h(x_{\pi 1}, \ldots, x_{\pi t}; y).$$

It is easy to see that h_{alt} is an alternating polynomial. Indeed, if we specialize $x_j \mapsto x_k$, then the same monomial would appear in pairs whose coefficients are $\mathrm{sgn}\,\pi$ and $\mathrm{sgn}((j\,k)\pi) = -\mathrm{sgn}\,\pi$, which cancel. In fact, all alternating polynomials are obtained in this way; cf. Exercise 18.

Example 23.10. The **standard polynomial** s_t is defined as h_{alt}, where we take $h = x_1 \cdots x_t$. Explicitly,

$$s_t = \sum_{\pi \in S_t} (\mathrm{sgn}\,\pi)\, x_{\pi 1} \cdots x_{\pi t}.$$

The **Capelli polynomial** c_t is h_{alt}, where $h = x_1y_1x_2y_2 \cdots x_ty_t$. Explicitly,

$$c_t = \sum_{\pi \in S_t} \mathrm{sgn}(\pi) x_{\pi 1} y_1 x_{\pi 2} y_2 \cdots y_{t-1} x_{\pi t} y_t.$$

The reason t-alternating polynomials are of such great interest is the following observation.

PROPOSITION 23.11. *Any t-alternating polynomial f is an identity for every algebra R spanned by fewer than t elements over its center.*

Proof. By Lemma 23.7, it suffices for us to check that f vanishes whenever $x_1, \ldots, x_t$ are specialized to a spanning set B of R. But choosing B to have $< t$ elements, two of the x_i must be specialized to the same element of B, so the evaluation of f is 0 by definition of t-alternating. □

COROLLARY 23.12. *For any commutative ring C, $M_n(C)$ is a PI-ring since it satisfies any (n^2+1)-alternating polynomial.*

To check the sharpness of Corollary 23.12, we want a method of checking that a polynomial

$$f = x_1 \cdots x_d + \sum_{1 \neq \pi \in S_d} c_\pi x_{\pi 1} \cdots x_{\pi d}$$

is not an identity of R. One obvious method is to find a substitution $r_1, \dots, r_d$ for which $r_1 \cdots r_d \neq 0$ but $r_{\pi 1} \cdots r_{\pi d} = 0$ for every $\pi \neq 1$ in S_d.

Remark 23.13. (i) The Capelli polynomial c_{n^2} is not an identity of $M_n(C)$, and in fact $e_{11} \in c_{n^2}(M_n(C))$. Indeed, we list the set of matrix units

$$a_1 = e_{11},\ a_2 = e_{12},\ \dots,\ a_n = e_{1n},\ a_{n+1} = e_{21},\ \dots$$
$$\dots,\ a_{n^2-1} = e_{n,n-1},\ a_{n^2} = e_{nn},$$

and also list the same set of matrix units in the unique order

$$b_1 = e_{11}, \quad b_2 = e_{21},\ \dots,\ b_n = e_{n2},\ \dots,\ b_{n^2-1} = e_{n-1,n}, \quad b_{n^2} = e_{n1}$$

for which $a_1 b_1 a_2 b_2 \cdots a_{n^2} \neq 0$. Then $a_1 b_1 a_2 b_2 \cdots a_{n^2} b_{n^2} = e_{11}$, whereas $a_{\pi 1} b_1 a_{\pi 2} b_2 \cdots a_{\pi(n^2)} b_{n^2} = 0$ for every $\pi \neq (1)$, so

$$c_{n^2}(a_1, \dots, a_{n^2}, b_1, \dots, b_{n^2}) = a_1 b_1 a_2 b_2 \cdots a_{n^2} b_{n^2} = e_{11} \neq 0.$$

(ii) In general, any multilinear polynomial $f(x_1, \dots, x_d)$ is not a PI of $M_n(R)$ unless $d \geq 2n$. (Indeed, otherwise we take

$$0 = f(e_{11}, e_{12}, e_{22}, e_{23}, \dots, e_{k-1,k}, e_{kk}, \dots) = e_{1m}$$

where $m = \left[\frac{d}{2}\right]$, a contradiction.) In particular, the standard polynomial $s_{2n-1} \notin \mathrm{id}(M_n(C))$. One of the first major theorems of PI-theory was the Amitsur-Levitzki Theorem [**AmL**], that $s_{2n} \in \mathrm{id}(M_n(C))$; cf. Exercise 22.

We can now begin to link representable algebras to polynomial identities. Although the identity s_{2n} is of minimal degree, often it is more convenient to work with the Capelli identities because of Remark 23.13(i).

Remark 23.14. Every representable algebra, being isomorphic to a subalgebra of $M_n(K)$ for some n (and K), satisfies the corresponding Capelli identity c_{n^2+1}. (The converse is false, by Exercise 8.) Any *proper* F-subalgebra of $M_n(F)$ has dimension $< n^2$, and thus satisfies the identity c_{n^2}.

This innocuous observation ties in nicely to group representations, as follows. Extending a group representation $\rho: G \to M_n(F)$ to a homomorphism $\hat{\rho}: F[G] \to M_n(F)$, we see that $\hat{\rho}(F[G])$ is a representable algebra and satisfies the identity c_{n^2} iff $\hat{\rho}$ is not onto, i.e., iff the representation ρ is reducible. In this way, the maximal degree of the irreducible representations of G is ascertained from the minimal Capelli identity of $F[G]$.

Let us also make more explicit the det-like property of alternating polynomials, for use in Appendix 23A.

PROPOSITION 23.15. *Suppose $f(x_1, \dots, x_d)$ is n-alternating and V is an n-dimensional subspace of R. Then for any $a \in M_n(F)$, $v_1, \dots, v_n \in V$, and $r_{n+1}, \dots, r_d$ in R, we have*

$$f(av_1, \dots, av_n, r_{n+1} \dots, r_d) = \det(a) f(v_1, \dots, v_n, r_{n+1}, \dots r_d).$$

Proof. The two sides are proportional, by the axiomatic development of the determinant as an alternating function, and the ratio is 1, as seen by taking $a = 1$. □

This observation has been extended quite far by Razmyslov and Zubrilin, as discussed in Belov-Rowen [**BelR, Chapter 2**].

Trace identities of matrices.

Although we have determined some identities of $M_n(F)$, the question remains of determining all identities of $M_n(F)$. This is an extremely difficult question that has motivated much of the PI-research from the birth of the subject; the full solution still remains out of reach. There is an answer of sorts, when one broadens the definition to permit the formal use of traces. Rather than develop the theory formally, let us consider the case of $M_2(F)$. Any matrix x satisfies its Hamilton-Cayley polynomial

$$x^2 - \mathrm{tr}(x)x + \det(x) = 0; \tag{23.1}$$

and, furthermore,

$$2\det(x) = \mathrm{tr}(x)^2 - \mathrm{tr}(x^2),$$

seen at once by reducing to diagonal matrices via Zariski density. Thus, when $\mathrm{char}\, F \neq 2$, $M_2(F)$ satisfies the **Hamilton-Cayley trace identity**

$$x^2 - \mathrm{tr}(x)x + \frac{1}{2}\big(\mathrm{tr}(x)^2 - \mathrm{tr}(x^2)\big).$$

In general, over any field F of characteristic 0, Newton's Formulas [**Row3, p. 153**] enable one to express the coefficients of the characteristic polynomial of x in terms of traces of powers of x, so for any n there is a corresponding Hamilton-Cayley trace identity for $M_n(F)$, depending only on n. Helling [**He**], Procesi [**Pr**], and Razmyslov [**Ra2**] independently proved that every trace identity (and thus every PI) of $M_n(F)$ is a formal consequence of the Hamilton-Cayley trace identity, and Razmyslov [**Ra2**] used this method to reprove the Amitsur-Levitzki Theorem. But this theorem still does not provide explicit answers to many other important computational PI-questions.

Multilinearization.

The Hamilton-Cayley trace identity is not linear, and we will come across various other useful non-linear identities. In order to deal with them, we need a procedure to transform an arbitrary identity into a multilinear identity.

Linearization procedure 23.16. Given an identity $f(x_1, \dots, x_m)$ with $\deg_i f > 1$, we introduce another indeterminate x' and consider the new identity

$$\Delta_i f(x_1, \dots, x_m, x') = f(x_1, \dots, x_i + x', \dots, x_m) \\ -f(x_1, \dots, x_i, \dots, x_m) - f(x_1, \dots, x', \dots, x_m).$$

$\Delta_i f$ has the same total degree d as f, but $1 \leq \deg_i(\Delta_i f) < \deg_i f$, so after a finite number $(d-m)$ of such linearization steps we arrive at a multilinear identity, called the **multilinearization** of f.

We call the homogeneous components of $\Delta_i f$ (and of any subsequent linearization step) the **partial linearizations** of f. Although the partial linearizations are of mainly technical interest in characteristic 0, in which case they are also identities, they also play a significant role for PI-algebras of characteristic > 0, as indicated in Appendix 23B.

Example 23.17. An algebra R is called **Boolean** if it satisfies the identity $x_1^2 - x_1$. Multilinearizing yields the multilinear identity

$$(x_1 + x_2)^2 - (x_1 + x_2) - (x_1^2 - x_1) - (x_2^2 - x_2) = x_1x_2 + x_2x_1;$$

i.e., R is anticommutative. Now specializing $x_2 \mapsto 1$ yields the identity $2x_1$, implying that $\mathrm{char}(R) = 2$. So we have proved that any Boolean algebra is commutative of characteristic 2.

Actually, we already have utilized this idea in Remarks 21.2 and 21B.8. The linearization procedure involves some fine points; cf. Exercises 9ff. An alternate approach is given in Exercise 15.

Central polynomials

In 1943, M. Hall observed that the polynomial $[x_1, x_2]^2$ takes on only scalar values in $M_2(F)$, as is clear from Equation (23.1), since $\mathrm{tr}[a, b] = 0$ for any $a, b \in M_2(F)$, yielding $[a, b]^2 = -\det[a, b] \in F$. This leads us to a new notion.

Definition 23.18. A polynomial $f(x_1, \dots, x_m)$ is R-**central** if $0 \neq f(R) \subseteq \mathrm{Cent}(R)$; i.e., f is *not* an identity of R but $f(r_1, \dots, r_m) \in \mathrm{Cent}(R)$ for all $r_i \in R$.

Remark 23.19. (i) A polynomial $f(x_1, \dots, x_m)$ is R-central iff $[x_{m+1}, f]$ (but not f) is an identity of R. Thus, two algebras satisfying precisely the same sets of multilinear identities also possess precisely the same multilinear central polynomials. This observation permits us to pass central polynomials from an algebra to related algebras.

(ii) By Lemma 23.7, when B spans R over C, a multilinear polynomial f is R-central iff $0 \neq f(B) \subseteq \mathrm{Cent}(R)$.

Example 23.20. (i) x_1 is central for any commutative ring.

(ii) The polynomial $[x_1, x_2]^2$ is central for $M_2(F)$ for any field F, and can be linearized to produce a multilinear central polynomial.

(iii) An example of a PI-algebra that does not have a central polynomial is the algebra of upper triangular $n \times n$ matrices; cf. Exercise 23.

Example 23.20(ii) led Kaplansky to ask whether multilinear central polynomials exist for arbitrary $n \times n$ matrix algebras. In 1972 Formanek [**Fo**] and Razmyslov [**Ra1**] discovered such central polynomials, thereby leading to the introduction of techniques from commutative algebra to PI-theory, culminating in a beautiful structure theory with applications to central simple algebras and, more generally, to Azumaya algebras (Appendix 25B). Following [**BelR**], we extract the main idea from Razmyslov's construction and simplify the remainder of his proof.

Definition 23.21. A k-**weak** identity of the matrix algebra $M_n(F)$ is a polynomial $f(x_1, \dots, x_m)$ that vanishes whenever $x_1, \dots, x_k$ are specialized to matrices of trace 0 in $M_n(F)$.

Example 23.22. (i) c_{n^2} is a multilinear n^2-weak identity of $M_n(F)$ (since it is n^2-alternating whereas the matrices of trace 0 comprise a subspace of dimension only $n^2 - 1$), although $c_{n^2} \notin \mathrm{id}(M_n(F))$ by Remark 23.13(i).

(ii) A polynomial $f(x_1, x_2, \dots, x_m)$ is a k-weak identity of $M_n(F)$ iff the polynomial

$$g(x_1, \dots, x_{m+1}) = f(x_1, \dots, x_{k-1}, [x_k, x_{m+1}], x_{k+1}, \dots, x_m)$$

is a $(k-1)$-weak identity of $M_n(F)$, since the matrices of trace 0 are spanned by the Lie commutators.

(iii) Applying (ii) $n^2 - 1$ times to (i) gives us a 1-weak identity of $M_n(F)$ that is not an identity.

Razmyslov found a correspondence between central polynomials and 1-weak identities, which is best explained in terms of another generalization of polynomial identity.

Definition 23.23. An expression $f = \sum_{i=1}^{k} a_i x b_i$, with a_i, b_i in R, is a **linear generalized identity** (LGI) for R if $\sum a_i r b_i = 0$, $\forall r \in R$; f is **linear generalized central** (LGC) for R if $\sum a_i r b_i \in \mathrm{Cent}(R), \forall r \in R$, but f is not an LGI.

Digression. A "generalized polynomial" is a noncommutative polynomial with coefficients from R interspersed throughout the x_i, and there is a well-developed theory of generalized polynomial identities; cf. Beidar-Martindale-Martindale [**BeiMM**] or Rowen [**Row2, Chapter 7**]. However, this very special case suffices for our purpose.

Next, recall the opposite ring R^{op} of Definition 13.19.

Remark 23.24. If $\sum a_u x b_u$ is an LGI of R, then $\sum b_u x a_u$ is an LGI of R^{op}.

LEMMA 23.25. *Suppose $f = \sum_i a_i x b_i$ is not an LGI of $R = M_n(F)$. Then $\sum b_i x a_i$ is LGC iff $f(r) = 0$ for every matrix r of trace 0.*

Proof. The condition is that

$$0 = f([r,s]) = \sum_u a_u r s b_u - \sum_u a_u s r b_u, \qquad \forall r, s \in R,$$

i.e., $\sum a_u x s b_u - \sum a_u s x b_u$ is an LGI for each s in R. By Remark 23.24, this is equivalent to $\sum s b_u x a_u - \sum b_u x a_u s = [s, \sum b_u x a_u]$ being an LGI of R^{op} for each s, which means $\sum b_u x a_u$ is R^{op}-central. But $R^{\mathrm{op}} \cong R$ since the transpose is an anti-automorphism of $M_n(F)$. □

We are ready for a major result.

THEOREM 23.26 (RAZMYSLOV). *There is a 1:1 correspondence between multilinear central polynomials of $M_n(F)$ and multilinear 1-weak identities that are not identities, given by*

$$\sum f_i(x_2, \dots, x_d) x_1 g_i(x_2, \dots, x_d) \mapsto \sum g_i(x_2, \dots, x_d) x_1 f_i(x_2, \dots, x_d).$$

Proof. Write a given multilinear polynomial

$$f(x_1, x_2, \dots, x_d) = \sum_{i=1}^{t} f_i(x_2, \dots, x_d) x_1 g_i(x_2, \dots, x_d).$$

Picking matrices $r_2, \dots, r_d$ arbitrarily, let $a_i = f_i(r_2, \dots, r_d)$ and $b_i = g_i(r_2, \dots, r_d)$, and apply Lemma 23.25 for each such choice. □

COROLLARY 23.27. *For any $n \in \mathbb{N}^+$, there is a multilinear, n^2-alternating, central polynomial for $M_n(F)$ for all fields F.*

Proof. This is Theorem 23.26 confronted with Example 23.22(iii). □

For further reference, we call this central polynomial g_n.

Remark 23.27′. $\mathrm{id}(M_n(F)) \subset \mathrm{id}(M_{n-1}(F))$ for each $n \geq 2$. More precisely, any polynomial f that is central or an identity for $M_n(F)$ is in $\mathrm{id}(M_k(F))$ for all $k < n$. (Indeed, one can embed $M_k(F)$ into the upper left corner of $M_n(F)$ as a ring *without* 1 by placing zeroes in all the extra entries; then each evaluation of f on $M_k(F)$ is a scalar matrix with zero in the n, n position, and thus is 0.)

The Grassmann algebra

After matrices, the most important example in PI-theory is the Grassmann (e.g., exterior) algebra defined in Example 18.40(iv). For convenience, we consider $G = E(V)$, where V is a countably infinite-dimensional vector space over $\mathbb{Q}$. (Although the letter G has been used often for a group, in this discussion it refers to the Grassmann algebra.) Recall that G has the base

$$B = \{1\} \cup \{e_{i_1} \cdots e_{i_m} : i_1 < i_2 < \cdots < i_m,\ m \in \mathbb{N}^+\}. \tag{23.2}$$

We say that the element $e_{i_1} \cdots e_{i_m}$ is **even** or **odd** depending on whether its length m is even or odd, respectively. Let G_0 (resp. G_1) denote the linear span of the even (resp. odd) elements. Then $G = G_0 \oplus G_1$ is $\mathbb{Z}/2$-graded, since $G_1^2 \subset G_0$. Furthermore, each even element commutes with every element of B, and thus $G_0 = \mathrm{Cent}(G)$. The odd elements anti-commute and thus have the following remarkable property:

Remark 23.28. (i) If $a, b \in G_1$, then $arb = -bra$ for every $r \in G$. (Indeed, one may write $r = r_0 + r_1$ for $r_i \in G_i$. Then

$$arb = ar_0b + ar_1b = -r_0ba + r_1ba = -br_0a - br_1a = -bra.)$$

(ii) $[G, G] \subseteq G_0$. (Indeed, if $a_i, b_i \in G_i$ for $i = 0, 1$, then $[a_i, b_j] = 0$ unless $i = j = 1$, so $[a_0 + a_1, b_0 + b_1] = [a_1, b_1] = 2a_1b_1 \in G_0$.)

(iii) G satisfies the **Grassmann identity** $[[x_1, x_2], x_3]$, in view of (ii). Hence, G satisfies the central polynomial $[x_1, x_2]$.

(iv) In characteristic 0, G fails to satisfy any Capelli identity, since if a_i are odd and b_i are even, then $c_t(a_1, \dots, a_t; b_1, \dots, b_t) = t! a_1 \cdots a_t b_1 \cdots b_t$. In particular, G cannot be representable! This observation of P.M. Cohn was viewed at first as a curiosity, but is now seen to be a foundation stone of PI-theory, playing a basic role in Kemer's proof of Specht's problem, as discussed in Appendix 23A.

Main theorems in PI-structure theory

Our next objective is to present the basic PI-structure theorems as expeditiously as possible. Accordingly, we rely on the Jacobson-Levitzki-Amitsur structure theory developed in Appendix 15A. One main theme is to match identities of a given algebra with identities of matrices.

Definition 23.29. An algebra R has PI-**class** n when both of the following two conditions hold:

1. R satisfies all multilinear PIs of $M_n(\mathbb{Z})$;
2. g_n (defined after Corollary 23.27) is R-central.

In view of Remark 23.19, a ring R has PI-class n whenever it satisfies precisely the same multilinear PIs over $\mathbb{Z}$ as $M_n(\mathbb{Z})$. However, the converse may fail: $\mathbb{Z}/2$ has PI-class 1 but also satisfies the multilinear PI $xy + yx$, which is not satisfied by $\mathbb{Z}$.

The point of studying PI-class is that once the algebra R is established to have PI-class n, we can use the identities of matrices, and especially the central polynomial g_n, to study the structure of R in terms of its center, applying techniques from commutative algebra. So our strategy is first to establish a wide range of algebras of PI-class n and then to examine the structural implications. We start with some basic observations.

Remark 23.29′. (i) By Remark 23.13(ii), if R has PI-class n and satisfies a PI of degree d, then $n \leq [\frac{d}{2}]$.

(ii) Suppose R is a central extension of an algebra W. Then R has PI-class n iff W has PI-class n, in view of Proposition 23.8.

(iii) Suppose R_i are PI-algebras of PI-class n_i for $i \in I$, with each n_i bounded by some given number n. Then $\prod_{i\in I} R_i$ has PI-class equal to $\max\{n_i : i \in I\} \leq n$.

(iv) Suppose a PI-algebra R is a subdirect product of PI-algebras of PI-class n_i. Then R has PI-class equal to $n = \max\{n_i : i \in I\}$. (Indeed, R satisfies all identities of $\prod R_i$, which has PI-class n. But taking i such that R_i has PI-class n, we have a surjection of R to R_i, implying that g_n is not an identity of R, and thus is R-central.)

We are ready to find algebras of PI-class n.

Proposition 23.30. *$M_n(C)$ has PI-class n, for any commutative ring C.*

Proof. $M_n(C) \simeq M_n(\mathbb{Z}) \otimes_{\mathbb{Z}} C$ satisfies all multilinear PIs of $M_n(\mathbb{Z})$ by Proposition 23.8. In particular, g_n is either central or an identity for $M_n(C)$. If $g_n \in \mathrm{id}(M_n(C))$, then g_n is an identity of $M_n(C)/M_n(P) \cong M_n(C/P)$ for any maximal ideal P of C, which is impossible in view of Corollary 23.27 (since C/P is a field). □

We finally have the first basic representability theorem in PI-theory.

Theorem 23.31 (Kaplansky's Theorem). *Any primitive ring R satisfying a PI of degree d is simple of dimension n^2 over its center, for some $n \leq [\frac{d}{2}]$. Also, R is representable of PI-class n.*

Proof. Take a simple R-module M. By the Density Theorem (15A.2), R is dense in $\mathrm{End}\, M_D$, viewing M as a vector space over D. If M were infinite-dimensional, we could take a base $a_1, a_2, \ldots$ of M over D and define $r_i \in R$ such that $r_i a_j = \delta_{ij} a_{j-1}$ for $1 \leq i, j \leq d+1$. Then $f(r_1, \ldots, r_d)a_{d+1} = a_1$, contrary to f being a PI of R. Hence, $[M : D] = t < \infty$ for some t, so $R \cong M_n(D)$ is simple Artinian.

Let $F = \mathrm{Cent}(R)$, and let F_1 be an algebraically closed field of cardinality greater than $[R : F]$. Then by Corollary 18.34, $R \otimes_F F_1$ remains simple, so we may replace R by $R \otimes_F F_1$ and F by F_1, and thereby assume that F is algebraically closed with $[R : F] < |F|$. Since R is simple Artinian by the last paragraph, we can write $R = M_n(D)$, and have $[D : F] < |F|$, implying that D is algebraic over F by Corollary 15A.8, and thus $D = F$. We conclude that $R = M_n(F)$, where $n \leq [\frac{d}{2}]$ by Remark 23.29′, and the assertion now is immediate. □

Kaplansky's Theorem thrusts PI-theory into the mainstream of algebra. Nil ideals also behave very well.

Proposition 23.32. *A semiprime PI-ring R has no nonzero left or right nil ideals.*

Proof. We prove more generally that if L is any nil left ideal of a semiprime ring R satisfying $Lf(L) = 0$ for some admissible multilinear polynomial f, then $L = 0$. This is by induction on $d = \deg f$. If $d = 1$, then we may assume that $f = x_1$, so $L^2 = 0$, implying that $L = 0$.

In the general case, write

$$f(x_1, \ldots, x_d) = x_1 g(x_2, \ldots, x_d) + h(x_1, \ldots, x_d),$$

where $x_1 g$ is the sum of all monomials of f starting with x_1, and no monomial of h starts with x_1. (Thus, g is admissible multilinear of degree $d-1$.) If $L \neq 0$, then $L^2 \neq 0$ (since R is prime), so there exists $a \in L$ with $La \neq 0$. Since a is nilpotent, there is $1 \leq m \in \mathbb{N}$ with $La^m \neq 0$ but $La^{m+1} = 0$. Thus, for all $r_1, r_2, \ldots, r_d$ in L,

$$0 = Lf(ar_1, r_2 a^m, \ldots, r_d a^m) = Lar_1\, g(r_2 a^m, \ldots, r_d a^m) + 0,$$

since some $r_j a^{m+1} r_1 = 0$ appears in the evaluation of each monomial of h. But then $LaL\, g(La^m) = 0$, implying that $La^m\, g(La^m) = 0$. (This is clear for $m > 1$; when $m = 1$, we note that

$$(Lag(La))^2 \subseteq (La)^3 g(La) \subseteq LaLg(La) = 0,$$

implying that $Lag(La) = 0$ since R is semiprime.) By induction, since $\deg g < d$, the left ideal $La^m = 0$, a contradiction. □

THEOREM 23.33. *Any semiprime PI-ring R satisfying a PI of degree d has PI-class n for some $n \leq [\frac{d}{2}]$, and every ideal A intersects the center nontrivially.*

Proof. R, and thus $R[\lambda]$, satisfy a multilinear PI of some degree d. By Proposition 23.32, R has no nonzero nil ideals, so by Theorem 15A.5, $\mathrm{Jac}(R[\lambda]) = 0$. Thus $R[\lambda]$ is a subdirect product of primitive PI-rings $\{R_i : i \in I\}$, each of which has some PI-class $n_i \leq [\frac{d}{2}]$, so $R[\lambda]$ (and thus R) has some PI-class n.

Let $\pi_i\colon R \to R_i$ be the canonical projection. Taking $I' = \{i \in I : \pi_i(A[\lambda]) \neq 0\}$ and noting that $\pi_i(A[\lambda])$ is an ideal of the simple ring R_i, we see that $\pi_i(A[\lambda]) = R_i$ for each $i \in I'$. Thus, letting $n' = \max\{n_i : i \in I'\}$, we see that $0 \neq g_{n'}(A[\lambda]) \subseteq A[\lambda] \cap \mathrm{Cent}(R[\lambda]) \subseteq (A \cap \mathrm{Cent}(R))[\lambda]$, implying that $A \cap \mathrm{Cent}(R) \neq 0$. □

We can now begin to apply commutative techniques to PI-algebras. Given any prime algebra R, we define the **ring of central fractions** $Q(R) = S^{-1}R$, where $S = \mathrm{Cent}(R) \setminus \{0\}$.

COROLLARY 23.34. *If R is prime PI with center C, then the ring of central fractions $Q(R)$ is simple and f.d. over the field of fractions F of C. In particular, R is representable of some PI-class n.*

Furthermore, if R is a (noncommutative) domain, then $Q(R)$ is a division ring.

Proof. Clearly $S^{-1}R$ is prime and its center $F = S^{-1}C$ is a field. By Theorem 23.33, $S^{-1}R$ is simple and is f.d. over F by Kaplansky's Theorem. Hence, $S^{-1}R$ has some PI-class n, which then is the PI-class of R.

When R is a domain, $Q(R)$ is a domain, and thus a division ring, in view of Remark 14.25(ii). □

Here is an example of the power of structure theory for PI-rings.

THEOREM 23.35. *Suppose R is semiprime PI of PI-class n, and S is a nil subset such that for any s_1, s_2 in S there is $\nu = \nu(s_1, s_2) \in \mathbb{Z}$ with $s_1 s_2 + \nu s_2 s_1 \in S$. Then $S^n = 0$.*

Proof. By Proposition 16.8, the intersection of the prime ideals of R is a nil ideal and thus 0, so one may assume that R is prime. Then $Q(R)$ is simple Artinian, and we are done by Theorem 15.23. □

It also follows at once from Exercise 5 that any semiprime, affine PI-algebra is representable.

Applications of alternating central polynomials.

Sometimes localizing at all of the center (excluding 0) is too crude a procedure. A bit more care gives more precise results. First, a few general observations about identities.

Remark 23.36. For any algebra homomorphism $\varphi\colon R \to R$ and any polynomial $f \in C\{X\}$,

$$\varphi(f(r_1, \ldots, r_d)) = f(\varphi(r_1), \ldots, \varphi(r_d)).$$

Let us apply this remark to matrices.

Remark 23.37. (i) If a 1-linear polynomial $f(x_1, \ldots, x_d)$ is *not* an identity of $M_n(C)$, then $f(e_{11}, a_2, \ldots, a_d) \neq 0$ for suitable $a_2, \ldots, a_d \in M_n(C)$. (Indeed by linearity, we have some $b = f(e_{ij}, a_2, \ldots, a_d) \neq 0$ for suitable i, j. Changing indices, we assume that $i = 1$. If $j = 1$ we are done, so we may assume that $f(e_{11}, a_2, \ldots, a_d) = 0$. But then $f(e_{11} + e_{1j}, a_1, \ldots, a_d) = b$, and $r = I + e_{1j}$ satisfies $r^{-1} = I - e_{1j}$ and $r^{-1} e_{11} r = e_{11} + e_{1j}$. Thus, $e_{11} = r(e_{11} + e_{ij})r^{-1}$, so

$$f(e_{11}, ra_1 r^{-1}, \ldots, ra_{d-1} r^{-1}) = rbr^{-1} \neq 0,$$

as desired.)

(ii) If R has PI-class n, then

$$h_n = g_n(c_{n^2}(x_1, \ldots, x_{n^2}, y_1, \ldots, y_{n^2}), y_{n^2+1}, y_{n^2+2}, \ldots, y_{n^2+d-1}) \tag{23.3}$$

is multilinear, R-central, and also n^2-alternating. (Indeed, by Remark 23.13, $e_{11} \in c_{n^2}(M_n(\mathbb{Z}))$, so making this substitution in g_n, we see by (i) that h_n is not an identity of $M_n(\mathbb{Z})$, so h_n is central. The multilinear and n^2-alternating properties are immediate from those of c_{n^2}.)

It is useful to view h_n as an alternator, in the terminology of Remark 23.9; namely, $h_n = g_{\mathrm{alt}}$ where

$$g(x_1, \ldots, x_{n^2}; y) = g_n(x_1 y_1 x_2 y_2 \cdots x_{n^2} y_{n^2}, y_{n^2+1}, y_{n^2+2}, \ldots, y_{n^2+d-1}). \tag{23.4}$$

This polynomial h_n, although cumbersome to write out, has some very useful properties, encapsulated in the next lemma.

Remark 23.38. Suppose $g(x_1,\dots,x_t;y)$ is a t-linear polynomial, and let $f(x_0,x_1,\dots,x_t;y)=g(x_1,\dots,x_t;y)x_0$, viewed as a $(t{+}1)$-linear polynomial in $x_0,\dots,x_t$.

If $h=g_{\rm alt}(x_1,\dots,x_t;y)$, then defining

$$\tilde{h}(x_1,\dots,x_t;y)=\sum_{i=0}^{t}(-1)^i h(x_1,\dots,x_{i-1},x_{i+1},\dots,x_t;y)x_i,$$

we see that $\tilde{h}=f_{\rm alt}(x_0,\dots,x_t;y)$, and thus $\tilde{h}$ is $(t{+}1)$-alternating.

LEMMA 23.38′. *Suppose R has PI-class n, and let $m=n^2+d-1$, notation as in Remark 23.37(ii). For any $r_1,\dots,r_{n^2},s_1,\dots,s_m,r$ in R, we have*

$$\begin{aligned}&h_n(r_1,\dots,r_{n^2},s_1,\dots,s_m)r\\ &=\sum_{i=1}^{n^2}(-1)^{i+1}h_n(r,r_1,\dots,r_{i-1},r_{i+1},\dots,r_n^2,s_1,\dots,s_m)r_i.\end{aligned}\tag{23.5}$$

Proof. By Remark 23.38, $\tilde{h}_n$ is (n^2+1)-alternating and thus is an identity of R, by hypothesis; Equation (23.5) follows at once. □

THEOREM 23.39. *Suppose R has PI-class n and center C, and there exist $r_1,\dots,r_{n^2},s_1,\dots,s_m$ in R such that $h_n(r_1,\dots,r_{n^2},s_1,\dots,s_m)=1$. Then*

(i) *R is a free C-module with base $r_1,\dots,r_{n^2}$;*

(ii) *There is a 1:1 correspondence between $\{$ideals of $R\}$ and $\{$ideals of $C\}$, given by $A\mapsto A\cap C$ for $A\lhd R$; the inverse correspondence is given by $I\mapsto IR$ for $I\lhd C$.*

Proof. (i) By Lemma 23.38′, for any r in R,

$$\begin{aligned}r&=h_n(r_1,\dots,r_{n^2},s_1,\dots,s_m)r\\ &=\sum_{i=1}^{n^2}\pm h_n(r,r_1,\dots,r_{i-1},r_{i+1},\dots,r_{n^2},s_1,\dots,s_m)r_i\in\sum Cr_i,\end{aligned}$$

proving that $r_1,\dots,r_{n^2}$ span R.

To prove independence, suppose $\sum_{i=1}^{n^2}c_ir_i=0$. Then

$$\begin{aligned}0=h_n(0,r_2,\dots,r_{n^2},s_1,\dots,s_m)&=h_n\Big(\sum c_ir_i,r_2,\dots,r_{n^2},s_1,\dots,s_m\Big)\\ &=\sum c_ih_n(r_i,r_2,\dots,r_{n^2},s_1,\dots,s_m).\end{aligned}$$

But $h_n(r_i,r_2,\dots,r_{n^2},s_1,\dots,s_m)=0$ for all $2\le i\le n^2$ since r_i repeats, so

$$0=c_1h_n(r_1,\dots,r_{n^2},s_1,\dots,s_m)=c_1.$$

The same argument (putting 0 in the j position) shows that $c_j=0$ for each $1\le j\le n^2$.

(ii) For any $A\lhd R$ and a in A,

$$\begin{aligned}a&=ah_n(r_1,\dots,r_{n^2},s_1,\dots,s_m)\\ &=\sum_{i=1}^{n^2}h_n(a,r_1,\dots,r_{i-1},r_{i+1},\dots,r_{n^2},s_1,\dots,s_m)r_i\in(A\cap C)R,\end{aligned}$$

proving that $A=(A\cap C)R$. Conversely, for any $I\lhd C$ and $c=\sum c_ir_i'\in IR\cap C$, for $c_i\in I$ and $r_i'\in R$,

$$\begin{aligned}c&=ch_n(r_1,\dots,r_{n^2},s_1,\dots,s_m)=h_n(cr_1,\dots,r_{n^2},s_1,\dots,s_m)\\ &=h(\sum c_ir_i'r_1,\dots,r_{n^2},s_1,\dots,s_m)\\ &=\sum c_ih(r_i'r_1,r_2,\dots,r_{n^2},s_1,\dots,s_m)\in I,\end{aligned}$$

proving that $I=IR\cap C$. Hence, the two correspondences given in the assertion of (ii) are inverses. □

Theorem 23.39 is so powerful that we would like to know how to broaden its applicability as far as possible. Fortunately, the naive approach is very successful here. Namely, suppose the polynomial h_n of (23.3) is R-central. If S is any multiplicative subset of $\mathrm{Cent}(R)$ containing some element of $h_n(R)$, then by the results of Chapter 8 of Volume 1 we can form the **central localization** $S^{-1}R$, which obviously satisfies the condition of Theorem 23.39; indeed, if $s=h_n(r_1,r_2,\dots)\in S$, then $1=h_n(\frac{r_1}{s},\frac{r_2}{1},\dots)$ in $S^{-1}R$.

Now, for any prime ideal $\mathfrak{p}$ of $C=\mathrm{Cent}(R)$ not containing $h_n(R)$. we can take $S=C\setminus\mathfrak{p}$, which satisfies the condition of the previous paragraph. In particular, it is easy to see that R has a unique prime ideal whose intersection with C is $\mathfrak{p}$, which localizes to the unique maximal ideal of $S^{-1}R$. For R semiprime of PI-class n, the set of such prime ideals is dense in $\mathrm{Spec}(R)$, and "generically" we conclude that the methods of commutative algebra are accessible for semiprime PI-algebras. The theory is far richer for affine PI-algebras, as to be seen in Appendix A, where this method is carried further.

Digression: PI-algebras without 1.

Throughout this book, our associative algebras are endowed with the unit element 1, since the structure theory is easier to carry out, and the

vast majority of examples of associative algebras have unit elements. One exception is in PI-theory, which often deals with questions of nilpotence; thus it is convenient to include algebras without 1 in the theory. (We did this in disguise in Proposition 23.32, which could have been stated more concisely in terms of algebras without 1. See Exercises 27 and 28 to connect with PI-theory.)

Definition 23.40. An algebra R_0 without 1 is **nil of bounded index** n iff $x^n \in \mathrm{id}(R_0)$; R_0 is **nilpotent of index** m iff $x_1 \cdots x_m \in \mathrm{id}(R_0)$.

Levitzki proved that if R_0 is affine (cf. Definition 17.1) and nil of bounded index n, then R_0 is nilpotent; this can be seen as an instant consequence of Shirshov's Dichotomy Lemma (Theorem 23A.3). In fact, in characteristic 0, Dubnov-Ivanov and Nagata showed that the same conclusion holds without the hypothesis of R_0 affine. Although there are elementary proofs of this result, Razmyslov [**Ra2**] found an elegant proof (with the stronger bound that R_0 has nilpotence index $\leq n^2$) using the technique of Young tableaux to be described in Appendix 23A.

Varieties and relatively free algebras

Having seen the structural consequences of a given algebra R having a PI, we take a more algebraic-geometric viewpoint and focus on *all* the identities of an algebra R, namely $\mathrm{id}(R)$.

Definition 23.41. Given a class $\mathcal{V}$ of algebras, we define $\mathrm{id}(\mathcal{V}) = \bigcap\{\mathrm{id}(R) : R \in \mathcal{V}\}$, the set of identities satisfied by each algebra from $\mathcal{V}$.

Conversely, given a subset $\mathcal{I} \subset C\{X\}$, we define the **variety** $\mathcal{V} = \mathcal{V}(\mathcal{I})$ to be the class of algebras R for which $\mathrm{id}(R) \supseteq \mathcal{I}$; i.e., every polynomial from $\mathcal{I}$ is an identity of R.

Remark 23.42. Any variety $\mathcal{V}$ is closed with respect to subalgebras, homomorphic images, and direct products (over a set of any cardinality). Conversely, any class of algebras having these properties can be displayed as the variety of a suitable set of polynomials; cf. Exercise 29.

Patently $\mathrm{id}(\mathcal{V}) \lhd C\{X\}$. However, $\mathrm{id}(\mathcal{V})$ satisfies an extra important property.

Definition 23.43. Suppose $\mathcal{I} \lhd C\{X\}$. We call $\mathcal{I}$ a T-**ideal**, and write $\mathcal{I} \lhd_T C\{X\}$ if $\varphi(\mathcal{I}) \subseteq \mathcal{I}$ for every homomorphism $\varphi: C\{X\} \to C\{X\}$ (or, in other words, if for any $f(x_1, \ldots, x_m) \in \mathcal{I}$, we have $f(h_1, \ldots, h_m) \in \mathcal{I}$ for each $h_1, \ldots, h_m \in C\{X\}$.)

Remark 23.44. In view of Remark 23.36, $\mathrm{id}(\mathcal{V}) \lhd_T C\{X\}$ for any class $\mathcal{V}$ of algebras.

Definition 23.45. Given a T-ideal $\mathcal{I}$, we define the **relatively free** algebra $\mathcal{F}_\mathcal{I} = C\{X\}/\mathcal{I}$. We write $\mathcal{F}_\mathcal{I} = C\{\bar{X}\}$, where $\bar{x}_i = x_i + \mathcal{I}$, the natural image of x_i in $\mathcal{F}_\mathcal{I}$.

$\mathcal{F}_\mathcal{I}$ is free in the variety $\mathcal{V}(\mathcal{I})$ in the following sense:

Remark 23.46. (i) Given an algebra $R \in \mathcal{V}(\mathcal{I})$ and any subset $\{r_i : i \in I\}$ of R, there is a unique homomorphism $\mathcal{F}_\mathcal{I} \to R$ satisfying $\bar{x}_i \mapsto r_i$ for each i. (Indeed, the natural homomorphism $C\{X\} \to R$ given by $x_i \mapsto r_i$ has kernel containing $\mathcal{I}$ and thus induces the desired map.)

(ii) If any other algebra $\tilde{\mathcal{F}}$ generated by $\{\tilde{x}_i : i \in I\}$ satisfies (i) (with $\tilde{x}_i$ in place of $\bar{x}_i$), then $\tilde{\mathcal{F}} \cong \mathcal{F}_\mathcal{I}$. (This is true since (i) is a universal property.)

(iii) If $f(\bar{x}_1, \ldots, \bar{x}_m) = 0$ in $\mathcal{F}_\mathcal{I}$, then $f(x_1, \ldots, x_m) \in \mathcal{I}$. (Indeed, the homomorphism $\psi : \mathcal{F} \to \mathcal{F}_I$ given by $x_i \mapsto \bar{x}_i$ satisfies

$$0 = f(\psi(x_1), \ldots, \psi(x_m)) = \psi(f(x_1, \ldots, x_m)).)$$

(iv) If $\mathcal{I} \lhd_T C\{X\}$, then $\mathcal{I} = \mathrm{id}(\mathcal{F}_\mathcal{I})$ by (iii).

(v) If $\mathcal{V} = \mathcal{V}(\mathcal{I})$ for $\mathcal{I} \lhd_T C\{X\}$, then $\mathcal{I} = \mathrm{id}(\mathcal{V})$ in view of (i) and (iv).

In summary, we have 1:1 correspondences among the following concepts:
1. T-ideals of $C\{X\}$;
2. relatively free C algebras;
3. varieties of C-algebras.

Of course, these definitions depend on the cardinality $|I|$ of the set of indeterminates $\{x_i : i \in I\}$. If $|I| = 1$, then $C\{X\}$ is commutative, and the theory degenerates. There is strong motivation for taking $|I|$ finite, since then the theory of affine PI-algebras becomes available, as described in Appendix 23A. If $|I| = 2$, we can recover the full structure theory by Exercise 1 but at the cost of losing the multilinear identities. One often takes I to be countably infinite in order to be able to multilinearize at will.

From this point of view, we would like to know when two algebras have the same sets of identities.

Definition 23.47. Two algebras R_1 and R_2 are PI-**equivalent** if $\mathrm{id}(R_1) = \mathrm{id}(R_2)$.

In characteristic 0, since, by Exercise 16, all identities can be recovered from their multilinearizations, we can check PI-equivalence using only multilinear identities, and thus refer to Proposition 23.8. The situation requires a more careful analysis in nonzero characteristic.

PROPOSITION 23.48. *If R is an algebra over an infinite field F, and H is any commutative F-algebra, then R is PI-equivalent to $R \otimes_F H$.*

Proof. Clearly R satisfies all identities of $R \otimes H$, since $R \hookrightarrow R \otimes H$ via $r \mapsto r \otimes 1$. It remains to show that every identity f of R remains an identity of $R \otimes H$. Take a base $\{b_i : i \in I\}$ of R over F. Write $f = f(x_1, \dots, x_m)$. Given $a_j \in R \otimes H$, $1 \le j \le m$, we need to show that $f(a_1, \dots, a_m) = 0$. One can show this by reducing to homogeneous components of f, but we take a lovely generic argument due to Amitsur.

Write $a_j = \sum_{k=1}^{t_j} b_{i_k} \otimes c_{jk}$ for $c_{jk} \in H$. Working in the ring $R \otimes F[\Lambda]$, where Λ is the (finite) set of commutative indeterminates $\{\lambda_{jk} : 1 \le j \le m, 1 \le k \le t_j\}$, let $\hat{a}_j = \sum_{k=1}^{t_j} b_{i_k} \otimes \lambda_{jk}$. Then $f(\hat{a}_1, \dots, \hat{a}_m) = \sum b_i \otimes p_i$, where $p_i = p_i(\lambda_{jk})$ are suitable polynomials in the λ_{jk}. But specializing these indeterminates λ_{jk} to arbitrary elements α_{jk} of F sends p_i to elements $p_i(\alpha_{jk}) \in F$, and the corresponding specialization

$$\sum b_i \otimes p_i(\alpha_{jk}) = \sum p_i(\alpha_{jk}) b_i \otimes 1 = f\left(\sum \alpha_{1k} b_{i_k}, \dots, \sum \alpha_{mk} b_{i_k}\right) \otimes 1$$

is 0 since $f \in \mathrm{id}(R)$. Since the b_i comprise a base, the coefficients $p_i(\alpha_{jk}) = 0$ for arbitrary specializations in F. But now Remark 0.7′(ii) of Volume 1 implies that each $p_i = 0$. Hence, specializing λ_{jk} to our original c_{jk} yields

$$f(a_1, \dots, a_m) = \sum b_i \otimes p_i(c_{jk}) = \sum b_i \otimes 0 = 0,$$

as desired. □

Thus, any variety defined over an infinite field F is also closed under extensions of scalars; hence, the theory of varieties often reduces to the case for which F is algebraically closed. However, Proposition 23.48 can fail for algebras over finite fields; cf. Exercises 11 and 12.

Example 23.48′. The most important case of a T-ideal is $\mathrm{id}(M_n(C))$, which we denote as $\mathcal{M}_{n,C}$. Obviously if $\mathrm{id}(R) = \mathcal{M}_{n,C}$, then R has PI-class n.

At this stage, it is convenient to bring in a theorem of Wedderburn [**Row3, Theorem 15 of Appendix B**], to be reproved below as Theorem 24.42: *Every finite division ring is commutative.*

It follows at once from the Wedderburn-Artin Theorem that every finite simple ring is split (cf. Definition 15.24); thus, when studying non-split simple PI-algebras, we may restrict our attention to algebras over an infinite field. Using Proposition 23.48, we now see that any semiprime algebra of PI-class n over a field F belongs to $\mathcal{V}(\mathcal{M}_{n,F})$.

More precisely, any prime algebra that is not PI-equivalent to a matrix algebra over an algebraically closed field has the form $M_n(F)$ for F a finite field! This fact makes PI-theory in characteristic $p > 0$ more palatable, although it still is much more difficult than in characteristic 0.

Relatively free algebras of f.d. algebras.

The relatively free algebra of a variety $\mathcal{V}$ is viewed as the "generic" algebra in $\mathcal{V}$, since many of its elementary properties can be transferred via homomorphic images to every other algebra in $\mathcal{V}$. An important instance of this generic property will be given in Theorem 24.54ff.; we prepare the foundation with an explicit description of the relatively free algebra for $\mathcal{M}_{n,C}$, which has important applications.

Definition 23.49. Take a set $\Lambda = \{\lambda_{jk}^{(i)} : 1 \le j, k \le n,\ i \in I\}$ of commuting indeterminates. Working in $M_n(C[\Lambda])$, we define the **algebra of generic $n \times n$ matrices** $C\{Y\}_n$ to be the C-subalgebra generated by the **generic matrices** $y_i = (\lambda_{jk}^{(i)})$ for $i \in I$; intuitively, each matrix y_i is the most general possible, since its entries are (commuting) indeterminates.

The index set I for the generic matrices can have any cardinality ≥ 2, at our convenience. We start by proving that $C\{Y\}_n$ satisfies the condition of Remark 23.46(iii).

LEMMA 23.50. *If $f(y_1, \dots, y_m) = 0$ in $C\{Y\}_n$, then $f(x_1, \dots, x_m) \in \mathcal{M}_{n,C}$.*

Proof. Specialize the $\lambda_{jk}^{(i)}$ to arbitrary elements of C, thereby specializing the y_i to arbitrary matrices in $M_n(C)$. □

PROPOSITION 23.51. *There is an isomorphism $\varphi : \mathcal{F}_{\mathcal{I}} \to C\{Y\}_n$, where $\mathcal{F}_{\mathcal{I}}$ is the relatively free PI-algebra $\mathcal{F}_{\mathcal{I}}$ with respect to $\mathcal{I} = \mathcal{M}_{n,C}$, given by $\bar{x}_i \mapsto y_i$.*

Proof. Obviously, $\mathcal{I} \subseteq \mathrm{id}(C\{Y\}_n)$, so, in view of Remark 23.46(i), we have a natural surjection φ. It remains to prove that $\ker \varphi = \mathcal{I}$ so that we can apply Noether I (Theorem 1.13 of Volume 1). If $\bar{f} = f(\bar{x}_1, \dots, \bar{x}_m) \in \ker \varphi$, then $f(y_1, \dots, y_m) = 0$; hence $f(x_1, \dots, x_m) \in \mathcal{I}$ by Lemma 23.50, implying $\bar{f} = 0$. □

COROLLARY 23.52. *$C\{Y\}_n$ is a prime ring for any integral domain C.*

Proof. Write $F(\Lambda)$ for the field of fractions of the polynomial algebra $C[\Lambda]$. Viewing $M_n(F[\Lambda]) \subset M_n(F(\Lambda))$, let $W = C\{Y\}_n F(\Lambda)$, a central extension of $C\{Y\}_n$ and an $F(\Lambda)$-subalgebra of $M_n(F(\Lambda))$. But $c_{n^2+1} \notin \mathrm{id}(W)$, so $W = M_n(F(\Lambda))$ by Remark 23.14; hence, $C\{Y\}_n$ is prime, by Remark 16.3′. □

This construction can be generalized.

Example 23.53. A construction of the relatively free algebra of an arbitrary f.d. algebra R. Pick any base $b_1, \dots, b_d$ of R, and working in the polynomial algebra $R[\Lambda]$, where $\Lambda = \{\lambda_{ij} : 1 \le j \le d,\ i \in I\}$, we define the **generic element** $\tilde{b}_i = \sum_{j=1}^{d} \lambda_{ij} b_j$. Let $\widetilde{R}$ be the subalgebra of $R[\Lambda]$ generated by the $\tilde{b}_i$. Clearly, given any $r_i \in R$, we can write $r_i = \sum \alpha_{ij} b_j$, and then define the specialization $\widetilde{R} \to R$ given by $\lambda_{ij} \mapsto \alpha_{ij}$, and thus $\tilde{b}_i \mapsto R_i$. By Remark 23.46(iii), $\widetilde{R}$ is relatively free for $\mathrm{id}(R)$.

Remark 23.54. Since $\widetilde{R}$ is a subalgebra of $R \otimes_F F(\Lambda)$, we conclude that the relatively free algebra of any f.d. algebra is representable.

Specht's problem posed.

Naturally, we are interested in generating T-ideals by means of as few identities as possible. Accordingly, Specht asked (in the case when C is a field F) whether every T-ideal is finitely generated as a T-ideal, or, equivalently, whether $F\{X\}$ satisfies the ACC on T-ideals. In a tour de force, Kemer [**Ke**] answered Specht's problem affirmatively when $\mathrm{Char}(F) = 0$. Put more colloquially, in characteristic 0, any set of identities is a "consequence" of a finite set of identities. For example, the identities of the Grassmann algebra all are consequences of the Grassmann identity $[[x_1, x_2], x_3]$; cf. Exercise 36. The depth of Kemer's result can be measured in part by the fact that we still do not have explicit generators for $\mathcal{M}_{n,\mathbb{Q}}$ for $n > 2$; even the fact that $\mathcal{M}_2(\mathbb{Q})$ is generated as a T-ideal by the obvious candidates s_4 and $[[x_1, x_2]^2, x_3]$ is surprisingly difficult to prove. Specht's conjecture fails in characteristic p; cf. Belov [**Bel**]. Kemer's Theorem is given in full detail in Belov-Rowen [**BelR**], where the proof requires close to 200 pages. In Appendix 23A, we discuss some of the main features of the proof, together with their roots in Shirshov's theory.

Viewing Kemer's Theorem in terms of the ACC on T-ideals, we can bring in Noetherian techniques to PI-theory, replacing ordinary ideals by T-ideals. For example, one sees in this manner that every relatively free algebra has a unique maximal nilpotent T-ideal that is a finite intersection of "T-prime" T-ideals; cf. Exercises 41 and 42.

PI-theory and the symmetric group

There is a quantitative approach to PI-theory, first initiated by Regev in order to prove that the tensor product of PI-algebras is a PI-algebra. We consider V_n, the vector subspace of $F\{X\}$ consisting of multilinear polynomials in $x_1, \dots, x_n$ over a field F. Thus, V_n has a base consisting of the monomials $\{x_{\pi 1} \cdots x_{\pi n} : \pi \in S_n\}$.

Definition 23.55. For any PI-algebra R, we define $\mathcal{I}_n(R) = V_n \cap \mathrm{id}(R)$ and its **codimension** $c_n(R) = \dim_F(V_n/\mathcal{I}_n(R))$.

Remark 23.56. Identifying each monomial $x_{\pi 1} \cdots x_{\pi n}$ with its corresponding permutation π enables us to identify V_n with the group algebra $F[S_n]$, and also as an $F[S_n]$-module, by the action $\sigma \cdot x_{\pi 1} \cdots x_{\pi n} = x_{\sigma\pi 1} \cdots x_{\sigma\pi n}$, yielding

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

(Indeed, writing $f(x_1, \dots, x_n) = \sum \alpha_\pi x_{\pi 1} \cdots x_{\pi n}$ and y_i for $x_{\sigma i}$, we have $y_{\pi i} = x_{\sigma\pi i}$, and thus

$$\begin{aligned} \sigma \cdot f(x_1, \dots, x_n) &= \sum \alpha_\pi \sigma \cdot x_{\pi 1} \cdots x_{\pi n} = \sum \alpha_\pi x_{\sigma\pi 1} \cdots x_{\sigma\pi n} \\ &= \sum \alpha_\pi y_{\pi 1} \cdots y_{\pi n} = f(y_1, \dots, y_n) = f(x_{\sigma 1}, \dots, x_{\sigma n}).) \end{aligned}$$

But then $\sigma \cdot \mathcal{I}_n(R) = \mathcal{I}_n(R)$, so $\mathcal{I}_n(R)$ is an $F[S_n]$-submodule of V_n, and by Maschke's Theorem is a sum of minimal left ideals when $\mathrm{char}(F) = 0$. We have computed these left ideals in Theorem 19.60; namely, they are generated by the idempotents e_T (or equivalently the a_T of Lemma 19.59). Thus, $\mathcal{I}_n(R)$ can be described completely in terms of Young tableaux, and we can compute the identities generating $\mathcal{I}_n(R)$.

Remark 23.56 opens the door to the use of the representation theory of S_n, and in particular Young tableaux, as a tool in the determination of multilinear identities of degree n for a given algebra R; namely, compute the polynomial a_T for each tableau T and evaluate it on R. A second action obtained by reversing Remark 23.56 is described in Belov-Rowen [**BelR**].

The behavior of the codimensions tells us much about PI-algebras. For example, if R satisfies a PI of degree $d < n$, then $c_n(R) < n!$, but various results from the theory of Young tableaux then imply

$$c_n(R) < (d-1)^{2n}. \tag{23.6}$$

This inequality immediately yields Regev's Theorem. Indeed, if R_i satisfies a PI of degree d_i for $i = 1, 2$, then

$$c_n(R_1 \otimes R_2) \leq c_n(R_1)c_n(R_2) < ((d_1 - 1)(d_2 - 1))^{2n} < n!$$

for n large enough, and one concludes that $R_1 \otimes R_2$ is a PI-algebra.

So far we have only utilized the left ideal structure of $F[S_n]$, by identifying a Young tableau with the polynomial arising from its semi-idempotent and checking whether this is an identity of our algebra. By considering two-sided ideals, Kemer and, independently, Amitsur-Regev, proved the following remarkable result:

THEOREM 23.57. *Suppose R satisfies a PI of degree d, and $u, v \in \mathbb{N}$ satisfy*

$$\frac{1}{u} + \frac{1}{v} \leq \frac{2}{e(d-1)^4},$$

where $e = 2.71828\cdots$. Then any multilinear polynomial of a Young tableau whose shape contains a $u \times v$ rectangle is an identity of R.

The proof is in Belov-Rowen [**BelR, Theorem 5.51**]; the main idea is given in Exercises 43–45 in conjunction with the "branching theorem" from Young theory. This is just a taste of the quantitative PI-theory, which is an active area of current research.

Appendix 23A: Affine PI-algebras

In this appendix, we describe the highlights of the combinatoric theory of affine PI-algebras and its role in Kemer's solution of Specht's problem in characteristic 0. Amazingly (and with considerable effort), the main theorems from the affine commutative theory have been carried over to affine PI-algebras. The history is quite interesting. When the pioneers of structure theory turned toward polynomial identities, they asked as a test question whether the Kurosh and Levitzki problems (cf. Appendix 17C) could be solved for affine PI-algebras. This was solved affirmatively by Kaplansky, building on work of Jacobson and Levitzki. In Russia, Shirshov proved a stronger result using combinatorial properties of the free algebra, which led Razmyslov and Schelter independently to the process of embedding a prime affine PI-algebra into one that is finite over an affine center, thereby enabling researchers to utilize the commutative affine theory. We start by presenting Shirshov's theory; note that it also works for algebras without 1.

Shirshov's Theorem.

We start with an arbitrary commutative base ring C. Consider the free affine algebra $C\{x_1, \ldots, x_\ell\}$ in ℓ generators. We work with the monomials, viewed as "words" in the "letters" x_i. We write $|w|$ for the length of a word w, written as a product of letters, and use the crossword dictionary order $\succ$ of Definition 17.9.

Definition 23A.1. A word w is **d-decomposable** if it can be written as a product of subwords $w = w'w_1w_2\cdots w_dw''$ with $w_1 \succ w_2 \succ \cdots \succ w_d$.

Shirshov's idea was to utilize the polynomial identity as follows: Let $R = C\{r_1, \ldots, r_\ell\}$ be a PI-algebra satisfying some multilinear PI

$$f(x_1, \ldots, x_d) = x_1 \cdots x_d + \sum_{1 \neq \pi \in S_d} \alpha_\pi x_{\pi 1} \cdots x_{\pi d}. \tag{23A.1}$$

Write $\overline{w}$ for the image of a word w under the specialization $x_i \mapsto r_i$. Then

$$0 = \overline{x}_1 \cdots \overline{x}_d + \sum_{1 \neq \pi \in S_d} \alpha_\pi \overline{x}_{\pi 1} \cdots \overline{x}_{\pi d},$$

which yields a reduction procedure whereby any d-decomposable word

$$w = w'w_1w_2\cdots w_dw''$$

can be replaced by

$$-\sum_{1 \neq \pi \in S_d} \alpha_\pi w'w_{\pi 1} \cdots w_{\pi d}w'', \tag{23A.2}$$

which by hypothesis is a linear combination of smaller words. Since we make no claims about the uniqueness of this reduction procedure, we do not need to worry about the word problem or other concerns raised in Appendix 17B. One observation is immediate.

LEMMA 23A.2. *The image $\overline{w}$ in R of any word w can be rewritten as a linear combination of images of d-indecomposable words, each of which is obtained by permutating the letters of w.*

Proof. Either w is d-indecomposable, in which case we are done, or else one rewrites $\overline{w}$ via (23A.2) and applies induction to each term, noting that there are only finitely many permutations of the letters of w. □

The key to Shirshov's theory is a purely combinatorial result.

THEOREM 23A.3 (SHIRSHOV'S DICHOTOMY LEMMA). *For any ℓ, d, k, there is $\beta = \beta(\ell, d, k) \in \mathbb{N}$ such that any word w of length $\geq \beta$ in ℓ letters is either d-decomposable or contains a repeating subword of the form u^k with $1 \leq |u| \leq d$.*

Several proofs are known. The idea of Shirshov's original proof is clear enough, but hard to implement because of its notational complexity. The proof given here is taken from Belov-Rowen [**BelR**].

Remark 23A.3′. The word w is d-decomposable if it contains a (lexicographically) descending chain of d disjoint subwords of the same length; i.e., if w is of the form

$$s_0 v_1 s_1 v_2 \cdots s_{d-1} v_d s_d$$

where $v_1 \succ v_2 \succ \cdots \succ v_d$ and $|v_1| = \ldots = |v_d|$. (Indeed, define $w_i = v_i s_i$ in Definition 23A.1.)

Thus, we examine subwords of equal length inside long enough words. To guarantee enough room, we return to hyperwords; cf. Definition 17.51. We say that a word v is **unconfined** in a hyperword h if v occurs infinitely many times as a subword of h, and v is **confined** in h if v occurs only finitely many times as a subword of h. Also, recall that $\nu_t(h)$ denotes the number of distinct words of length t that appear as subwords in a given hyperword h.

PROPOSITION 23A.4. *If a hyperword h has fewer than m unconfined subwords of length m, then $h = vu^\infty$ for suitable words u, v, with $|u| < m$.*

Proof. Since $\nu_m(h) < \ell^m < \infty$, h must have a (finite) initial subword w containing all the confined subwords of length m. Writing $h = wh'$, we see by assumption that the only subwords of length m that survive in h' are unconfined; i.e., $\nu_m(h') < m$. But obviously $\nu_1(h') \geq 1$, so in view of Remark 17.53, there is some $t \leq m$ for which $\nu_t(h') = \nu_{t+1}(h') < m$; Lemma 17.54 says that h', and thus h, has the desired form vu^∞. □

Following Remark 23A.3′, we say that a hyperword is **d-decomposable** if it contains a (lexicographically) descending chain of d disjoint subwords of the same length.

THEOREM 23A.5. *Any hyperword h is either d-decomposable or has the form vu^∞ for some initial subword v and some subword u with $|u| < d$.*

Proof. If h has $< d$ unconfined subwords of length d, then we are done by Proposition 23A.4. Thus, we may assume that h has d unconfined subwords $v_1, \ldots, v_d$ of length d, which we list in decreasing lexicographic order.

(This is possible since they have the same length.) We take the first occurrence of v_1 and then take the first occurrence of v_2 that starts after this occurrence of v_1 ends; proceeding inductively, we obtain a subword of h in which the subwords $v_1, \ldots, v_d$ appear disjointly in that order. Hence, by definition, h is d-decomposable.

Proof of Shirshov's Dichotomy Lemma. A variant of the König graph theorem (Proposition 17.25), which we spell out. Theorem 23A.5 says that for any hyperword h on ℓ letters there is a number $\beta(\ell, d, k, h)$ for which the initial d-indecomposable subword of length $\geq \beta(\ell, d, k, h)$ contains a nonempty word u^k with $|u| \leq d$. We claim that $\beta(\ell, d, k, h)$ is bounded as h runs over all hyperwords on ℓ letters. Otherwise, for any number β, let

$$P_{j,\beta} = \{\text{Initial subwords of length } j \text{ of hyperwords } h \text{ with } \beta(\ell, d, k, h) > \beta\},$$

and for any word v of length j belonging to some $P_{j,\beta}$, let $\gamma(v)$ denote the largest β' for which $v \in P_{j,\beta'}$. If $v \in P_{j,\beta'}$ for all β', we write $\gamma(v) = \infty$. Let $Q_j = \{v : \gamma(v) = \infty\}$.

Since each $P_{j,\beta}$ is finite nonempty this means $Q_j \neq \emptyset$. If $v_j \in Q_j$, then we can find v_{j+1} in Q_{j+1} starting with v_j, so we continue this process by induction to obtain a hyperword h. But the initial subword of length $\beta(\ell, d, k, h)$ of this hyperword h already contains a subword u^k with $|u| \leq d$, contrary to hypothesis on the v_j for $j = \beta(\ell, d, k, h)$. We conclude by taking $\beta(\ell, d, k)$ to be our bound on the $\beta(\ell, d, k, h)$. □

Having Shirshov's Dichotomy Lemma under our belt, we introduce a second reduction procedure — namely, if $\bar{u}$ is integral of degree k, say $\bar{u}^k = \sum_{i=0}^{k-1} \alpha_i \bar{u}^i$, then we may replace u^k by the corresponding combination $\sum_{i=0}^{k-1} \alpha_i u^i$ of words of smaller length. Combining these two reduction procedures easily yields:

THEOREM 23A.6 (SHIRSHOV'S FIRST THEOREM). *If $R = C\{r_1, \ldots, r_\ell\}$ satisfies a PI of degree d, and each word in the r_i of length $\leq d$ is integral (say of degree $\leq k$), then R is f.g. as a C-module by $\leq \beta(\ell, k, d)$ elements.*

THEOREM 23A.7. *If an affine algebra R without 1 satisfies a PI of degree d and each word in the generators of length $\leq d$ is nilpotent, then R is nilpotent.*

Thus, we see some of the power of Shirshov's Dichotomy Lemma. There is an even stronger version, called Shirshov's Height Theorem, which is crucial to the study of growth of PI-algebras, implying at once that any affine PI-algebra has polynomially bounded growth. See Belov-Rowen [**BelR**] for a proof of this theorem and a detailed discussion of growth in PI-algebras.

Representable affine algebras revisited.

Shirshov's First Theorem has a key application to representable algebras.

Example 23A.8 (The Characteristic Closure). Suppose the affine algebra $R = F\{r_1, \dots, r_\ell\}$ is representable. Viewing R in $M_n(K)$, we see that each word in the r_i of length $\leq d$ satisfies its characteristic polynomial, so adjoining the finite number ($t = n\ell^d$) of these characteristic coefficients (from K) to F and to R, respectively, gives us respective rings $\hat{F} \subset \hat{R}$ such that, in $\hat{R}$, every word of length $\leq d$ in the r_i is integral over $\hat{F}$. By Shirshov's First Theorem, $\hat{R}$ is f.g. as an $\hat{F}$-module. The structure of $\hat{R}$ is rather nice; since $\hat{F}$ is a commutative affine domain and thus Noetherian, we see that $\hat{R}$ is also Noetherian.

When R has PI-class n (for example, when R is semiprime), we get more by appealing to alternating central polynomials:

Proposition 23A.9. *Let $t = n^2$. Suppose h_n is the t-alternating, central polynomial of Corollary 23.27. We view $a \in M_n(F)$ as a $t \times t$ matrix via the regular representation, and write the Hamilton-Cayley polynomial of a as $\lambda^t - \sum_{u=1}^{t} \alpha_u \lambda^{t-u}$. For any $r_1, \dots, r_d$ in $M_n(F)$ and any $1 \leq u \leq t$, we have*

$$\sum_{\substack{j_1+\cdots+j_t=u;\\ 0\leq j_i\leq 1}} h_n(a^{j_1}r_1, \dots, a^{j_t}r_n, r_{n+1}, \dots, r_d) = \alpha_u h_n(r_1, \dots, r_d). \tag{23A.3}$$

Proof. Applying Proposition 23.15 to the transformation $\tilde{a} = \lambda I - a$ yields

$$\begin{aligned} h_n(\tilde{a}r_1, \dots, \tilde{a}r_n, r_{n+1}, \dots, r_d) &= \det(\tilde{a}) h_n(r_1, \dots, r_d) \\ &= \sum_{u=1}^{t} \alpha_u \lambda^{t-u} h_n(r_1, \dots, r_d), \end{aligned}$$

and we get (23A.3) by matching coefficients of λ. □

Corollary 23A.10. *If $A = Rh_n(R) \lhd R$, then also $A \lhd \hat{R}$.*

In other words, R and $\hat{R}$ have a common nonzero ideal, so much of the structure of $\hat{R}$ can be passed to R. Using this technique, Schelter showed for example that affine PI-algebras are catenary.

Since we can handle representable affine algebras readily enough, we can start to extend the theory to an arbitrary affine PI-algebra R. One major step in this direction, outside the scope of this treatment, is the theorem of Braun-Kemer-Razmyslov that $\mathrm{Jac}(R)$ is nilpotent. (Amitsur had shown that $\mathrm{Jac}(R)$ is locally nilpotent.)

Kemer's solution of Specht's conjecture in characteristic 0

In the next few pages we sketch Kemer's solution to Specht's problem, which takes up all of Kemer [**Ke**] and most of Belov-Rowen [**BelR**].

Kemer's Finite-Dimensionality Theorem.

The first main step in Kemer's solution is the following striking theorem:

Theorem 23A.11 (Kemer). *Any affine PI-algebra over a field F of characteristic 0 is PI-equivalent to a finite-dimensional algebra.*

The main reason why F is assumed to have characteristic 0 is that, as mentioned earlier, all identities in characteristic 0 are determined by multilinear identities. This enables us to use the representation theory of S_n in conjunction with combinatoric techniques such as Shirshov's Theorem, as well as to assume that F is algebraically closed, since varieties over infinite fields are not affected by tensor extensions.

The proof of Theorem 23A.11 is much too intricate to present here in detail. It starts with a theorem of Lewin implying that any T-ideal $\mathcal{I}$ contains $\mathrm{id}(R)$ for a suitable finite-dimensional algebra R. The goal is to modify R, enlarging its set of identities until $\mathrm{id}(R) = \mathcal{I}$. This is achieved by examining alternating polynomials on finite-dimensional algebras. Here is the main idea:

Since the base field F is assumed to be algebraically closed, Wedderburn's Principal Theorem (Theorem 15.26) says that $R = \bar{R} \oplus J$, where $J = \mathrm{Jac}(R)$ is nilpotent and $\bar{R} \cong R/J$ is semisimple. Thus, any evaluation of a multilinear polynomial $f(x_1, \dots, x_d)$ on R is the sum of evaluations $f(r_1, \dots, r_d)$ where each $r_i \in \bar{R} \cup J$. We say that r_i is a **semisimple** (resp. **radical**) specialization if $r_i \in \bar{R}$ (resp. $r_i \in J$). Lifting sets of orthogonal idempotents from $\bar{R}$ to R and applying the Peirce decomposition (Exercise 13.9), Kemer managed to describe $\dim_F \bar{R}$ and the nilpotence index of J in terms of properties of semisimple and radical substitutions in alternating polynomials. This gives us a pair of numbers (β, γ) called the **Kemer index** of the T-ideal. Since the Kemer index only depends on polynomials. The proof of Theorem 23A.11, based on induction on the Kemer index, involves an intricate manipulation of alternating polynomials.

Corollary 23A.12. *The relatively free algebra of any affine PI-algebra of characteristic 0 is representable.*

Proof. Follows at once from Theorem 23A.11 and Remark 23.54. □

Theorem 23A.11 also implies at once that every affine PI-algebra satisfies a Capelli identity. (To be fair, this result is used en route to proving

the Braun-Kemer-Razmyslov Theorem, which is a prerequisite to Kemer's Theorem.)

It is not difficult to deduce from Theorem 23A.11 that the free affine algebra in characteristic 0 satisfies the ACC on T-ideals. However, passing to non-affine PI-algebras requires another sublime idea.

Superidentities and Kemer's correspondence.

Remark 23A.13. We can grade the free algebra by labelling some indeterminates even and the others odd; we rename the even indeterminates y_i and the odd indeterminates z_i, thereby yielding the **free superalgebra** $F\{Y, Z\}$ whose elements are called **superpolynomials**. This gives rise to a theory of **superidentities** of any superalgebra A, where the y_i (resp. z_i) are specialized to even (resp. odd) elements of A.

Example 23A.14. Recall that the Grassmann algebra $G = G_0 \oplus G_1$ is a superalgebra. The superpolynomials $y_1z_1 - z_1y_1$ and $y_1y_2 - y_2y_1$ are superidentities of G, since all even elements are central. Likewise, $z_1z_2 + z_2z_1$ is a superidentity of G. We wrote this all more concisely in Equation (18.9); this property is called **supercommutativity**.

Remark 23A.15. If R and T are superalgebras, then $R \otimes T$ is a superalgebra under the grading

$$(R \otimes T)_0 = R_0 \otimes T_0 + R_1 \otimes T_1; \qquad (R \otimes T)_1 = R_0 \otimes T_1 + R_1 \otimes T_0.$$

In particular, if we view an ungraded algebra R trivially as a superalgebra with $R_0 = R$ and $R_1 = 0$, then $R \otimes G$ is a superalgebra with $(R \otimes G)_i = R \otimes G_i$.

Remark 23A.15 yields a functor from varieties to supervarieties, but to see this, we must understand the passage from identities to superidentities. Given a superalgebra R, we write $\mathrm{id}_2(R)$ to denote the set of superidentities of R. $\mathrm{id}_2(R)$ is a T_2-ideal of $F\{Y, Z\}$ in the sense that it is invariant under any endomorphism of $F\{Y, Z\}$ that sends Y to even elements and Z to odd elements.

We recall the standard base B of the Grassmann algebra G from (23.2).

Definition 23A.16. Given an n-tuple $\mathbf{b} = (b_1, \dots, b_n)$ where each $b_i \in B$, we write $\mathbf{b}^-$ for $\{i : b_i \text{ is odd}\}$.

The following observation is immediate from Example 23A.14.

Remark 23A.17. Given $I \subset \{1, \dots, n\}$ and $\pi \in S_n$, there is a number $\varepsilon(\pi, I) = \pm 1$ such that for any $\mathbf{b} = (b_1, \dots, b_n)$ with $\mathbf{b}^- = I$,

$$b_{\pi 1} \cdots b_{\pi n} = \varepsilon(\pi, I) b_1 \cdots b_n.$$

We are ready for a key definition.

Definition 23A.18 (The Grassmann involution on polynomials). Given a multilinear polynomial

$$f = f(x_1, \dots, x_n) = \sum_{\pi \in S_n} \alpha_\pi x_{\pi 1} \cdots x_{\pi n}$$

and $I \subseteq \{1, \dots, n\}$, define

$$f_I = \sum_{\pi \in S_n} \varepsilon(\pi, I) \alpha_\pi x_{\pi 1} \cdots x_{\pi n}.$$

Given $f(x_1, \dots, x_n)$ and $I \subseteq \{1, \dots, n\}$, define $f_I{}^*$ to be the superpolynomial obtained by substituting z_i for x_i whenever $i \in I$ and substituting y_i for x_i whenever $i \notin I$.

THEOREM 23A.19. *For any PI-algebra R, the following assertions are equivalent for any multilinear polynomial*

$$f = f(x_1, \dots, x_n) = \sum_{\pi \in S_n} \alpha_\pi x_{\pi 1} \cdots x_{\pi n} :$$

(1) $f \in \mathrm{id}(R)$;
(2) $f_I{}^* \in \mathrm{id}_2(R \otimes G)$ *for some subset* $I \subseteq \{1, \dots, n\}$;
(3) $f_I{}^* \in \mathrm{id}_2(R \otimes G)$ *for every subset* $I \subseteq \{1, \dots, n\}$.

Proof. Both directions (1) $\Rightarrow$ (3) and (2) $\Rightarrow$ (1) of the proof easily follow from the following equality:

$$f_I(r_1 \otimes b_1, \dots, r_n \otimes b_n) = f(r_1, \dots, r_n) \otimes b_1 \cdots b_n \tag{23A.5}$$

whenever $r_1, \dots, r_n \in R$ and $b_1, \dots, b_n \in B$ satisfy $(b_1, \dots, b_n)^- = I$. Equation (23A.5) holds since

$$\begin{aligned} f_I(r_1 \otimes b_1, \dots, r_n \otimes b_n) &= \sum_\pi \varepsilon(\pi, I) \alpha_\pi (r_{\pi 1} \otimes b_{\pi 1}) \cdots (r_{\pi n} \otimes b_{\pi n}) \\ &= \sum_\pi \varepsilon(\pi, I) \alpha_\pi r_{\pi 1} \cdots r_{\pi n} \otimes b_{\pi 1} \cdots b_{\pi n} \\ &= \left(\sum_\pi \alpha_\pi r_{\pi 1} \cdots r_{\pi n} \right) \otimes b_1 \cdots b_n \\ &= f(r_1, \dots, r_n) \otimes b_1 \cdots b_n \end{aligned}$$

□

COROLLARY 23A.20. *Suppose* $\mathrm{char}(F) = 0$. *There is an inclusion-preserving map* $\Phi : \{T\text{-ideals}\} \to \{T_2\text{-ideals}\}$ *given by* $\mathrm{id}(R) \to \mathrm{id}_2(R \otimes G)$.

Better yet, Φ can be viewed as a functor from varieties to supervarieties.

The Grassmann envelope.

Kemer's next observation was to find an inverse functor, thereby elevating the Grassmann algebra to a position of prominence in PI-theory.

Definition 23A.21. The **Grassmann envelope** of a superalgebra R is the following subalgebra of $R \otimes G$:

$$\mathcal{G}(R) = (R \otimes G)_0 = R_0 \otimes G_0 \oplus R_1 \otimes G_1.$$

THEOREM 23A.22. *Let R be a PI-superalgebra, and let $f = f(x_1, \dots, x_n) = \sum_{\pi \in S_n} \alpha_\pi x_{\pi 1} \cdots x_{\pi n}$. Then $f \in \mathrm{id}(\mathcal{G}(R))$, iff $f_I{}^* \in \mathrm{id}_2(R)$ for every subset $I \subseteq \{1, \dots, n\}$.*

Proof. To check whether $f \in \mathrm{id}(\mathcal{G}(R))$, we check substitutions $a \otimes b$, where $a \in R$ and $b \in G$ are both homogeneous of the same parity; the argument is just as in the proof of Theorem 23A.19, so we leave the details as an exercise. □

THEOREM 23A.23. *There is a correspondence Ψ from $\{$varieties of superalgebras$\}$ to $\{$varieties of algebras$\}$ given by $R \mapsto \mathcal{G}(R)$, which is the inverse of Φ of Corollary 23A.20 in the sense that*

$$\mathrm{id}(R_0) = \mathrm{id}(\mathcal{G}(R_0 \otimes G)) \tag{23A.6}$$

and

$$\mathrm{id}_2(R) = \mathrm{id}_2(\mathcal{G}(R) \otimes G). \tag{23A.7}$$

Proof. We have proved everything except that Φ and Ψ are inverses, i.e., the two displayed equations. But

$$\mathcal{G}(R_0 \otimes G) \cong R_0 \otimes (G_0 \otimes G_0) \oplus R_0 \otimes (G_1 \otimes G_1) \cong R_0 \otimes (G_0 \otimes G_0 \oplus G_1 \otimes G_1),$$

and $(G_0 \otimes G_0) \oplus (G_1 \otimes G_1)$ is a commutative algebra, yielding (23A.6) in view of Proposition 23.48.

To see (23A.7) one notes by Theorem 23A.22 that $f_I{}^* \in \mathrm{id}_2(\mathcal{G}(R) \otimes G)$ iff $f \in \mathrm{id}(\mathcal{G}(R)) = \mathrm{id}(R_0 \otimes G_0 + R_1 \otimes G_1)$, iff $f_I{}^* \in \mathrm{id}_2(R)$. □

Conclusion of Kemer's solution to Specht's problem.

One starts over again from Theorem 23A.11, this time with superalgebras, developing the theory of superidentities in parallel to the theory of identities (although the proofs are twice as complicated, since one must keep track of even and odd elements separately). In particular, the theory of representations of the symmetric group is applicable to superidentities, and also Kemer's finite-dimensionality theorem can be modified to show that any affine superalgebra satisfies the same superidentities as a suitable finite-dimensional superalgebra.

We need one more piece to assemble the proof of Kemer's solution to Specht's problem. This is an argument that arbitrary superidentities follow from superidentities of affine algebras. The proof relies on a certain asymmetry in Young tableaux and is applicable only in characteristic 0. (In fact, this may well be the only portion of the proof which is unsalvageable in nonzero characteristic.) Indeed, Theorem 23.57 implies that any nonidentity of a superalgebra R can be described in terms of a Young tableau lying in a certain infinite L-shape, by which we mean that the heights of all columns far enough out to the right are bounded, as are the widths of all rows far enough down. We call the infinite horizontal strip the **arm** and the infinite vertical strip the **leg**. Since the arm and leg of the L-shape can extend indefinitely far, one would imagine that infinitely many indeterminates are required to describe the identity. But the following trick enables us to make the number of indeterminates finite.

Suppose the leg has width d (eventually). Since the horizontal rows in this part of the leg correspond to polynomials of the form $\sum_{\pi \in S_d} x_{\pi 1} \cdots x_{\pi d}$, each row can be viewed as the multilinearization of x^d, as in Exercise 23.10. Once the width of the leg reaches d, we repeat this indeterminate as far as we want. Unfortunately no such trick works for the arm, since the columns correspond to standard polynomials and cannot be obtained via linearization. But we can apply the Grassmann involution to eliminate the appearances of $\operatorname{sgn} \pi$ and thus make Exercise 23.10 applicable. We conclude that any *superalgebra* satisfies the same superidentities as a suitable affine superalgebra, and thus of a finite-dimensional superalgebra. Hence, we obtain the ACC on T_2-ideals, which by Theorem 23A.23 yields the ACC on T-ideals.

Some reflections in view of the success of Kemer's correspondence and the Grassmann envelope: First of all, just as we derived superidentities via the Grassmann envelope, one can "superize" many concepts and results from the structure theory by translating from a superalgebra R to its Grassmann envelope. For example, R is supercommutative iff $\mathcal{G}(R)$ is commutative, and there is a rich theory of supercommutative algebra.

Secondly, since Kemer had to pass to superalgebras to solve Specht's problem, the T-ideal theory may be handled more easily by studying T_2-ideals of superalgebras and then translating back. For example, Kemer noted that the T-ideals corresponding to the T_2-ideals of graded-simple superalgebras turn out to be the **T-prime** T-ideals (in characteristic 0) in the sense that the product of two nonzero T-ideals is nonzero. As an illustration, $\mathrm{id}(G)$ is a T-prime T-ideal in this sense. The complete list of T-prime ideals of $F\{X\}$ (for $\mathrm{char}(F) = 0$) consists of the sets $\mathrm{id}(R)$, where R has the form $M_n(F)$, $M_n(G)$, or one of the superalgebras of Exercise 23.41.

Appendix 23B: Identities of nonassociative algebras

Having already introduced identities of nonassociative algebras in Appendix 21B, let us merge that discussion with the associative theory described in this chapter. The main initial difference is that, lacking associativity, we must work in the free nonassociative algebra $\mathcal{F}$ of Definition 21B.3 and use parentheses to indicate in what order we multiply. Since there are so many different ways of placing parentheses in multiplication, it is difficult to compute with words of the free nonassociative algebra. Nevertheless, if we keep track of the parentheses, the multilinearization process works in analogy to the associative case, and in fact we have already applied this idea to various nonassociative identities, such as anticommutativity of Lie algebras and the alternative identities. In characteristic 0, analogously to the associative theory, the identities of a nonassociative algebra are determined by its multilinear identities.

Nonassociative algebras often are studied from the point of view of varieties. Associative algebras themselves comprise a variety (defined by the associator identity of Example 21B.5(ii)); varieties of nonassociative algebras include Lie algebras, alternative algebras, and Jordan algebras. Any variety $\mathcal{V}$ of nonassociative algebras is defined in terms of the set of its identities $\mathcal{I} = \mathcal{I}(\mathcal{V}) \subset \mathcal{F}$, which is a nonassociative version of T-ideal and gives rise to the analogous **relatively free** nonassociative algebra $\mathcal{F}_{\mathcal{I}} = \mathcal{F}/\mathcal{I}$.

Example 23B.1. (i) The relatively free algebra with respect to the associative identity is just the free associative algebra.

(ii) The **free Lie algebra** $\mathcal{FL}$ is defined as the relatively free nonassociative algebra with respect to the identities of Definition 21.1 (writing $[fg]$ for the product. Thus, for anticommutativity, we need to factor out all $[ff]$; to get the Jacobi identity, we need to factor out all $[f[gh]] + [g[hf]] + [h[fg]]$.

(iii) The **free Jordan algebra** $\mathcal{FJ}$, defined by taking the free nonassociative algebra and factoring out the Jordan identities, plays a leading role in Jordan theory, taking up all of Chapter 3 in Jacobson [**Jac6**].

(iv) The **free special Jordan algebra** $\mathcal{FSJ}$ is defined as $F\{X\}^+$, where $F\{X\}$ denotes the free associative algebra. There is a natural map $\mathcal{FJ} \to \mathcal{FSJ}$ given by $x_i \mapsto x_i$; its kernel K is comprised precisely of those Jordan polynomials that are identities of every special Jordan algebra. (Indeed, for any special Jordan algebra $J = R^+$, the natural homomorphism $F\{X\} \to R$ induces a natural homomorphism $\mathcal{FSJ} \to R^+$.)

Recall that $K \neq 0$ from Exercise 21B.37 (which is used to show that the Albert algebra is not special). Nevertheless, Shirshov proved that K does not contain any 2-variable polynomials, and MacDonald extended this to 3-variable polynomials that are linear in x_1. In particular, any Jordan algebra generated by two elements is special.

Central polynomials (in particular, the analog to Hall's polynomial) play a key role in the theories of alternative and Jordan algebras; cf. Exercise B1.

Identities of Lie algebras and the Restricted Burnside Problem

Recall the Restricted Burnside Problem (RBP) from Appendix 17C: Does $B(m, n)$, the free group of exponent n generated by m elements, have only finitely many homomorphic images? In view of the Hall-Higman reduction, we may assume that n is a prime power.

Our aim here is to develop enough theory to be able to describe Zelmanov's solution (cf. Zelmanov [**Ze2**]), although we do not have enough room for the computational details. We start with some identities of Lie algebras, which then are exploited by means of constructing a Lie algebra from a p-group. In an effort to make the notation manageable, we start with some observations about **Lie words** in the free Lie algebra $\mathcal{FL}$ in the alphabet $X = \{x_0, x_1, \dots\}$; these are formed by means of repeated Lie multiplication, such as $[[x_1x_2][x_2x_4]]$.

LEMMA 23B.2. *Any Lie word w linear in $x_0, x_1, x_2, \dots, x_n$ can be written as a sum of words of the form*

$$[[\cdots[[x_0x_{\pi 1}]x_{\pi 2}]\cdots]x_{\pi n}] \tag{23B.1}$$

for some permutation $\pi \in S_n$.

Proof. By induction on n, since the assertion is clear for $n = 1$ by anticommutativity. Suppose $w = [w_1w_2]$, where w_1, w_2 also are Lie words. Replacing $[w_1w_2]$ by $-[w_2w_1]$ if necessary, we may assume that x_0 appears in w_1. By induction, w_1 has the form (23B.1); we are done unless w_2 has length > 1 and thus has the form $[w_2'w_2'']$ for words w_2', w_2'' of shorter length.

Now we claim the result by a secondary induction on $|w_2|$. Indeed, Jacobi's identity shows that

$$[w_1w_2] = [w_1[w_2'w_2'']] = [[w_1w_2']w_2''] - [[w_1w_2'']w_2'],$$

each of which can be reduced to (23B.1) by the secondary induction. □

In view of Lemma 23B.2, we write $[a_0a_1 \dots a_n]$ for $[\cdots[[a_0a_1]\cdots]a_n]$. Likewise, $[a, b^n]$ denotes $[a, b, \dots, b]$, where b appears n times. We have proved that any multilinear polynomial can be rewritten in the form

$$\sum_{\pi \in S_n} \alpha_\pi [x_0 x_{\pi 1} \cdots x_{\pi n}]. \tag{23B.2}$$

Writing X_i for ad_{-x_i} so that $X_i x_0 = [x_0, x_i]$, we can rewrite (23B.2) as $f(x_0)$, where

$$f = \sum_{\pi \in S_n} \alpha_\pi X_{\pi n} \cdots X_{\pi 1}. \tag{23B.3}$$

A Lie identity written in the form (23B.3) is called an **ad-identity** of the Lie algebra L. For example, L is Lie nilpotent of index $k+1$ iff $X_1 \cdots X_k$ is an ad-identity of L.

The study of multilinear identities of Lie algebras is thereby reduced to ad-identities, giving the theory an associative flavor since ad-identities are described in the associative algebra $\mathrm{End}_C L$.

Definition 23B.3. Write X^n for $(\mathrm{ad}_{-x})^n$. The **n-Engel identity** e_n is X_1^n, i.e., $[x_0, x_1, \dots, x_1]$, with x_1 applied n times.

The **multilinearized n-Engel identity** $\tilde{e}_n$ is $\sum_{\pi \in S_n} X_{\pi 1} \cdots X_{\pi n}$.

A natural analog to Kurosh's Problem for Lie algebras is:

Engel Problem. If a Lie algebra L satisfies the n-Engel identity e_n for some n, then is L Lie nilpotent?

See Theorem 21.27, where the Engel problem was answered positively for the case that L is a linear Lie algebra. Furthermore, in view of the discussion in Appendix 23A, this problem has a positive answer when $L = R^-$ for R an associative affine algebra, but the situation is more mysterious in the general case.

Example 23B.4 (Solution of the Engel Problem for $n = 2$). Suppose L satisfies the ad-identity X^2 and thus also $\tilde{e}_2 = X_2X_1 + X_1X_2$, by linearization. Applying Jacobi's identity and anticommutativity (repeatedly) yields

$$\begin{aligned} 0 &= (X_3X_1 + X_1X_3)(x_2) = [x_2x_1x_3] + [x_2x_3x_1] \\ &= -([x_1x_3x_2] + [x_3x_2x_1]) - [x_3x_2x_1] \\ &= [x_3x_1x_2] - 2[x_3x_2x_1] \\ &= (X_2X_1 - 2X_1X_2)(x_3); \end{aligned}$$

so $X_2X_1 - 2X_1X_2$ is also an ad-identity, thereby yielding the ad-identity $3X_1X_2$. Thus, for $\mathrm{char}(F) \neq 3$, L is Lie nilpotent of class ≤ 3. A different argument fills out the picture in characteristic 3; cf. Exercise B4.

The same sort of argument, although more complicated, shows that every Lie algebra of characteristic > 5 satisfying e_3 satisfies the ad-identity $X_1^2X_2^2$, and thus is Lie nilpotent of class ≤ 6. Zelmanov solved the Engel Problem in characteristic 0 (cf. Kostrikin [**Ko, Chapter 6**]), but the answer in general is negative; the most direct counterexample, due to P.M. Cohn, is found in the same reference.

Nevertheless, there are positive results in the case when L is f.g. (i.e., finitely generated as a Lie algebra). (Note that f.g. Lie algebras are a wide generalization of f.d. Lie algebras studied in Chapter 21.)

THEOREM 23B.5 (KOSTRIKIN-ZELMANOV). *If L is a f.g. Lie algebra over a field of characteristic p and satisfies e_{p-1}, then L is Lie nilpotent.*

For arbitrary n, we need a harder result about restricted Lie algebras; cf. Definition 21.21′.

THEOREM 23B.6 (ZELMANOV). *Suppose L is a f.g. restricted Lie algebra over a field of characteristic p, and L satisfies e_n and all of its partial linearizations. Then L is Lie nilpotent.*

We delay our discussion of the proofs of Theorems 23B.5 and 23B.6 because of their difficulty. Here is an immediate consequence:

COROLLARY 23B.7. *Suppose L is as in Theorem 23B.6, and R is an associative enveloping algebra (without 1) of L satisfying $a^n = 0$ for all $a \in L$. Then R is nilpotent.*

Proof. By Theorem 23B.6, L is Lie nilpotent of some index k. Hence, L contains a central Lie ideal, which thus generates a nilpotent ideal $N \lhd R$. Passing to R/N, we conclude by induction on k. □

Unfortunately, as we shall see, Zelmanov's Theorem still is not enough to yield the RBP, but he also obtained the following stronger result, which does suffice. Suppose the a Lie algebra L is generated by a given set $S = \{u_i : i \in I\}$. We define the **underlying Lie monoid** $L_0 = L_{0,S}$ to be the subset of L consisting of all Lie words evaluated on the u_i. Thus, L_0 contains $[u_1u_2]$ but not necessarily $u_1 + u_2$. The **weak Engel condition** $e_{S,n}$ is that $\mathrm{ad}_a^n = 0$ for all $a \in L_0$.

Clearly e_n implies $\tilde{e}_n$ and $e_{S,n}$, but not conversely.

THEOREM 23B.8. *Suppose L is a f.g. Lie algebra over $\mathbb{Z}/p$. If L satisfies the multilinearized Engel identity $\tilde{e}_m$ and the weak Engel condition $e_{S,n}$ for some m, n and some finite generating set S, then L is (Lie) nilpotent.*

This theorem sounds rather technical, but the reason we need to weaken the hypotheses will become clear.

Definition 23B.9. A subset S of a nonassociative algebra A is **nilpotent** of index k if **every** product of k elements of S is 0, regardless of the placement of the parentheses. A is **locally nilpotent** if any finite subset of A is nilpotent.

It follows easily from Zorn's Lemma that any algebra A has a maximal locally nilpotent ideal N. We would like N to be unique in the sense that it contains every locally nilpotent ideal, and also we would like A/N not to contain any nonzero locally nilpotent ideal. This is false even for f.d. Lie algebras; cf. Exercise 21.33. The theory runs quite smoothly in the presence of an Engel identity.

LEMMA 23B.10. *If a Lie algebra $L = N + Fa$, where $N \lhd L$ is locally nilpotent and $a \in L$ is* ad*-nilpotent, then L is locally nilpotent.*

Proof. First, recall from Remark 21.7′(ii) that $\mathrm{ad}\, a$ acts as a derivation on N; by hypothesis, $\mathrm{ad}_a^n = 0$ for some n. We need to show that any finite subset $S = \{c_1+\alpha_1 a, \dots, c_m+\alpha_m a : c_i \in N,\ \alpha_i \in F\}$ in L is nilpotent. Let S_0 be the set of nonzero evaluations of Lie words in $c_1, \dots, c_m$; by hypothesis, S_0 is finite and clearly spans a nilpotent Lie subalgebra of N. Let $\hat{S}_0 = \{\mathrm{ad}_a^j(s) : s \in S_0,\ 0 \le j < n\}$, a finite set that spans a Lie subalgebra N_0 of N, on which ad_a acts as a derivation. By hypothesis, N_0 is Lie nilpotent of some index k; thus any evaluation of a Lie word of length k in $\hat{S}_0$ is 0.

Now consider any nonzero evaluation of a Lie word w in $S_0 \cup \{a\}$. There must be at most k substitutions in elements of S_0, since we can incorporate the appearances of $\mathrm{ad}\, a$ into words evaluated in $\hat{S}_0$. Thus, in view of Lemma 23B.2, any nonzero evaluation is a sum of expressions of the form

$$[s_1, a, \dots, a, s_2, a, \dots, a, \dots, s_{k-1}, a, \dots, a].$$

But there can be at most $n-1$ consecutive occurrences of a, since $\mathrm{ad}_a^n = 0$, implying that the longest possible nonzero evaluation has length $(k-1)n$; hence, S generates a nilpotent Lie algebra of index $\le (k-1)n+1$. □

PROPOSITION 23B.11. *Any Lie algebra L satisfying e_n has a unique maximal locally nilpotent Lie ideal N, called the* **locally nilpotent radical of** *L, that contains all locally nilpotent Lie ideals, but such that L/N has no nonzero locally nilpotent ideals.*

Proof. Take a Lie ideal N, maximal with respect to being locally nilpotent. If $I \lhd L$ such that $(I+N)/N$ is locally nilpotent in L/N, then any finitely generated Lie subalgebra J of I has an element a that is central modulo N, implying that $N + Fa$ is a locally nilpotent ideal of J, by the lemma; continuing up the central series, one sees that J is locally nilpotent. This proves that I is locally nilpotent, and therefore $I \subseteq N$, as desired. □

The Lie ring of a nilpotent group.

The key to Zelmanov's solution of the RBP lies in a reduction from nilpotent groups to Lie algebras. We write $\gamma_i = \gamma_i(G)$ according to Definition 17.17′.

Definition 23B.12. Given a nilpotent group G of class t, let us define

$$L_i = \gamma_i/\gamma_{i+1}, \quad 1 \le i \le t,$$

an Abelian group. The **Lie ring** $L = L_\gamma(G)$ **of the group** G is the direct sum $\bigoplus_{i=1}^t L_i$, with Lie addition defined as the given group operation of L, rewritten in additive notation; Lie multiplication is defined on the cosets as follows:

$$[a_i\gamma_{i+1}, b_j\gamma_{j+1}] = (a_i, b_j)\gamma_{i+j+1},\ \forall a_i \in L_i,\ b_j \in L_j.$$

This is defined on each component and extended to all of $L_\gamma(G)$ via distributivity.

To check that our candidate for Lie multiplication is defined, we utilize Lemma 17.18.

THEOREM 23B.13. *$L = L_\gamma(G)$ is a Lie algebra and is $\mathbb{N}$-graded in the sense that $[L_iL_j] \subseteq L_{i+j}$. As a Lie algebra, L is Lie nilpotent of the same index t as the nilpotence class of the group G.*

Proof. Lemma 17.18 shows that L is graded. We check the defining identities of a Lie algebra. To ease notation, we write an element $\bar{a} = \sum \bar{a}_i$

as a sum of homogeneous components in L, where each $\bar{a}_i = a_i\gamma_{i+1} \in L_i$. Anticommutativity follows from

$$[\bar{a},\bar{a}] = \Big[\sum \bar{a}_i, \sum \bar{a}_j\Big] = \sum_i [\bar{a}_i,\bar{a}_i] + \sum_{i<j}([\bar{a}_i,\bar{a}_j]+[\bar{a}_j,\bar{a}_i]);$$

we note that $[\bar{a}_i,\bar{a}_i] = 0$ since $(a_i,a_i) = 1$, and $[\bar{a}_i,\bar{a}_j] + [\bar{a}_j,\bar{a}_i] = 0$ since $(a_i,a_j)(a_j,a_i) = 1$.

The Jacobi identity requires verifying

$$[\bar{a}_i[\bar{b}_j\bar{c}_k]] + [\bar{c}_k[\bar{a}_i\bar{b}_j]] + [\bar{b}_j[\bar{c}_k\bar{a}_i]] = 0,$$

or

$$(a_i,(b_j,c_k))(c_k,(a_i,b_j))(b_j,(c_k,a_i)) \in \gamma_{i+j+k+1}. \tag{23B.4}$$

But

$$(a_i^{c_k},(b_j,c_k)) = ((c_k,a_i)a_i,(b_j,c_k)) = ((c_k,a_i),(b_j,c_k)^{a_i})(a_i,(b_j,c_k))$$

by Equation (17.2), and $((c_k,a_i),(b_j,c_k)^{a_i}) \in \gamma_{i+j+k+1}$ and thus can be ignored; the other terms in Equation (17.3) are treated analogously, so (23B.4) reduces to showing that

$$(a_i^{c_k},(b_j,c_k))(c_k^{b_j},(a_i,b_j))(b_j^{a_i},(c_k,a_i)) \in \gamma_{i+j+k+1},$$

which is true by Equation (17.3).

The Lie nilpotence of L follows by comparing with Definition 21.21. □

Remark 23B.14. Since multiplication in G has been changed to addition in each L_i, we see that if G has exponent n, then L is a Lie algebra over $\mathbb{Z}/n$. Also, the Lie algebra L is f.g., as seen by implying induction to Lemma 17.64.

For $n = p$ prime, the multilinearized Engel ad-identity $\tilde{e}_{p-1}$ holds in L. This important result was known rather early in the game, following directly from Exercise 17C.4. (The reason we should first obtain a multilinearized ad-identity is that in verifying any identity in L, we may check only homogeneous elements.) But now, specializing all the indeterminates to the same indeterminate X, we get the ad-identity $(p-1)!e_{p-1}$; since p is prime, we can divide out by $(p-1)!$ and conclude that L satisfies e_{p-1}. (This also implies that L is a restricted Lie algebra.) L is nilpotent by Theorem 23B.5, completing the proof of the RBP for prime exponent.

Unfortunately, when $n = p^k$ for $k > 1$, we do not even know whether L satisfies $\tilde{e}_{n-1}$, so we need to change our Lie algebra construction. Instead of the lower central series $\gamma_i(G)$ we must deal with the **lower p-central series**

$$G = \hat{\gamma}_1 \supset \hat{\gamma}_2 \supset \cdots,$$

where $\hat{\gamma}_i$ is the subgroup of G generated by γ_i and all powers $\{a_j^{p^u} : jp^u \geq i\}$. One can check (cf. Exercise 5) that G now gives rise to a graded restricted Lie algebra

$$L_{\hat{\gamma}}(G) = \bigoplus \hat{\gamma}_i/\hat{\gamma}_{i+1} \tag{23B.5}$$

and furthermore, if the group G is f.g., then the corresponding elements of $\hat{\gamma}_1/\hat{\gamma}_2$ generate $L_{\hat{\gamma}}(G)$. Although $L_{\hat{\gamma}}(G)$ still fails to satisfy e_{n-1}, it does satisfy both $\tilde{e}_{n-1}$ and $e_{S,n}$, as shown respectively by Higman and Sanov. The slightly weaker result that $L_{\hat{\gamma}}(G)$ satisfies $\tilde{e}_n$ and $e_{0,2n}$ is given in Exercise 6, and suffices to conclude the proof of the RBP by Zelmanov's Theorem 23B.8.

A taste of the proof of Zelmanov's Theorem: Sandwiches.

The proofs of Theorems 23B.5, 23B.6, and 23B.8 are based on the following idea.

Definition 23B.15. A **sandwich** of a Lie algebra L is an element $a \in L$ such that $\mathrm{ad}_a^2 = 0 = \mathrm{ad}_a \,\mathrm{ad}\, L\, \mathrm{ad}_a$. L is a **sandwich algebra** if L is generated by finitely many sandwiches.

Theorem 23B.16 (Kostrikin and Zelmanov). *Any sandwich algebra is Lie nilpotent.*

Zelmanov's proof of Theorem 23B.16 requires a refinement of the definition of a sandwich; an outline of the proof is given in Exercises B7ff. To prove Theorem 23B.8 from Theorem 23B.16, it is more convenient to prove the equivalent assertion that any Lie algebra L satisfying the appropriate Engel-type identities is locally nilpotent. Assume on the contrary that L is not locally nilpotent. Modding out the locally nilpotent radical, one may assume that L has no nonzero locally nilpotent ideal. Suppose $f(x_1,\ldots,x_m)$ is a Lie polynomial that is not a Lie identity, but such that $f(r_1,\ldots,r_m)$ is a sandwich for any $r_i \in L$. The linear span of values of f turns out to be a Lie ideal, which by Theorem 23B.16 is locally nilpotent, a contradiction.

Although Zelmanov's Theorem would imply that such sandwich-valued polynomials f exist, Zelmanov could not construct one directly, so he considered sandwich-valued linear operators in order to conclude the proof. The details are far beyond the scope of this book. Zelmanov's original proof utilizes Jordan algebras; a self-contained treatment is given in [**Zel2**] for $p \neq 2$.

Chapter 24

Central Simple Algebras and the Brauer Group

So far, we have emphasized algebras over an algebraically closed field, exploiting the fact that every f.d. simple algebra over an algebraically closed field is split. In this chapter, we consider f.d. simple algebras over fields that need not be algebraically closed. The ensuing theory is very rich, with deep connections to geometry, K-theory, and cohomology theory. Some of the theory is to be generalized in Appendix 25B to separable algebras and Azumaya algebras.

Definition 24.1. A **central simple algebra over** F, denoted F**-csa** (or csa if F is understood), is a simple f.d. F-algebra center is F. The matrix algebra $M_n(F)$ is called the **split csa**; cf. Definition 15.24. A **central division algebra**, denoted **cda**, is a csa that is also a division algebra.

By the Wedderburn-Artin Theorem, any csa $R \cong M_t(D)$ for some t, where D is a cda, called the **underlying division algebra** of R. (Strictly speaking, this case is due to Wedderburn alone.) Recall from Corollary 15.9 that D is unique up to isomorphism and t is uniquely determined. Consequently, if $M_n(R_1) \cong M_n(R_2)$, then $R_1 \cong R_2$. (Indeed, writing $R_i = M_{t_i}(D_i)$, we get $D_1 \cong D_2$ and $nt_1 = nt_2$, implying $t_1 = t_2$.)

Basic examples

By Proposition 14.26, for F algebraically closed, the only cda is F itself. Thus the theory of cda's emerges only when F is not algebraically closed. So far we have seen only one noncommutative cda, namely Hamilton's algebra $\mathbb{H}$ for $F = \mathbb{R}$, given in Example 14.29. Since it is customary to develop a theory only after having at least two examples, let us start by scrounging around for a few more noncommutative cda's. We note that $\mathbb{H}$, whose dimension over $\mathbb{R}$ is 4, has many subfields isomorphic to $\mathbb{C}$, whose dimension over $\mathbb{R}$ is 2. This gives us the idea of building a division algebra by means of a cyclic Galois extension.

Example 24.2 (Cyclic algebras). Suppose K/F is a cyclic Galois field extension of dimension n with $\mathrm{Gal}(K/F) = \langle\sigma\rangle$, and pick nonzero $\beta \in F$ arbitrarily. Formally define $R = \bigoplus_{i=0}^{n-1} Kz^i$, viewed as an n-dimensional vector space over K with base $\{1, z, \dots, z^{n-1}\}$, and define multiplication by

$$(az^i)(bz^j) = \begin{cases} a\sigma^i(b)z^{i+j}, & i+j < n; \\ \beta a\sigma^i(b)z^{i+j-n}, & i+j \geq n. \end{cases} \tag{24.1}$$

One could check directly from (24.1) that multiplication in R is associative, but here is a more conceptual argument.

Recall the skew polynomial ring $W = K[\lambda;\sigma]$ from Example 13A.3. The element $\lambda^n - \beta$ is central in W, since $(\lambda^n - \beta)a = \sigma^n(a)\lambda^n - a\beta = a\lambda^n - a\beta$ for each $a \in K$. Thus $W/\langle\lambda^n - \beta\rangle$ is an algebra, in which we denote the image of λ as z. Clearly $1, z, \dots, z^{n-1}$ are linearly independent over F and (24.1) holds, so R is an algebra.

Write $K = F[a]$. We claim that R is an F-csa. Although this can be seen using the structure of skew polynomial rings (cf. Exercise 2), here is a direct proof. Suppose that $0 \neq I \lhd R$ and take $0 \neq r = \sum a_i z^i \in I$ with a minimal number of nonzero coefficients including $a_u \neq 0$. Then

$$\sum a_i \left(\sigma^u(a) - \sigma^i(a)\right) z^i = \sigma^u(a)\left(\sum a_i z^i\right) - \left(\sum a_i z^i\right) a \in I,$$

and has fewer nonzero coefficients (since the coefficient of z^u is 0), implying that each coefficient is 0, and thus $a_i = 0$, $\forall i \neq u$. (Indeed, $\sigma^u \neq \sigma^i$ implies $\sigma^u(a) - \sigma^i(a) \neq 0$.) Hence, $r = a_u z^u$, which has inverse $z^{n-u}\beta^{-1}a_u^{-1}$, thereby yielding $1 \in I$. Hence, R is simple.

It remains to show that $\mathrm{Cent}(R) = F$. If $c = \sum_{i=0}^{n-1} a_i z^i \in \mathrm{Cent}(R)$ for $a_i \in K$, then

$$\sum_{i=0}^{n-1} aa_i z^i = ac = ca = \sum_{i=0}^{n-1} a_i z^i a = \sum_{i=0}^{n-1} \sigma^i(a) a_i z^i;$$

matching coefficients, we conclude that $a_i = 0$ for all $i > 0$, since $\sigma^i(a) \neq a$ whenever $0 < i < n$, and thus $c = a_0 \in K$. But then $c = zcz^{-1} = \sigma(c)$, implying that $c \in F$, as desired.

We write this algebra R as (K, σ, β) and call it a **cyclic algebra**. Note that the cyclic algebra (K, σ, β) is unique up to isomorphism, since its multiplication table is determined by Formula (24.1).

Remark 24.3 (Symbol algebras). In Example 24.2, when the base field F contains a primitive n-th root ζ of 1, Example 4.93(ii) of Volume 1 shows that $K = F[y]$ for some $y \in K$ such that $\sigma(y) = \zeta y$, and then $\alpha = y^n \in F$. Now the cyclic algebra $R = (K, \sigma, \beta)$ satisfies

$$zyz^{-1} = \sigma(y) = \zeta y,$$

and we can describe multiplication in $R = \sum_{i,j=0}^{n-1} F y^i z^j$ more concisely in terms of the rules

$$y^n = \alpha, \qquad z^n = \beta, \qquad zy = \zeta yz.$$

One thereby writes the cyclic algebra R as the **symbol** $(\alpha, \beta; F; \zeta)_n$. For convenience we often suppress ζ and F and write $(\alpha, \beta)_n$ when the field F and the n-th root ζ of 1 are given.

Symbol algebras provide many examples of cda's.

Example 24.4. (i) Hamilton's division algebra $\mathbb{H}$ of quaternions is the symbol $(-1, -1)_2$, defined over $\mathbb{R}$. Here we take $y = i$ and $z = j$.

(ii) A **generalized quaternion algebra** is the symbol algebra

$$(\alpha, \beta)_2 = F + Fy + Fz + Fyz,$$

where $y^2 = \alpha$, $z^2 = \beta$, and $yz = -zy$ (or, equivalently, $(yz)^2 = -\alpha\beta$). The algebra $(\alpha, \beta)_2$ is a division algebra iff $\gamma_1^2 - \alpha\gamma_2^2 - \beta\gamma_3^2 + \alpha\beta\gamma_4^2 \neq 0$ for all $\gamma_i \in F$ (not all zero), since then

$$(\gamma_1 + \gamma_2 y + \gamma_3 z + \gamma_4 yz)^{-1} = \frac{1}{\gamma_1^2 - \alpha\gamma_2^2 - \beta\gamma_3^2 + \alpha\beta\gamma_4^2}(\gamma_1 - \gamma_2 y - \gamma_3 z - \gamma_4 yz).$$

(This can be checked directly in analogy to $\mathbb{H}$ or by means of an involution.) But the denominator $\gamma_1^2 - \alpha\gamma_2^2 - \beta\gamma_3^2 + \alpha\beta\gamma_4^2 = 0$ iff

$$\beta = \frac{\gamma_1^2 - \alpha\gamma_2^2}{\gamma_3^2 - \alpha\gamma_4^2} = N_{K/F}\left(\frac{\gamma_1 + \gamma_2 y}{\gamma_3 + \gamma_4 y}\right),$$

so a more concise way of phrasing this condition is that β is not a norm from $K = F[y]$ to F.

When $F = \mathbb{Q}$, for example, one gets infinitely many symbol division algebras by means of some elementary number theory; cf. Exercise 4.

So far our constructions have been built from a given field F that is to be the center of the division ring. However, one can take the opposite point of view, utilizing Corollary 23.34, which says that if R is any domain satisfying a polynomial identity, then the ring of central fractions $Q(R)$ of R is a cda over the field of fractions of $\mathrm{Cent}(R)$. In particular, this is true for any domain that is f.g. as a module over its center, in view of Proposition 23.12. Let us give some applications of this method.

Example 24.5 (The generic symbol). Suppose F_1 has a primitive root ζ of 1. (Explicitly, start with a field F_0 which is either $\mathbb{Q}$ or $\mathbb{Z}/q$ for q a prime number not dividing n. Then the polynomial $x^n - 1$ is separable over F_0, and so has n distinct roots, one of which is ζ; we take any field F_1 containing $F_0(\zeta)$.) We define an F_1-automorphism σ of $K = F_1[\mu]$ by putting $\sigma(\mu) = \zeta\mu$. Since σ fixes $F_1[\mu^n]$, the skew polynomial ring $K[\lambda; \sigma]$ has center containing $F_1[\lambda^n, \mu^n]$, and so is f.g. over its center and thus is a PI-ring. The ring of central fractions is a division ring with base $\{\lambda^i\mu^j : 0 \leq i, j < n\}$ over its center $F(\mu^n, \lambda^n)$, and thus is easily identified with the symbol algebra $(\mu^n, \lambda^n)_n$, called the **generic symbol**.

Example 24.6. Another striking example is obtained when we take R to be the algebra of generic matrices of Definition 23.49, which by Corollary 23.52 is a prime ring. Its ring of central fractions, denoted as $UD(n, F)$, plays a key role, to be described after we develop some more theory. In particular, we will see that $UD(n, F)$ is the most "generic" example of a division algebra of degree n (although its center is much larger than F).

Example 24.7 (Crossed products). Generalizing the cyclic algebra construction, suppose K/F is an arbitrary Galois extension of dimension n with $G = \mathrm{Gal}(K/F)$, and formally define $R = \bigoplus_{\sigma \in G} K z_\sigma$, viewed as an n-dimensional vector space over K with base $\{z_\sigma : \sigma \in G\}$. We want to make R into a simple (associative) ring such that $z_\sigma a = \sigma(a) z_\sigma, \forall a \in K$, and $z_\sigma z_\tau \in K z_{\sigma\tau}, \forall \sigma, \tau \in G$.

Toward this end, suppose $z_\sigma z_\tau = c_{\sigma,\tau} z_{\sigma\tau}$ for $c_{\sigma,\tau} \in K$. Then associativity of the ring R would require $(z_\sigma z_\tau) z_\rho = z_\sigma (z_\tau z_\rho)$, or

$$\begin{aligned} c_{\sigma,\tau} c_{\sigma\tau,\rho} z_{\sigma\tau\rho} &= c_{\sigma,\tau} z_{\sigma\tau} z_\rho = z_\sigma z_\tau z_\rho \\ &= z_\sigma c_{\tau,\rho} z_{\tau\rho} = \sigma(c_{\tau,\rho}) z_\sigma z_{\tau\rho} = \sigma(c_{\tau,\rho}) c_{\sigma,\tau\rho} z_{\sigma\tau\rho}, \end{aligned}$$

yielding the **factor set** condition

$$c_{\sigma,\tau}c_{\sigma\tau,\rho} = \sigma(c_{\tau,\rho})c_{\sigma,\tau\rho}, \quad \forall \sigma,\tau,\rho \in G. \tag{24.2}$$

Thus, we define multiplication in R via the rule

$$(az_\sigma)(bz_\tau) = a\sigma(b)c_{\sigma,\tau}z_{\sigma\tau}.$$

The same sort of verification as in Example 24.2 shows that R is a csa, called a **crossed product** and denoted as $(K, G, (c_{\sigma,\tau}))$. As in Example 24.2, we can use this information to reconstruct the multiplication table. Thus the crossed product $(K, G, (c_{\sigma,\tau}))$ is unique up to isomorphism. This definition has an important connection with a second cohomology group, to be described in Chapter 25; also see Example 24.49.

The Brauer group

Fixing the field F, there is a group structure, called the Brauer group, which we get by considering the class of all csa's. It is difficult to exaggerate the importance of the Brauer group in the subject of csa's.

Tensor products of simple algebras.

The first step in defining the Brauer group is the introduction of tensor products, because of the following results:

LEMMA 24.8. *If A, B are arbitrary F-algebras, then*

$$\mathrm{Cent}(A \otimes_F B) = \mathrm{Cent}(A) \otimes_F \mathrm{Cent}(B).$$

Proof. $(\supseteq)$ is clear, so we need only prove that $(\subseteq)$. Suppose

$$z = \sum_{i=1}^{t} a_i \otimes b_i \in \mathrm{Cent}(A \otimes B),$$

written with t minimal. Then for any a in A we have

$$0 = (a \otimes 1)z - z(a \otimes 1) = \sum (aa_i - a_i a) \otimes b_i,$$

implying that each $aa_i - a_i a = 0$ by Proposition 18.12, since the b_i are F-independent. Thus each $a_i \in \mathrm{Cent}(A)$; likewise, each $b_i \in \mathrm{Cent}(B)$. □

PROPOSITION 24.9. *If R is a csa and W is a simple F-algebra (not necessarily central), then $R \otimes_F W$ is simple with center* $\mathrm{Cent}(W)$.

Proof. Combine Lemma 24.8 with Corollary 18.34. □

COROLLARY 24.10. *If R_1 and R_2 are csa's, then $R_1 \otimes_F R_2$ is also a csa.*

Proof. Proposition 24.9 says that $R_1 \otimes_F R_2$ is simple with center F. □

Digression 24.11. To deal more generally with nonassociative algebras, it is easy to modify the proofs of Lemma 24.8 and Corollary 18.34, and thus prove Corollary 24.10 for nonassociative algebras. However, although the tensor product of associative algebras is associative, other famous (nonassociative) varieties are not closed under tensor products; for example, the tensor product of alternative algebras need not be alternative!

Corollary 24.10 also provides an important way of generating new cda's. Namely, if D_1 and D_2 are cda's over F, then $D_1 \otimes_F D_2$ is a csa, and thus has the form $M_t(D)$ for some cda D. (One must be careful, for we shall see that the tensor product $D_1 \otimes D_2$ of division algebras need not be a division algebra.) Furthermore, the set of isomorphism classes of csa's constitutes a monoid under tensor multiplication, where the identity element is F. On the other hand, since the dimension increases under taking tensor products, this monoid has no invertible elements other than F itself. To obtain a group structure, we introduce a very important equivalence relation.

Definition 24.12. Two csa's R_1 and R_2 are **(Brauer) equivalent**, written $R_1 \sim R_2$, if their underlying division algebras are isomorphic. We write $[R]$ for the Brauer class of R under this equivalence. Thus, any Brauer class $[R]$ contains a *unique* division algebra (up to isomorphism), namely the underlying division algebra D of R, and we often identify $[R]$ with D. In particular, if R is split, then $[R] = [F]$; in this case, we write $R \sim 1$.

Remark 24.12′. If $e \neq 0$ is an idempotent of a csa R, then eRe (as a csa with multiplicative unit e) is Brauer equivalent to R, by Remark 15.8(ii).

We define $\mathrm{Br}(F)$ to be the set of Brauer equivalence classes of csa's over F, endowed with multiplication

$$[R_1][R_2] = [R_1 \otimes_F R_2].$$

The group structure now follows from the following theorem, which arguably is the most basic theorem of the subject.

Remark 24.13. Suppose R is an algebra over a commutative ring C. Given $a \in R$, we recall the left multiplication map $\ell_a \colon R \to R$ given by

$r \mapsto ar$ and the right multiplication map $\mathfrak{r}_a\colon R \to R$ given by $r \mapsto ra$. Thus $\ell_a, \mathfrak{r}_a \in \mathrm{End}_C R$. The left regular representation (Example 13.49) is the homomorphism $R \to \mathrm{End}_C R$ given by $a \mapsto \ell_a$, since $\ell_{ab} = \ell_a\ell_b$. On the other hand, $\mathfrak{r}_a\mathfrak{r}_b = \mathfrak{r}_{ba}$, so the right regular representation, given by $a \mapsto \mathfrak{r}_a$, is an anti-homomorphism or, equivalently, a homomorphism $R^{\mathrm{op}} \to \mathrm{End}_C R$. Note that

$$\ell_a\mathfrak{r}_b(r) = a(rb) = (ar)b = \mathfrak{r}_b\ell_a(r), \quad \forall r \in R,$$

so $\ell_a\mathfrak{r}_b = \mathfrak{r}_b\ell_a$ for all $a \in R$ and $b \in R^{\mathrm{op}}$, and thus Remark 18.23 yields an algebra homomorphism $\Phi\colon R \otimes_C R^{\mathrm{op}} \to \mathrm{End}_C R$ given by $(a, b) \mapsto \ell_a\mathfrak{r}_b$.

Of course, we are now interested in the case $C = F$, a field, although the more general situation is needed in Appendix 25B.

THEOREM 24.14. *If R is a csa, then $\Phi\colon R \otimes_F R^{\mathrm{op}} \to \mathrm{End}_F R$ is an isomorphism.*

Proof. $R \otimes_F R^{\mathrm{op}}$ is simple by Proposition 24.9, so Φ is injective. But $\mathrm{End}_F R \cong M_{n^2}(F)$ has dimension n^2, as does $R \otimes_F R^{\mathrm{op}}$, by Corollary 18.13, implying that Φ is also onto. □

COROLLARY 24.15. $\mathrm{Br}(F)$ *is a group, where* $[R]^{-1} = [R]^{\mathrm{op}}$.

Proof. $[R][R^{\mathrm{op}}] = [R \otimes_F R^{\mathrm{op}}] = [\mathrm{End}_F R] = [M_{n^2}(F)] = [F]$. □

Definition 24.16. $\mathrm{Br}(F)$ is called the **Brauer group**.

Example 24.17. (i) When F is algebraically closed, every csa is split, so $\mathrm{Br}(F) = \{[F]\}$ is trivial. Finite fields and fields of transcendence degree 1 over an algebraically closed field also have trivial Brauer group, as seen in Theorem 24.42 and Exercise 17.

(ii) $\mathrm{Br}(\mathbb{R}) = \{[\mathbb{R}], [\mathbb{H}]\}$ by Frobenius' theorem, which says the only cda's over $\mathbb{R}$ are $\mathbb{R}$ and $\mathbb{H}$. (Although this can be seen directly, cf. Exercise 1, the proof becomes easy after we develop more structure theory; cf. Example 24.45.)

Also see Theorem 24.42. Incidentally, there are very few instances of a field whose Brauer group is a nontrivial finite group. The basic group-theoretic properties of $\mathrm{Br}(F)$ becomes the pillar of the theory.

Splitting fields.

Let us compare Brauer groups of different fields. We write $[R : F]$ for $\dim_F R$.

Remark 24.18. Suppose R is an F-algebra. Then, for any field extension L of F, $R \otimes_F L$ is an L-algebra and, by Remark 18.27,

$$[R\colon F] = [R \otimes_F L : L].$$

R is an F-csa iff $R \otimes_F L$ is an L-csa. (Indeed, ($\Rightarrow$) is by Lemma 24.8 and Corollary 18.34; conversely, if $R \otimes_F L$ is an L-csa, then $\mathrm{Cent}(R) = F$ by Lemma 24.2, and R is simple by Remark 18.28.)

Thus, $\_ \otimes_F L$ yields a group homomorphism $\mathrm{Br}(F) \to \mathrm{Br}(L)$, called the **restriction map** $\mathrm{res}_{L/F}$, which need not be 1:1; for example, take $F = \mathbb{R}$ and $L = \mathbb{C}$. The kernel of $\mathrm{res}_{L/F}$ is denoted $\mathrm{Br}(L/F)$ and is called the **relative Brauer group**. $[R] \in \mathrm{Br}(L/F)$ iff $R \otimes_F L \sim L$; in this case L is called a **splitting field** of R, in the sense of Definition 18.30.

Remark 24.18′. Br can be viewed via Remark 24.18 as a functor from the category of fields (whose morphisms are field embeddings) to the category of groups; indeed, given an embedding $F \to L$, we have the corresponding map $\mathrm{res}_{L/F}\colon \mathrm{Br}(F) \to \mathrm{Br}(L)$.

Any csa Brauer equivalent to R has the same splitting fields as R; more generally, any splitting field of R_1 and of R_2 is also a splitting field of $R_1 \otimes R_2$.

Remark 24.19. The algebraic closure $\bar{F}$ of F is clearly a splitting field of any csa R and thus, for suitable n, $R \otimes \bar{F} \cong M_n(\bar{F})$. In other words, the map $r \mapsto r \otimes 1$ defines an injection $R \hookrightarrow M_n(\bar{F})$. Hence, by Remark 18.27,

$$[R\colon F] = [R \otimes_F \bar{F} : \bar{F}] = [M_n(\bar{F}) : \bar{F}] = n^2,$$

a square.

Likewise, by Remark 18.27, for any $r \in R$, $\deg r = \deg_{\bar{F}}(r \otimes 1) \le n$ (since every matrix satisfies its characteristic polynomial).

Definition 24.20. We call $n = \sqrt{[R\colon F]}$ the **degree** of R, written $\deg(R)$. The **index** of R, denoted $\mathrm{ind}(R)$, is defined as the degree of the underlying division algebra of R.

If $R = M_t(D)$, then $\deg(R) = t \cdot \mathrm{ind}(R)$. In particular, R is a cda iff $\mathrm{ind}(R) = \deg(R)$.

We still would like to know when arbitrary cda's are split! Here is one basic but useful criterion. (Later, in Remark 24.46(iii), we establish a field-theoretic criterion of Wedderburn for cyclic algebras to be split.)

Remark 24.20′. Suppose $R = M_t(D)$ is a csa with $\deg R = n$, and L is a minimal left ideal of R. In view of Corollary 15.9 (and Remark 15.9′), we may assume that $L = Re_{11}$ and $D = \mathrm{End}_R L$. L can be viewed as a right D-module, and we defined $[L:D]$ in Remark 13.16. Then $[L:D] = t$, so $[L:F] = [L:D][D:F] = t(\frac{n}{t})^2 = \frac{n^2}{t} \geq n$.

It follows that R has a left ideal of dimension n iff $\frac{n}{t} = 1$, iff R is split.

Subfields and centralizers

By Remark 14.30, any cda D of degree > 1 has a wealth of subfields (although many may be isomorphic). Subfields work in tandem with the next definition.

Definition 24.21. The **centralizer** $C_R(A)$ of a subalgebra A of R is $\{r \in R : ra = ar, \forall a \in A\}$, easily seen also to be a subalgebra.

Thus, given a subfield $K \supset F$ of R, we can examine $C_R(K)$, whose dimension as a K-algebra is less than $[R:F]$. So by obtaining enough of a connection between R and $C_R(K)$, we might be able to establish an inductive procedure for investigating R. To do this, we need some more information about centralizers.

Example 24.22. (i) $C_R(R) = \mathrm{Cent}(R)$;

(ii) $C_R(\mathrm{Cent}\, R) = R$;

(iii) If A is a maximal commutative subring of R, then $C_R(A) = A$; indeed, if $r \in C_R(A)$, then A and r generate a commutative subring containing A, so $r \in A$ by hypothesis.

(iv) Every commutative F-subalgebra of a cda D is a field, since it is a f.d. integral domain over F. Thus, a subfield K of D is maximal iff $C_D(K) = K$.

Some words of caution in general: the $\mathbb{C}$-csa $M_n(\mathbb{C})$ has no subfields other than the center $\mathbb{C}$, since $\mathbb{C}$ is algebraically closed. On the other hand, $M_2(\mathbb{C})$ has the commutative subalgebra $\mathbb{C} \oplus \mathbb{C}$ along the diagonal. Hence, $\mathbb{C}$ is maximal as a subfield of $M_2(\mathbb{C})$ but not as a commutative subring. As will become clearer in Appendix 25B, the theory works best if we consider commutative algebras that are semisimple, i.e., direct products of subfields. We are ready for a generalization of Theorem 24.14 (which one recovers by taking $K = F$):

Proposition 24.23. $\mathrm{End}_K R \cong C_R(K) \otimes_F R^{\mathrm{op}}$ *as K-algebras, for any F-subfield K of R.*

Proof. We view the elements of K via the regular representation as left multiplication maps in $\mathrm{End}_F R$; then $\mathrm{End}_K R$ is the centralizer of K in $\mathrm{End}_F R$.

We can view this centralizer from a different perspective. Identifying $\mathrm{End}_F R$ with $R \otimes_F R^{\mathrm{op}}$, we identify K with $K \otimes_F F$ and claim that its centralizer Y in $R \otimes_F R^{\mathrm{op}}$ is $C_R(K) \otimes R^{\mathrm{op}}$. Indeed, clearly $C_R(K) \otimes R^{\mathrm{op}} \subseteq Y$. For the reverse inclusion, suppose $y \in Y$. Writing $y = \sum r_i \otimes v_i$ such that the $v_i \in R^{\mathrm{op}}$ are independent over F and taking arbitrary $a \otimes 1 \in K \otimes_F F$ yields

$$0 = \sum ((a \otimes 1)(r_i \otimes v_i) - (r_i \otimes v_i)(a \otimes 1)) = \sum (ar_i - r_i a) \otimes v_i,$$

implying (by Proposition 18.12) that $ar_i - r_i a = 0$ for each $a \in K$ and thus $r_i \in C_R(K)$ for each i. Hence, $y \in C_R(K) \otimes R^{\mathrm{op}}$, as desired.

Matching the two shows that $\mathrm{End}_K R \cong C_R(K) \otimes R^{\mathrm{op}}$. □

Corollary 24.24. *For any subfield K of a csa R, $C_R(K)$ is a K-csa and $[C_R(K):F] = [R:K]$.*

Proof. $\mathrm{End}_K R \cong M_{[R:K]}(K)$ is simple. Hence, in view of Remark 18.28, $C_R(K)$ is simple with center K, by Lemma 24.8. The last assertion is seen by comparing dimensions in Proposition 24.23:

$$[C_R(K):F][R:F] = [\mathrm{End}_K R:F] = [R:K]^2[K:F] = [R:K][R:F],$$

so $[C_R(K):F] = [R:K]$. □

Corollary 24.25. $R \otimes_F K \sim C_R(K)$ *in* $\mathrm{Br}(K)$.

Proof. Proposition 24.23 shows in $\mathrm{Br}(K)$ that

$$\begin{aligned} R \otimes_F K &\sim (R \otimes_F K) \otimes_K \mathrm{End}_K R \sim R \otimes_F K \otimes_K (C_R(K) \otimes_F R^{\mathrm{op}}) \\ &\cong C_R(K) \otimes_K (K \otimes_F R \otimes_F R^{\mathrm{op}}) \sim C_R(K) \otimes_K M_{n^2}(K) \sim C_R(K). \end{aligned}$$ □

We are ready to determine which subfields of a csa are splitting fields.

Corollary 24.26 (Brauer-Noether). *Suppose K is a subfield of a csa R of degree n.*

(i) $\deg_K C_R(K) \approx \frac{n}{[K:F]}$.

(ii) *K is a splitting field of R iff $C_R(K)$ is split.*

(iii) *$C_R(K) = K$ iff $[K:F] = n = [R:K]$.*

(iv) *For R a cda, K is a splitting field of R iff K is maximal as a subfield, iff $[K:F] = n$.*

Proof. (i) Let $m = [K{:}F]$. By Corollary 24.24, $[C_R(K){:}F] = [R{:}K] = \frac{n^2}{m}$, implying $[C_R(K){:}K] = \frac{n^2}{m^2} = (\frac{n}{m})^2$.

(ii) is immediate from Corollary 24.25. To see (iii), note that

$$n^2 = [R{:}F] = [R{:}K][K{:}F] = [C_R(K){:}F][K{:}F] \geq [K{:}F]^2,$$

so $n = [K{:}F]$ iff equality holds at the last step, i.e., $C_R(K) = K$.

(iv) Since R has no zero-divisors, $C_R(K)$ is split iff $C_R(K) = K$, iff the subfield K is maximal; so we are done by (ii) and (iii). □

In other words, the splitting subfields of a cda are precisely the maximal subfields, and these become the focus of our investigation. Here is a useful observation in this direction.

Corollary 24.27. *Suppose $K \supseteq F$ is a subfield of a cda D. Then any maximal subfield of $C_D(K)$ is a maximal subfield of D.*

Proof. Let $n = \deg D$. Any maximal subfield L of $C_D(K)$ satisfies

$$[L{:}F] = [L{:}K][K{:}F] = \deg_K C_D(K)[K{:}F] = \frac{n}{[K{:}F]}[K{:}F] = n. \qquad \square$$

The Double Centralizer Theorem.

We can generalize these results even further, using Theorem 24.14 as our point of departure.

Proposition 24.28. *Suppose $R = AB$ for F-subalgebras A, B of R, and $ab = ba$ for every $a \in A$ and $b \in B$. If A is a csa, then the natural map $\Phi\colon A \otimes_F B \to R$ given by $a \otimes b \mapsto ab$ is an isomorphism. Conversely, if R is a csa and Φ is an isomorphism, then A and B are csa's.*

Proof. Φ is an algebra homomorphism by Remark 18.23 and is onto by hypothesis. Given that A is a csa, we need to show that $\ker \Phi = 0$; but otherwise, by Proposition 18.33, $\ker \Phi$ would contain a nonzero element of the form $1 \otimes b$, so $0 = \Phi(1 \otimes b) = b$, a contradiction.

The last assertion follows at once from Remark 18.28. □

Corollary 24.29. *In the setup of Proposition 24.28, if R and A are csa's, then $B = C_R(A)$ and $A = C_R(B)$.*

Proof. Since Φ is an isomorphism, $[R{:}F] = [A{:}F][B{:}F]$. But replacing B by $C_R(A) \supseteq B$ in the hypothesis yields $[R{:}F] = [A{:}F][C_R(A){:}F]$, implying that $[B{:}F] = [C_R(A){:}F]$ and thus $B = C_R(A)$. Furthermore, B is a csa, by Proposition 24.28, so we may interchange the roles of A and B and obtain $A = C_R(B)$. □

Example 24.30. (i) Suppose R is a csa with a subalgebra isomorphic to $M_u(F)$; then $R \cong M_u(F) \otimes R_1$ for a suitable csa algebra R_1. (Indeed, noting that R has a set of $u \times u$ matric units, we can write $R \cong M_u(R_1)$ by Proposition 13.9, and the scalars R_1 commute with the matric units and thus with $M_u(F)$.)

(ii) As a special case of (i), $M_u(D) \simeq M_u(F) \otimes D$ for any division ring D.

Proposition 24.31. *If A is a central simple subalgebra of the csa R, then $R \cong A \otimes_F C_R(A)$.*

Proof. If A is split, we are done by Example 24.30. In general, $A^{\mathrm{op}} \otimes R$ contains $A^{\mathrm{op}} \otimes A \cong M_m(F)$, where $m = [A{:}F]$; cf. Theorem 24.14. Hence, $A^{\mathrm{op}} \otimes R \cong M_m(B)$ for a suitable csa B, implying that

$$M_m(R) \cong (A \otimes A^{\mathrm{op}}) \otimes R \cong A \otimes M_m(B) \cong A \otimes M_m(F) \otimes B \cong M_m(A \otimes B),$$

yielding $R \cong A \otimes B$, by Corollary 15.9. $B = C_R(A)$ by Corollary 24.29. □

To handle subalgebras that are not necessarily central, we first observe that if $A \subseteq T \subseteq R$, then $C_T(A) = T \cap C_R(A)$; in particular, if $C_R(A) \subseteq T$, then $C_T(A) = C_R(A)$. We are ready to put everything together.

Theorem 24.32 (Double Centralizer Theorem). *Suppose A is a simple F-subalgebra of a csa R, and $K = \mathrm{Cent}(A)$. Then*

(i) *$C_R(A)$ is simple with center K;*
(ii) *$C_R(K) \cong A \otimes_K C_R(A)$;*
(iii) *$C_R(C_R(A)) = A$;*
(iv) *$[A{:}F][C_R(A){:}F] = n^2$.*

Proof. Let $T = C_R(K)$, a K-csa by Corollary 24.24. Note that $A \subseteq T$ and $C_R(A) \subseteq C_R(K) = T$. Hence, $C_T(A) = C_R(A)$, and Proposition 24.31 yields

$$C_R(K) = T \cong A \otimes_K C_T(A) = A \otimes_K C_R(A).$$

This proves (i) and (ii), and $A = C_T(C_R(A)) = C_R(C_R(A))$, proving (iii).

(iv) $[C_R(K){:}K] = [A{:}K][C_R(A){:}K]$ by (ii); hence, by Corollary 24.26,

$$\begin{aligned} n^2 &= \left(\frac{n}{[K{:}F]}\right)^2 [K{:}F] = [C_R(K){:}K][K{:}F]^2 \\ &= [A{:}K][C_R(A){:}K][K{:}F]^2 \\ &= [A{:}F][C_R(A){:}F]. \end{aligned}$$ □

An alternate proof of (iii) by means of the Density Theorem yields a more general result; cf. Exercise 19.39.

COROLLARY 24.32′. *Notation as in Theorem 24.32, if B centralizes A and $[A:F][B:F] = n^2$, then $B = C_R(A)$ and $A \otimes_K B \cong C_R(K)$.*

Proof. $B \subseteq C_R(A)$ but $[B:F] = [C_R(A):F]$, implying that $B = C_R(A)$. □

Index reduction.

Even if a field does not split a csa R, we still want to know how it affects the index, defined in Definition 24.20. We fix the following notation: R is a csa with underlying division algebra D, and L is a field containing F with $m = [L:F]$.

Remark 24.33. $D \otimes_F L \cong M_u(D_1)$ for some division algebra D_1, and some u. Then

$$u \deg(D_1) = \deg(D) = \operatorname{ind}(R),$$

so $u = \operatorname{ind}(R)/\operatorname{ind}_L(R \otimes_F L)$. In particular, $u = \operatorname{ind}(R)$ iff L is a splitting field for R; on the other hand, $u = 1$ iff $D \otimes_F L$ is a division algebra.

Definition 24.33′. The number u of Remark 24.33 is called the **index reduction factor** (of R under L).

THEOREM 24.34 (INDEX REDUCTION THEOREM). *Notation as above, the index reduction factor u divides the g.c.d. of* $\operatorname{ind}(R)$ *and $m = [L:F]$.*

Proof. We may assume that $R = D$. In view of Remark 24.33, it remains to show that $u \mid m$. Embedding L in $M_m(F)$ via the regular representation, we have

$$M_m(D) \cong D \otimes M_m(F) \supseteq D \otimes_F L = M_u(D_1), \tag{24.3}$$

so $u \mid m$ by Example 24.30(ii) (for which $m = ku$). □

COROLLARY 24.35. *Suppose D is a cda and $L \supset F$ is any field. If* $\deg D$ *and $[L:F]$ are relatively prime, then $D \otimes_F L$ remains a division algebra.*

Proof. The index reduction factor divides both $\deg(D)$ and $[L:F]$, and thus is 1.

□

We write $R^{\otimes k}$ for $R \otimes_F R \otimes_F \cdots \otimes_F R$, where R appears k times.

COROLLARY 24.36. $\operatorname{ind}(R^{\otimes k})$ *divides* $\operatorname{ind}(R)$ *for any k.*

Proof. We may assume that R is a cda D. We take a maximal subfield K of D. Then $D^{\otimes k} \otimes_F K \cong (D \otimes_F K)^{\otimes k} \sim 1$, so $\operatorname{ind}(D^{\otimes k})$ divides $[K:F] = \operatorname{ind}(D)$, by Theorem 24.34. □

COROLLARY 24.37. *Hypotheses as in Theorem 24.34, L is isomorphic to a subfield of $M_{m/u}(D)$, which is maximal if L is a splitting field of D.*

Proof. With the setup of Remark 24.33, let $\hat{R} = M_m(D)$. Clearly

$$M_{m/u}(D) = C_{\hat{R}}(M_u(F)) \supseteq C_{\hat{R}}(M_u(D_1)) = C_{\hat{R}}(D \otimes_F L) \supseteq L.$$

Furthermore, when L is a splitting field of D, we have $\deg(D) = u$, implying that $[L:F] = m = \frac{m}{u}u = \deg M_{m/u}(D)$; thus, L is a maximal subfield of $M_{m/u}(D)$ by Corollary 24.26(iv). □

We now have an extension of Corollary 24.26 for arbitrary splitting fields.

COROLLARY 24.38. *If L is a splitting field of a cda D and $[L:F] = \deg(D)$, then L is isomorphic to a maximal subfield of D.*

Division algebras described in terms of maximal subfields

Now that we can recognize maximal subfields, we turn to the internal structure of a cda D in terms of a maximal subfield. We start by asking, "Which maximal subfields have the 'nicest' field-theoretic structure, and how does this enable us to determine the structure of D?" To address this question, we turn again to tensor products.

Remark 24.39. If A and W are subalgebras of R, then the natural A, W-bimodule structure of R makes R naturally an $A \otimes W^{\mathrm{op}}$-module, i.e.,

$$(a \otimes w)r = arw.$$

Recall that $R^\times$ denotes the group of invertible elements of R. We say that elements a_1 and a_2 of a csa R are **conjugate** if $a_2 = ua_1u^{-1}$ for suitable $u \in R^\times$. Likewise, subsets K_1 and K_2 are **conjugate** with respect to u if $K_2 = \{uau^{-1} : a \in K_1\}$. An automorphism φ of R is called **inner** if it is given by conjugation by some $u \in R^\times$; i.e., $\varphi(a) = uau^{-1}, \forall a \in R$.

THEOREM 24.40 (SKOLEM-NOETHER THEOREM). *Suppose A_1 and A_2 are isomorphic simple subalgebras of a csa R. Any F-algebra isomorphism $\varphi\colon A_1 \to A_2$ is given by conjugation by some $u \in R^\times$.*

Proof. Let $T = A_1 \otimes_F R^{\mathrm{op}}$, a simple algebra. Let $M_1 = R$, viewed as a T-module via Remark 24.39, and $M_2 = R$, viewed as a T-module under the action $(a \otimes r)y = \varphi(a)yr$ for $y \in M_2$. Being simple Artinian, T has a unique simple module M_0 (up to isomorphism), whose dimension as a vector space over F is some number d. Both M_1 and M_2 are direct sums of copies of M_0, the number of copies being precisely $\frac{[R:F]}{d}$ in each case. Hence, there is a module isomorphism $f\colon M_1 \to M_2$, so this satisfies

$$f(ayr) = \varphi(a)f(y)r, \qquad \forall a \in A_1, \quad \forall r, y \in R.$$

Taking $a = y = 1$ shows that $f(r) = f(11r) = f(1)r$ for all r in R, and $u = f(1)$ is invertible since f is an isomorphism (whereby one gets u^{-1} by applying the same argument to f^{-1}); now taking $y = r = 1$ shows that

$$ua = 1f(1)a = f(a) = f(a11) = \varphi(a)f(1)1 = \varphi(a)u,$$

as desired. □

Corollary 24.41. *For any F-subfield K of a csa R, each F-automorphism of K lifts to an (inner) automorphism of R.*

Here is a surprising application.

Theorem 24.42 (Wedderburn). *Every finite division ring is a field; in other words,* $\mathrm{Br}(F)$ *is trivial for any finite field F.*

Proof. Suppose D were a finite division ring of finite degree $n > 1$ over F. Thus, $[D : F] = n^2$. Let $m = |D| = |F|^{n^2}$. Any maximal subfield K has dimension n over F and thus has order $u = |F|^n$; then $|K^\times| = u - 1$. The number of multiplicative subgroups of $D^\times$ conjugate to $K^\times$ is at most $\frac{m-1}{u-1}$. Since each subgroup contains 1, the union of these subgroups contains at most $1 + \frac{m-1}{u-1}(u-2) < m - 1$ elements, so there is an element d of $D^\times$ not conjugate to any element of K. Let L be a maximal subfield containing d. Then $|L| = |F|^n = |K|$, implying that $L \cong K$, contrary to the Skolem-Noether Theorem. □

This proof was fast but not very enlightening. A more conceptual proof based on a theorem of Tsen is given in Exercise 17.

More criteria for an algebra to be split.

Bringing in some theory to bear on Remark 24.20′, we are ready for a more sophisticated method of detecting split algebras.

Lemma 24.43. *Suppose $K = F[a]$ (not necessarily a field) and the minimal polynomial $f \in F[\lambda]$ of a is separable and has a linear factor (over F). Then $K \cong F_1 \times \cdots \times F_m$, for fields F_i such that $F_1 \cong F$.*

Proof. $K \cong F[\lambda]/F[\lambda]f$. But we can write $f = f_1 \cdots f_m$ for distinct irreducible polynomials f_j, which thus are relatively prime. Also, we may take f_1 to be the linear factor. Hence, by the CRT,

$$K \cong F[\lambda]/F[\lambda]f \cong \prod_{j=1}^{m} F[\lambda]/F[\lambda]f_j,$$

and $F[\lambda]/F[\lambda]f_1 \cong F$ since f_1 is linear. □

Proposition 24.44. *Suppose $|F|$ is infinite. The following conditions are equivalent for a csa R of degree n:*

(i) *R is split;*
(ii) *R contains an element r of degree n over F, whose minimal polynomial f splits into n linear factors over F;*
(iii) *R contains an element r of degree n over F, whose minimal polynomial has a linear factor.*

Proof. (i) ⇒ (ii). Take $r = \alpha_1 e_{11} + \alpha_2 e_{22} + \cdots + \alpha_n e_{nn}$ for $\alpha_i \in F$ distinct. Then $f = \prod(\lambda - \alpha_i)$.

(ii) ⇒ (iii). Obvious.

(iii) ⇒ (i). Write $R = M_t(D)$ (where clearly $t \le n$). First we handle the case when f splits into linear factors over F. By the lemma, $F[a]$ has n orthogonal idempotents $e_1, \ldots, e_n$. Thus, R has composition length at least n, cf. Proposition 13.5 of Volume 1, so $t \ge n$, implying $t = n$ and $D = F$.

In general, write $f = f_1 \cdots f_m$ with the f_j irreducible and f_1 linear. By the lemma, $F[a]$ has orthogonal idempotents $e_1, \ldots, e_m$ with $F_j = F[a]e_j$ for each j. We claim that $[Re_1 : F] = n$, which would yield the result in view of Remark 24.20′.

Let L be a minimal left ideal of R; by Remark 24.20′, $[L:F] = \frac{n^2}{t}$. Also, $Re_1 \approx L^{(j)}$ for some j, and then $R(1-e_1) \approx L^{(t-j)}$. Let $R' = K \otimes_F R$, where K is a splitting field of f, and let $L' = K \otimes_F L$. Then $[L' : K] = [L : F] = \frac{n^2}{t}$, so

$$R'e_1 \cong K \otimes_F L^{(j)} \cong (K \otimes_F L)^{(j)} = (L')^{(j)},$$

implying $[R'e_1 : K] = j\frac{n^2}{t}$. But $[R'e_1 : K] = n$ by the first paragraph, since f splits over K. Thus $j\frac{n^2}{t} = n$, implying $jn = t$; hence, $j = 1$ and $t = n$. Hence $[L : F] = n$, implying that R is split, again by Remark 24.20′. □

Cyclic algebras and symbols, revisited.

With the Skolem-Noether Theorem in hand, we can take a much deeper look at cyclic algebras and crossed products. We start with cyclic algebras, using the notation of Example 24.2.

Example 24.45. A csa R of degree n is cyclic, iff it has a (maximal) subfield K that is cyclic Galois of dimension n over F.

Indeed, to see ($\Leftarrow$), write $G = \mathrm{Gal}(K/F) = \langle\sigma\rangle$. Thus, $n = [K:F] = |G| = \circ(\sigma)$, and, by Theorem 24.40, there is $z = z_\sigma$ such that $zaz^{-1} = \sigma(a)$ for each $a \in K$. Iterating, we get $z^i a z^{-i} = \sigma^i(a)$, or

$$z^i a = \sigma^i(a) z^i, \qquad \forall i \geq 1; \tag{24.4}$$

in particular,

$$z^n a = \sigma^n(a) z^n = a z^n, \quad \forall a \in K,$$

implying that $z^n \in C_R(K) = K$; then $z^n = z z^n z^{-1} = \sigma(z^n)$, implying that z^n is some element β of the fixed subfield $K^G = F$. We claim that $R \cong (K, \sigma, \beta)$.

Indeed, there is a surjection $K[\lambda;\sigma] \to R$ given by $\lambda \mapsto z$, whose kernel contains $\lambda^n - \beta$. The first Noether isomorphism theorem gives us a surjection $\varphi\colon K[\lambda;\sigma]/\langle\lambda^n - \beta\rangle \to R$. But by Example 24.2, $K[\lambda;\sigma]/\langle\lambda^n - \beta\rangle \cong (K, \sigma, \beta)$ is simple, so $\ker\varphi = 0$; i.e., φ is an isomorphism.

We are ready to establish a well-known condition, due to Wedderburn, for a cyclic algebra to be split.

Remark 24.46.

(i) The cyclic algebra $(K, \sigma, 1)$ is split for any cyclic Galois extension K of F. (Indeed, this follows from Proposition 24.44(iii), since the minimal polynomial of z over F is $\lambda^n - 1$, which has the linear factor $\lambda - 1$.)

(ii) In the cyclic algebra $R = (K, \sigma, \beta)$, we could replace z by za_0 for any $a_0 \in K$, since

$$za_0 a(za_0)^{-1} = za_0 a a_0^{-1} z^{-1} = zaz^{-1};$$

then β is replaced by

$$(za_0)^n = \sigma(a_0)\cdots\sigma^n(a_0) z^n = N_{K/F}(a_0)\beta.$$

Conversely, suppose we have $\tilde{z}$ such that $\tilde{z}a\tilde{z}^{-1} = \sigma(a) = zaz^{-1}, \forall a \in K$. Then $z^{-1}\tilde{z}a = az^{-1}\tilde{z}$, i.e., $z^{-1}\tilde{z} \in C_R(K) = K$, implying that $\tilde{z} = za_0$ for some $a_0 \in K$. Thus, once the maximal subfield K of R is chosen, the only flexibility in β is multiplying by $N_{K/F}(a_0)$ for some $a_0 \in K$.

(iii) (Wedderburn) In view of (i) and (ii), the cyclic algebra $R = (K, \sigma, \beta)$ is split iff $\beta = N_{K/F}(a)$ for some $a \in K$. (Indeed, if $\beta = N_{K/F}(a)$ and $z^n = \beta$, then $(za^{-1})^n = \beta N_{K/F}(a)^{-1} = 1$, implying that R is split, by (i). The direction ($\Rightarrow$) is trickier. By (i), R, being split, is isomorphic to $(K, \sigma, 1)$,

and thus has a subfield $K_1 \cong K$ and z_1 inducing the automorphism σ on K_1 such that $z_1^n = 1$. By the Skolem-Noether Theorem we may apply a suitable automorphism of R to assume that $K_1 = K$, and we are done by (ii).

We return to the question of when a cyclic algebra (K, σ, β) is a division algebra. Wedderburn showed that this holds if none of the β^m $(1 \leq m < n)$ is a norm from K to F, cf. Exercise 8, but the theory still lacks a useful necessary and sufficient condition.

Symbols, a special case of cyclic algebras defined in Remark 24.3, turn out to be the building blocks of the theory, so let us consider them briefly.

Example 24.47. Any division algebra D of degree 2 over an arbitrary field F of characteristic $\neq 2$ is a generalized quaternion algebra. Indeed, by Corollary 24.26(iv), D has a maximal subfield K of dimension 2 over F, which has the form $F[\sqrt{\alpha}]$ by Remark 4.10 of Volume 1; by Example 24.45, it follows that D is a symbol algebra of the form $(\alpha, \beta)_2 = F + Fy + Fz + Fyz$ where $y^2 = \alpha$, $z^2 = \beta$, and $yz = -zy$.

PROPOSITION 24.48. *Suppose F has a primitive n-th root ζ of 1.*

(i) $(\alpha, 1)_n \sim 1$ *for any* $\alpha \in F$.
(ii) $(\alpha, \beta) \sim 1$ *iff* β *is a norm from* $F[\sqrt[n]{\alpha}]$ *to* F.
(iii) $(\alpha, \beta_1) \otimes (\alpha, \beta_2) \sim (\alpha, \beta_1\beta_2)$.
(iv) $(\alpha, \beta) \cong (\beta, \alpha^{-1}) \cong (\beta, \alpha)^{\mathrm{op}}$.

Proof. (i) and (ii) follow from Remark 24.46.

(iii) Let $R_i = (\alpha, \beta_i)$ for $i = 1, 2$, and take $y_i \in K$, $z_i \in R_i$ with $z_i y_i = \zeta y_i z_i$, $y_i^n = \alpha$, and $z_i^n = \beta_i$. Let $\hat{z} = z_1 \otimes z_2$, which together with $y_1 \otimes 1$ generate a subalgebra A of $R_1 \otimes R_2$ isomorphic to $(\alpha, \beta_1\beta_2)$. By direct computation, the centralizer B of A contains $y_1 \otimes y_2^{-1}$, whose minimal polynomial $\lambda^n - 1$ is divisible by $\lambda - 1$; hence, B is split by Proposition 24.44.

(iv) If $zy = \zeta yz$, then $y^{-1}z = \zeta z y^{-1}$, implying that $(\alpha, \beta) \cong (\beta^{-1}, \alpha)$, and the last isomorphism follows from (i) and (iii). □

Let us generalize Proposition 24.48 for arbitrary cyclic algebras.

PROPOSITION 24.48′. $(K, \sigma, \beta_1) \otimes (K, \sigma, \beta_2) \sim (K, \sigma, \beta_1\beta_2)$.

Proof. Let $R = (K, \sigma, \beta_1) \otimes (K, \sigma, \beta_2)$, which contains $\tilde{K} = K \otimes_F K$. We define the map $\varphi\colon \tilde{K} \to K$ by $a \otimes a' \mapsto aa'$. Since $\tilde{K}$ is semisimple (cf. Example 18.29), $\ker\varphi = \tilde{K}e'$ for some idempotent e', a sum of $n - 1$ orthogonal primitive idempotents. Now let $\hat{\sigma} = \sigma \otimes \sigma$. If $\sum a_i \otimes a_i' \in \ker\varphi$, then

$$\varphi\left(\hat{\sigma}\left(\sum a_i \otimes a_i'\right)\right) = \sum \sigma(a_i)\sigma(a_i') = \sigma\left(\sum a_i \otimes a_i'\right) = \sigma(1) = 1.$$

Hence $\ker\varphi$ is invariant under $\hat\sigma$, so $\hat\sigma(e')$ is also an idempotent of $\ker\varphi$, implying $\hat\sigma(e') = e'$. Thus $\hat\sigma$ also fixes $e = 1 - e'$, also an idempotent of $\hat K$. But $\sigma^i \otimes \sigma^j$ does not fix e for $0 \le i < j \le n$, since $\sigma^i \otimes \sigma^j(a \otimes a^{-1}) = \sigma^i(a)\sigma^j(a^{-1}) \ne 1$ when $\sigma(a) \ne a$. Since $R = \bigoplus \tilde K(z_1^i \otimes z_2^j)$, the centralizer of e in R is generated by $\tilde K$ and $\hat z = z_1 \otimes z_2$. Let $\hat K = Ke \cong K$. Clearly, eRe is generated by $\hat K$ and $e\hat z e$, which acts as $\hat\sigma$ on $\hat K$. This displays eRe as $(\hat K, \hat\sigma, \beta_1\beta_2)$, so we are done by Remark 24.12′. □

Separable subfields.

Since not every csa is cyclic, we look for other field extensions with which to define csa's. Crossed products were already constructed in Example 24.7.

Example 24.49. A csa R is a crossed product, iff R has a maximal subfield K that is a Galois extension of the center F.

Indeed, let $G = \mathrm{Gal}(K/F)$. By the Skolem-Noether Theorem, each $\sigma \in G$ is given by conjugation by some $z_\sigma \in R^\times$; i.e., $z_\sigma a = \sigma(a) z_\sigma, \forall a \in K$. But then

$$z_\sigma z_\tau z_{\sigma\tau}^{-1} a = z_\sigma z_\tau (\sigma\tau)^{-1}(a) z_{\sigma\tau}^{-1} = a z_\sigma z_\tau z_{\sigma\tau}^{-1},$$

implying that $z_\sigma z_\tau z_{\sigma\tau}^{-1} \in C_R(K) = K$. Letting

$$c_{\sigma,\tau} = z_\sigma z_\tau z_{\sigma\tau}^{-1},$$

we see that the $c_{\sigma,\tau}$ satisfy the factor set conditions (24.2) just as in Example 24.7 (since R is associative). The proof that $R = \oplus K z_\sigma$ is analogous to the verification of Example 24.42; cf. Exercise 22.

A major question for many years was, "Is every cda a crossed product?" This is true in general for degrees 2, 3, 4, 6, and 12, cf. Exercise 20, but was answered negatively by Amitsur [**Am4**] for n divisible by 8 or by the square of an odd prime, as to be discussed below. However, every cda does have maximal subfields separable over F. This is obvious in characteristic 0, for then every finite field extension is separable. Thus, our treatment must involve field theory in characteristic p. It is convenient to bring in polynomial identities at this stage.

Proposition 24.50. *Any F-csa R is PI-equivalent to $M_n(F)$ for $n = \deg R$.*

Proof. In view of Theorem 24.42, one may assume that F is infinite. But now, for $\bar F$ the algebraic closure of F, Proposition 23.44 implies that R is PI-equivalent to $R \otimes_F \bar F \cong M_n(\bar F)$, which is PI-equivalent to $M_n(F)$. □

Theorem 24.51 (Koethe-Noether-Jacobson). *Any separable subfield L of a cda D is contained in a separable maximal subfield of D.*

Proof. In view of Corollary 24.27, we may replace D by $C_D(L)$, and thereby assume that $F = L$. Likewise, take a separable subfield K of D of greatest possible dimension; replacing D by $C_D(K)$, we may assume that $F = K$. Let $n = \deg D$. Take arbitrary $d \in D$, and take its minimal polynomial. By Remark 4.36 of Volume 1, the minimal polynomial of any element d of D has the form $g(\lambda^{p^t})$ with g separable. Then $g(d^{p^t}) = 0$, implying that d^{p^t} is separable over F, contrary to our assumption unless $d^{p^t} \in F$. This means $\deg d = p^t \le n$; applying this argument for each $d \in D$ shows that there is some t such that $d^{p^t} \in F$ for all $d \in D$. In other words, D satisfies the central polynomial x^{p^t}.

But now Proposition 24.50 implies that x^{p^t} is a central polynomial for $M_n(F)$, which is absurd since the matrix unit e_{11} is idempotent. □

Note that Theorem 24.51 holds over all fields. Another proof using derivations is given in Exercises 9–11. A third proof (actually of a stronger assertion) is given in Exercise 18.

Corollary 24.52. *Every csa is similar to a crossed product.*

Proof. Let E be the Galois closure of a separable maximal subfield K of D. Then E/F is Galois, but E is a maximal subfield of $M_{[E:K]}(D)$ by Corollary 24.37. □

Here is another quick application of Proposition 24.50 for later use.

Proposition 24.53. *Any csa R has elements $a, r_1, \dots, r_n$ such that $\{a^{i-1} r_j : 1 \le i, j \le n\}$ comprise a base of R over F.*

Proof. Otherwise, by Proposition 23.12, the polynomial

$$c_{n^2}(x_1, x_1x_0, \dots, x_1x_0^{n-1}, \dots, x_n, x_nx_0, \dots, x_n^{n-1}x_0; y_1, \dots, y_{n^2})$$

is an identity of R and thus of $M_n(F)$. But this is false, as seen by specializing $x_0 \mapsto a = e_{n1} + \sum_{i=1}^{n-1} e_{i,i+1}$ and $x_j \mapsto r_j = e_{jj}$ for $1 \le j \le n$. □

Central simple algebras in characteristic p.

As with the other topics we have discussed, the characteristic p theory has its own flavor, sometimes richer than the characteristic 0 theory. As we noted above, the center F of any cda D of characteristic p must be infinite. In the important case for $D = (K, \sigma, \beta)$ cyclic of degree $n = p$, then K/F is a cyclic (separable) field extension, whereas $F[\beta^{1/n}]/F$ is a purely inseparable field extension. This interplay between separable and inseparable produces a more malleable theory than in the characteristic 0 case.

For example, we define a csa R to be a p-**algebra** when $\deg R$ is a power of $p = \mathrm{char}(F)$. It is known that every p-algebra is similar in the Brauer group to a cyclic algebra, whereas this is false in characteristic 0. Exercises 26–31 provide some of the flavor of this theory; also see p. 479.

The generic division algebra.

The quickest example of a noncrossed product is the **generic** or **universal** division algebra $\mathrm{UD}(n, F)$, defined in Example 24.6 as the ring of central fractions of the algebra of generic matrices $F\{Y\}_n$ of Definition 23.49 and Corollary 23.52.

THEOREM 24.54. *$\mathrm{UD}(n, F)$ is a division algebra of degree n (over its center) for every n and every field F of characteristic prime to n.*

Proof. By Proposition 23.51, $F\{Y\}_n$ is relatively free for the class of algebras PI-equivalent to $M_n(F)$, and in particular is PI-equivalent to any F-division algebra of PI-class n.

We need to show that $\mathrm{UD}(n, F)$ is a division ring. First we note by Corollary 23.52 that $F\{Y\}_n$ is a prime ring. (of dimension n^2) over its center by Corollary 23.34, so to conclude the proof, it suffices to show that $UD(n, F)$ has no nonzero nilpotent elements.

Write a nilpotent element as fc^{-1} for $f \in F\{Y\}_n$ and $c \in \mathrm{Cent}(F\{Y\}_n)$. Then $f^n = 0$. Now take any F-division algebra D of degree n (not necessarily with center F), such as the generic symbol of Example 24.5. If f were nonzero, then f would correspond to a non-identity of D, so we could find a homomorphism $F\{Y\}_n \to D$ for which the image $\overline{f}$ of f is nonzero. But $\overline{f}^n = \overline{f^n} = 0$, contrary to D being a division algebra. Hence f must be 0.

□

The center of $F\{Y\}_n$ is much larger than F, since it contains all evaluations of central polynomials. Let us elaborate on a key point in the last proof.

Remark 24.55. Any given set of t nonzero elements of $\mathrm{UD}(n, F)$ can be written in the form $\frac{f_1(y_1,\dots,y_m)}{g(y_1,\dots,y_m)}, \dots, \frac{f_t(y_1,\dots,y_m)}{g(y_1,\dots,y_m)}$, where $0 \neq f_i, g \in F\{Y\}_n$ with g central. Let $f = f_1 \cdots f_m g \neq 0$. Then, for any cda D over F there are $d_1, \dots, d_m \in D$ such that $f(d_1, \dots, d_m) \neq 0$, i.e.,

$$\frac{f_1(d_1,\dots,d_m)}{g(d_1,\dots,d_m)}, \dots, \frac{f_t(d_1,\dots,d_m)}{g(d_1,\dots,d_m)} \neq 0. \qquad \square$$

In other words, any given finite set of elements of $\mathrm{UD}(n, F)$ satisfying certain equalities and inequalities can be transferred to elements of any cda D.

Example 24.56 (Amitsur's noncrossed product). If $\mathrm{UD}(n, F)$ is a crossed product with respect to a certain group G, then every cda D of degree n is a crossed product with respect to G; cf. Exercise 32. But for $n = p^t$ and $\mathrm{char}(F) \neq p$, we see in Exercise 36 that the tensor product of generic symbols of degree p (cf. Example 24.5, built with distinct indeterminates) is a crossed product **only** with respect to the group ${C_p}^{(t)}$, the direct product of t cyclic groups of order p. On the other hand, by the same argument, for $t \geq 3$, the generic symbol of degree n is **never** a crossed product with respect to ${C_p}^{(t)}$. Confronting $\mathrm{UD}(n, F)$ with these two examples shows that for $\mathrm{char}(F) \neq p$, $\mathrm{UD}(n, F)$ cannot be a crossed product whenever $p^3 \mid n$. (This argument can be extended to characteristic p but requires more work; cf. Saltman [**Sal1**] and McKinnie [**McK**].)

These ideas are very powerful, but leave us in the dissatisfying position of having a stunning example, $\mathrm{UD}(n, F)$, for which we know little about its center. For $n = 2$, one can describe its center as a purely transcendental field extension of F; cf. Exercise 37. However, this case is deceptively easy; the analogous assertion, proved (with considerably more effort) by Formanek for $n = 3$ and $n = 4$, remains open in general.

The exponent

$\mathrm{Br}(F)$ is a torsion group. More precisely, the order of any element in $\mathrm{Br}(F)$ divides its index. In preparation for the proof, we return to subfields.

THEOREM 24.57. *Suppose D is a cda of degree $p^u q$, where p is prime, $p \nmid q$, and K is a maximal separable subfield of D. Then the normal closure of K contains a field extension L of F with $p \nmid [L:F]$, as well as a splitting field $L_u \supseteq L$ of D together with a sequence of subfields*

$$L_0 = L \subset L_1 \subset L_2 \subset \cdots \subset L_u \tag{24.5}$$

for which $\mathrm{ind}(D \otimes_F L_i) = p^{u-i}$ for each $0 \leq i \leq u$, and each L_i/L_{i-1} is cyclic Galois of dimension p. (Thus, $[L_i : L] = p^i$ for each i.)

Proof. $[K : F] = \deg D = p^u q$. The normal closure E of K is a Galois extension whose Galois group G has some order $p^{u'}q'$, where $u' \geq u$ and q' is prime to p. Let $L = E^H$, where H is a p-Sylow subgroup of G. Then $[L:F] = [G:H] = q'$.

$D \otimes_F L$ is split by E, and so has index dividing $[E:L] = p^{u'}$, a p-power. But the index reduction factor of D under L divides $[L : F]$ and thus is prime to p, so we conclude that $\mathrm{ind}(D \otimes_F L) = p^u$. Let $L_u = KL \subseteq E$. Then $[L_u{:}F]$ is divisible both by $[K{:}F] = p^u q$ and by $[L{:}F] = q'$, implying that

$[L_u{:}F] = p^u q'$; hence, $[L_u{:}L] = p^u$. We have the desired sequence (24.5), in view of Lemma 4.86 of Volume 1.

The index reduction factors of each $D \otimes L_i$ under L_{i+1} divide p, and their product is p^u since L_u splits D. Hence, the index reduction factor at every stage is p, and the theorem follows. □

COROLLARY 24.58. *If D is a cda of degree $p^u q$, then there exists a field extension L_{u-1} of dimension $p^{u-1}q'$ over F with $p \nmid q'$, such that $D \otimes_F L_{u-1}$ is cyclic of index p.*

Proof. Take L_{u-1} in the theorem. □

Definition 24.59. The **exponent** of a csa R, written $\exp(R)$, is the order of $[R]$ in $\mathrm{Br}(F)$. The exponent of R is also called the **order** or the **period**.

In other words, $\exp(R)$ is the smallest positive integer such that $R^{\otimes m}$ is split. Since we may replace R by its underlying division algebra, we assume that R is a division algebra D. The next case, although rather easy, is the critical one.

LEMMA 24.60. *Any cyclic division algebra D of index p has exponent p.*

Proof. Write $D = (K, \sigma, \beta)$, where $\beta \in F$. Then, by Proposition 24.48′ and Remark 24.46(iii), $D^{\otimes p} \sim (K, \sigma, \beta^p) \sim 1$, since $\beta^p = N_{K/F}(\beta)$. □

This leads readily to the following generalization.

LEMMA 24.61. *If $\mathrm{ind}(D) = p^u q$ with p prime and $p \nmid q$, then $\mathrm{ind}(D^{\otimes p^u})$ is not divisible by p.*

Proof. Induction on u; the case $u = 0$ is immediate, so assume that $u \geq 1$. By Corollary 24.58, there is a field L' of dimension $p^{u-1}q'$ over F with $p \nmid q'$ such that $D \otimes_F L'$ is cyclic of degree p (over L'). Thus

$$D^{\otimes p} \otimes_F L' \cong (D \otimes L')^{\otimes p} \sim L',$$

implying that the index of $D^{\otimes p}$ divides $[L'{:}F]$ whose p-power part is p^{u-1}. By induction, p does not divide

$$\mathrm{ind}((D^{\otimes p})^{\otimes p^{u-1}}) = \mathrm{ind}(D^{\otimes p^u}).$$ □

THEOREM 24.62.

(i) *For every csa R, $\exp(R)$ divides $\mathrm{ind}(R)$. In particular, $\mathrm{Br}(F)$ is a torsion Abelian group.*

(ii) *If a prime number p divides $\mathrm{ind}(R)$, then p divides $\exp(R)$.*

Proof. (i) Write $\mathrm{ind}(R) = n = p_1^{u_1} \cdots p_t^{u_t}$. We need to show that $R^{\otimes n} \sim 1$. Fixing j for the moment, let $m = p_j^{u_j}$. By Lemma 24.61, $\mathrm{ind}(R^{\otimes m})$ is not divisible by p_j; hence, $\mathrm{ind}(R^{\otimes m})$ divides $\frac{\mathrm{ind}(R)}{m} = p_1^{u_1} \cdots p_{j-1}^{u_{j-1}} p_{j+1}^{u_{j+1}} \cdots p_t^{u_t}$. By induction on t, $\exp(R^{\otimes m})$ divides $\mathrm{ind}(R^{\otimes m})$. It is clear that $\exp(R^{\otimes n})$ divides $\exp(R^{\otimes m})$ and thus divides $p_1^{u_1} \cdots p_{j-1}^{u_{j-1}} p_{j+1}^{u_{j+1}} \cdots p_t^{u_t}$; since this holds for each j, $\exp(R^{\otimes n}) = 1$.

(ii) Using Corollary 24.58, take L such that $\mathrm{ind}(R \otimes_F L) = p$. Then $\exp(R \otimes_F L)$ is p (since it cannot be 1), so p divides $\exp(R)$. (This is clear when we view $\mathrm{res}_{L/F}$ as a homomorphism $\mathrm{Br}(F) \to \mathrm{Br}(L)$.) □

Definition 24.63. $\mathrm{Br}(F)_m$ is the subgroup of $\mathrm{Br}(F)$ consisting of algebras of exponent m.

COROLLARY 24.64. *If $[L{:}F]$ is relatively prime to m, then $\mathrm{res}_{L/F}$ yields a 1:1 map from $\mathrm{Br}(F)_m$ to $\mathrm{Br}(L)_m$.*

Proof. If $[D] \in \ker \mathrm{res}_{L/F}$, then L splits D, but the index reduction factor divides $[L{:}F]$ and $\mathrm{ind}(D)$, which are relatively prime in view of Theorem 24.62(ii); thus, $D \sim 1$. □

We can squeeze out even more information.

COROLLARY 24.65. *If D_1 and D_2 are cda's of relatively prime index, then $D = D_1 \otimes_F D_2$ is a division algebra.*

Proof. Let $n_i = \mathrm{ind}\, D_i$ and $n = \mathrm{ind}(D)$. Clearly, $n \mid n_1 n_2$. Let K be a maximal subfield of D_1.

$$D \otimes_F K \cong M_{n_1}(K) \otimes_K (K \otimes_F D_2) \sim D_2 \otimes_F K,$$

a division algebra of degree n_2 over K (since $[K{:}F] = n_1$ is prime to $\deg D_2$), implying that $n_2 \,|\, n$. Likewise, $n_1 \,|\, n$, yielding $n = n_1 n_2$. □

THEOREM 24.66. *Any cda D is isomorphic to the tensor product of cda's of prime power index.*

Proof. Let $n = \mathrm{ind}(D)$. It suffices to show that if $n = n_1 n_2$ for n_1 and n_2 relatively prime, then $D \cong D_1 \otimes D_2$, where $\mathrm{ind}\, D_i = n_i$. But letting $m = \exp(D)$, which divides n, we have $m = m_1 m_2$ where $m_i = \gcd(m, n_i)$, so a trivial observation in group theory (Remark 0.0 of Volume 1) shows that $D \sim D_1 \otimes D_2$ in the Brauer group, for division algebras D_i of exponent m_i. But Corollary 24.65 says that $D_1 \otimes D_2$ is a division algebra, which must then be D. □

Thus, the theory of cda's reduces to the prime power case. From this point of view, the key to the theory is the case of prime exponent p.

In Exercise 33, we take the tensor product of two generic quaternion algebras to build a noncyclic algebra of degree 4 and exponent 2. Taking tensor products of generic symbols, one likewise sees that for any prime powers m and n, where $m \mid n$, there is a division algebra of degree n and exponent m; cf. Exercise 35. Thus, in light of Theorem 24.66, Theorem 24.62 gives all possible numerical restrictions on the degree and exponent of a csa.

Techniques generalized from field theory

We bring in some standard techniques from commutative algebra which, surprisingly, can be applied even in the noncommutative setting.

The reduced characteristic polynomial.

We start with a modification of the characteristic polynomial of a matrix. The method follows Remark 4.104ff. of Volume 1, but here we use the injection $R \hookrightarrow M_n(\bar{F})$ of Remark 24.19 in conjunction with the regular representation.

Remark 24.67. Suppose R is a csa of degree n, and fix a base $b_1, \dots, b_{n^2}$ of R. As in Example 23.53, we build the **generic element**

$$\hat{r} = \sum_{i=1}^{n^2} \lambda_i b_i$$

in the polynomial algebra $R[\Lambda] = R[\lambda_1, \dots, \lambda_{n^2}]$ whose center clearly is $F[\Lambda] = F[\lambda_1, \dots, \lambda_{n^2}]$.

(i) For any element $a = \sum \alpha_i b_i \in R$, we have the substitution homomorphism $\psi_a \colon R[\Lambda] \to R$ given by $\lambda_i \mapsto \alpha_i$; clearly $\psi_a(\hat{r}) = a$.

(ii) Let $F(\Lambda) = F(\lambda_1, \dots, \lambda_{n^2})$, the field of fractions of $F[\Lambda]$, and denote $R(\Lambda) = R \otimes_F F(\Lambda)$. Viewing R inside $M_{n^2}(F)$ via the regular representation, we also view $R(\Lambda) \subseteq M_{n^2}(F(\Lambda))$. We denote the characteristic polynomial of $\hat{r}$ (in $M_{n^2}(F(\Lambda))$) as $\tilde{f} = \det(xI - \hat{r})$, a monic polynomial in $F[\Lambda][x]$, seen via the usual formula for determinant. Actually, we are interested in the minimal monic polynomial of $\hat{r}$ over $F(\Lambda)$, which we denote as $\hat{f}$. Clearly $\hat{f}$ divides $\tilde{f}$ in $F(\Lambda)(x)$; hence, $\hat{f} \in F[\Lambda][x]$ by Gauss' Lemma, since $F[\Lambda]$ is normal (Example 6.45 of Volume 1). Having shown that the coefficients $\hat{f}$ are in $F[\Lambda]$, we may forget $M_{n^2}(F(\Lambda))$ for the remainder of this discussion.

(iii) Now we turn to the algebraic closure $\bar{F}$ of F. Identifying $M_n(\bar{F})$ with $R \otimes_F \bar{F}$, we view R as $R \otimes_F 1 \subset M_n(\bar{F})$. Remark 24.19 implies that $\hat{f}$

remains the minimal polynomial of $\hat{r}$ over $\bar{F}$, and in particular $\deg \hat{f} \leq n$. Likewise, $b_1, \dots, b_{n^2}$ is also a base for $M_n(\bar{F})$ over $\bar{F}$, so $\hat{r}$ also serves as a generic element in $M_n(\bar{F}[\Lambda])$, and $M_n(\bar{F})$ obviously has an element a of degree n. By (i), $\hat{r}$ specializes to a; hence, $n = \deg a \leq \deg \hat{f} \leq n$, implying that $\deg \hat{f} = n$. On the other hand, $\hat{r}$, viewed in $M_n(\bar{F}[\Lambda])$, is a root of its characteristic polynomial, which has degree n and thus must be $\hat{f}$. This proves that $\hat{f}$ is the characteristic polynomial of $\hat{r}$, viewed in $M_n(F[\Lambda])$.

(iv) For any $a \in R$, we define its **reduced characteristic polynomial** f_a to be the monic polynomial $\psi_a(\hat{f}) \in F[x]$. By (iii), f_a is the characteristic polynomial of $a \otimes 1$ in $M_n(\bar{F})$, which is well-defined, independent of the choice of base of R.

(v) If R is a division algebra, the reduced characteristic polynomial f_a is a power of the minimal polynomial of a, by Remark 4.106 of Volume 1, applied to the injection $R \hookrightarrow M_n(\bar{F})$.

Definition 24.67′. Writing $f_a = \lambda^n + \sum_{i=0}^{n-1} \alpha_i \lambda^i$ for the reduced characteristic polynomial of a, define the **reduced trace** $\mathrm{tr}_{\mathrm{red}}(a)$ to be $-\alpha_{n-1}$ and the **reduced norm** $\mathrm{N}_{\mathrm{red}}(a)$ to be $(-1)^n \alpha_0$.

In particular, $\mathrm{tr}_{\mathrm{red}}(r), \mathrm{N}_{\mathrm{red}}(r) \in F$. The reduced trace and norm are powerful tools, as seen in Exercises 15–20.

Wedderburn's factorization method.

Perhaps the main technique used in Volume 1 for studying algebraic field extensions is by means of polynomials and their factorizations. We considered polynomials over division rings briefly in Proposition 13.53ff. For the reader's convenience, we review some notation from [**Row3, p. 227**] concerning polynomials over a division ring D. Given $f = \sum d_i \lambda^i \in D[\lambda]$ and $d \in D$, we write $f(d)$ for $\sum d_i d^i$, and say d is a **root** of f if $f(d) = 0$. We say a polynomial $g \in D[\lambda]$ **(right) divides** f if $f = hg$ for some $h \in D[\lambda]$. The Euclidean algorithm shows that $\lambda - d$ divides f iff d is a root of f. Wedderburn proved a noncommutative analog of the Fundamental Theorem of Algebra:

THEOREM 24.68 ([**Row3, Appendix B, Theorem 10**]). *Suppose D is a cda. If $a \in D$ is a root of a monic irreducible polynomial $f \in F[\lambda]$ of degree n, then*

$$f = (\lambda - a_n) \cdots (\lambda - a_1) \tag{24.6}$$

in $D[\lambda]$, where each a_i is a conjugate of a.

Proof. Rather than repeat the computational proof given in [**Row3**], we present a proof of Jacobson, that utilizes the structure theory of rings and modules.

By Proposition 13.53, $D[\lambda]$ is a principle left ideal domain, and any (two-sided) ideal has the form $D[\lambda]f$ for $f \in F[\lambda]$. It follows at once that $D[\lambda]f$ is a maximal ideal iff f is irreducible in $F[\lambda]$. The simple ring $W = D[\lambda]/D[\lambda]f$ has rank n as a free module over D, so W has finite composition length, and thus is a simple Artinian ring. Since $f(a) = 0$, the polynomial $\lambda - a$ (right) divides f, implying that $M = D[\lambda]/D[\lambda](\lambda - a)$ is a W-module. But $M \cong D$ is simple; hence, every simple W-module is isomorphic to M.

Now we follow the proof of Proposition 3.8 of Volume 1. We take a composition series

$$W = W_0 \supset W_1 \supset \cdots \supset W_n = 0.$$

Since $D[\lambda]$ is a PLID, each W_i has the form $D[\lambda]f_i/D[\lambda]f$, where $f_0 = 1$ and f_{i-1} divides f_i for each i. Dividing out by the leading coefficient in D, we may take each f_i monic. We write $f_i = g_i f_{i-1}$ for monic $g_i \in D[\lambda]$. Since W is simple Artinian, the factors $W_{i-1}/W_i \cong D[\lambda]f_{i-1}/D[\lambda]f_i$ are simple and thus isomorphic to M.

The map $D[\lambda] \to W_{i-1}/W_i$ given by $h \mapsto hf_{i-1} + W_i$ has kernel $D[\lambda]g_i$, so, by Noether's isomorphism theorem,

$$D[\lambda]/D[\lambda]g_i \cong D[\lambda]f_{i-1}/D[\lambda]f_i \cong W_{i-1}/W_i \cong M.$$

Hence, $\deg g_i = 1$ and $g_i = \lambda - a_i$ for $a_i \in D$.

Now $f = g_n f_{n-1} = \cdots = g_n \cdots g_1 = (\lambda - a_n)\cdots(\lambda - a_1)$. It remains to verify the last assertion. By Remark 13.54(ii), we have $u_i, v_i \in D$ with

$$u_i\lambda - u_i a = u_i(\lambda - a) = (\lambda - a_i)v_i = v_i\lambda - a_i v_i.$$

Comparing coefficients of λ shows $u_i = v_i$ and thus $v_i a = a_i v_i$, proving that a_i is conjugate to a. □

This remarkable result was pushed even further by Jacobson; cf. Exercises 45 and 46, and has many striking applications, some of which are given in Exercise 39ff. Here is an application to the reduced characteristic polynomial.

Corollary 24.69. *For any cda D of degree n and any $a \in D$, $\mathrm{tr}_{\mathrm{red}}(a)$ is a sum of n conjugates of a, and $\mathrm{N}_{\mathrm{red}}(a)$ is a product of n conjugates of a.*

This has a cute application in Exercise 44.

Galois descent and the corestriction map

We return to the Brauer group. Corollary 24.64 leads us to try to reverse the map $\mathrm{res}_{L/F}$. In other words, given a csa A over L, can we find a csa R over F such that $A \cong R \otimes_F L$? In the text, we focus on the case where $L = E$ is a Galois extension of F; see Exercise 47 for the separable case. In the Galois case, for $A = R \otimes_F E$ and for every $\sigma \in \mathrm{Gal}(E/F)$, we see that $1_R \otimes \sigma$ is an automorphism of A. Conversely, we have the following result.

Proposition 24.70. *Suppose G is a group of automorphisms on a (not necessarily simple) E-algebra A, which restricts to distinct automorphisms of E, and let $F = E^G$. Then*

$$A \cong A^G \otimes_F E. \tag{24.7}$$

Proof. First we show that the multiplication map $\mu\colon A^G \otimes_F E \to A$ is 1:1. Assume that $\sum r_i \otimes \gamma_i \in \ker \mu$ for $r_i \in A^G$ and $\gamma_i \in E$; we may assume that the r_i are independent over F. Multiplying on the right by arbitrary $\gamma \in E$, applying μ, and taking the trace with respect to the extension E/F yields

$$0 = \mathrm{tr}_{E/F}\Big(\sum_i r_i\gamma_i\gamma\Big) = \sum_i \sum_{\sigma\in G} \sigma(\gamma_i\gamma)r_i = \sum_i \mathrm{tr}(\gamma_i\gamma)r_i,$$

implying for each i that $0 = \mathrm{tr}(\gamma_i\gamma) = \sum_{\sigma\in G}\sigma(\gamma_i)\sigma(\gamma)$ for each $\gamma \in E$, and hence $\sigma(\gamma_i) = 0$ for each $\sigma \in G$, by Dedekind's Independence Theorem (Theorem 4.58 of Volume 1); in particular, each $\gamma_i = 0$. Thus $\ker\mu = 0$.

Next, we claim that μ is onto. Otherwise, taking $a \in A \setminus A^G E$, one could define an E-vector space map $f\colon A \to E$ with $f(a) = 1$ and $f(A^G E) = 0$; then $\sum_{\sigma\in G} f(\sigma(a))\sigma(\gamma) = f(\mathrm{tr}(\gamma a)) = 0$ for any $\gamma \in E$, again implying by Dedekind's Independence Theorem that $f(\sigma(a)) = 0$ for each σ, contrary to $f(a) = 1$. □

Corollary 24.71. *Hypotheses as in Proposition 24.70, if A is an E-csa, then A^G is an F-csa and $[A : E] = [A^G : F]$.*

Proof. The isomorphism (24.7) shows that A^G is simple, and

$$\mathrm{Cent}(A) \cong \mathrm{Cent}(A)^G \otimes_F E$$

implies $\mathrm{Cent}(A^G) = F$. The equality of dimensions follows from (24.7). □

Unfortunately, $\mathrm{Gal}(E/F)$ need not extend to a group of automorphisms of an E-csa A, so we need to work harder.

Definition 24.72. Suppose E is a (finite) Galois extension of the field F. For any vector space V over E and for $\sigma \in \mathrm{Gal}(E/F)$, define $\sigma(V)$ to be V as an F-vector space, but with new scalar multiplication over E given by

$$\gamma \cdot v = \sigma^{-1}(\gamma)v, \qquad \forall \gamma \in E, \forall v \in V. \tag{24.8}$$

Let $G = \mathrm{Gal}(E/F) = \{\sigma_i : 1 \le i \le t\}$, where $t = [E:F]$. Given any E-csa A, G acts naturally on $\widehat{A} = \bigotimes_{i=1}^{t} \sigma_i(A)$ as follows: Given $\sigma \in G$, write $\sigma_i \sigma = \sigma_{\pi_\sigma i}$ for a suitable permutation $\pi_\sigma \in S_t$, and define

$$\sigma\left(\otimes_i a_i\right) = \otimes_i a_{\pi_\sigma i}, \qquad a_i \in \sigma_i(A). \tag{24.9}$$

By Corollary 24.71, the fixed subalgebra $\widehat{A}^G$, which we call the **corestriction** $\mathrm{cor}_{E/F}(A)$, is an F-csa whose degree is $(\deg A)^t$.

PROPOSITION 24.73. *$\mathrm{cor}_{E/F}$ induces a homomorphism of Brauer groups.*

Proof. We need to show that if A_1 and A_2 are central simple over E, then

$$\widehat{A_1 \otimes_E A_2}^{\,G} \cong \widehat{A}_1^G \otimes_F \widehat{A}_2^G, \tag{24.10}$$

where G acts diagonally on $\widehat{A_1 \otimes_E A_2}$. The inclusion ($\supseteq$) in (24.10) is clear, so we prove ($\subseteq$). Corollary 24.71 implies

$$[\widehat{A}_1^G : F][\widehat{A}_2^G : F] = [\widehat{A}_1 : E][\widehat{A}_2 : E] = [A_1 \otimes_E A_2 : E] = [\widehat{A_1 \otimes_E A_2}^{\,G} : F],$$

so the assertion follows from Corollary 24.32′ (taking $K = F$). □

We are aiming for the following important property of the corestriction:

THEOREM 24.74. *For any csa R over F and any Galois extension E of F, $\mathrm{cor}_{E/F}\,\mathrm{res}_{E/F}\,R \cong R^{\otimes [E:F]}$.*

By Exercise 14, for any m prime to $[E:F]$, the map $[R] \mapsto [R^{\otimes [E:F]}]$ is an automorphism of the group $\mathrm{Br}(F)_m$; Theorem 24.74 is particularly useful in this case.

Remark 24.74′. A few preliminary words about the proof of Theorem 24.74. Let $A = \mathrm{res}_{E/F}\,R = R \otimes_F E$. There is an isomorphism $\widehat{A} \to R^{\otimes t} \otimes_F E$, given by

$$\bigotimes_{i=1}^{t}(r_i \otimes \beta_i) \mapsto \left(\bigotimes_{i=1}^{t} r_i\right) \otimes \prod_{i=1}^{t} \sigma_i^{-1}(\beta_i). \tag{24.11}$$

(This is easily seen to be an algebra surjection and thus is an isomorphism since the dimensions of both sides are equal.) Hence, $\mathrm{cor}_{E/F}\,A \approx \widehat{A}^G \approx (R^{\otimes t} \otimes_F E)^G$, which we want to identify with $R^{\otimes t}$. Although $E^G = F$, we cannot conclude the proof so easily, since the action of G on the components is twisted by the various automorphisms σ_i. Our task is to find some method of "untwisting" the action.

Remark 24.75. For any matrix $r \in M_n(F)$, the trace map satisfies the formula $\mathrm{tr}(r) \cdot 1 = \sum_{i,j} e_{ij} r e_{ji}$.

To generalize this observation to an arbitrary csa R, we turn to the isomorphism $\mathrm{End}_F R \cong R \otimes_F R^{\mathrm{op}}$ of Theorem 24.14. (Actually, we are borrowing an idea from proofs in Azumaya algebras; cf. [**KnuO**] and [**Sal2**].) Identifying R^{op} with R as vector spaces over F, we have an F-vector space isomorphism $\mathrm{End}_F R \cong R \otimes_F R$. Since the reduced trace $\mathrm{tr}_{\mathrm{red}} \colon R \to F$ can be viewed as an endomorphism from R to R, it corresponds to a *unique* element of $R \otimes_F R$.

Definition 24.76. The element of $R \otimes_F R$ corresponding to $\mathrm{tr}_{\mathrm{red}}$ is called the **Goldman element for** R, which we denote t_R.

In other words, $t_R = \sum a_i \otimes b_i$ satisfies $\sum a_i r b_i = \mathrm{tr}_{\mathrm{red}}(r)$, $\forall r \in R$.

Example 24.77. In view of Remark 24.75, the Goldman element $t_{M_n(F)} = \sum_{i,j} e_{ij} \otimes e_{ji}$. Note that $t^2_{M_n(F)} = \sum_{i,j} e_{ii} \otimes e_{jj} = 1$.

PROPOSITION 24.78. *The Goldman element t_R satisfies the following properties:*

(i) $(u \otimes v) t_R = t_R (v \otimes u)$, $\forall u, v \in R$;
(ii) $t_R^2 = 1$.

Proof. Write $t_R = \sum a_i \otimes b_i$. Then $(u \otimes v)\, t_R = \sum u a_i \otimes v b_i$ corresponds to the map

$$r \mapsto \sum u a_i r v b_i = u\, \mathrm{tr}_{\mathrm{red}}(rv),$$

whereas $t_R\,(v \otimes u) = \sum a_i v \otimes b_i u$ corresponds to the map

$$r \mapsto \sum a_i v r b_i u = \mathrm{tr}_{\mathrm{red}}(vr) u = u\, \mathrm{tr}_{\mathrm{red}}(rv),$$

proving (i).

(ii) Let K be a splitting field for R. The Goldman element $t_{R \otimes K} = t_R \otimes 1$, but (being unique, and since $R \otimes K$ is split) $t_{R \otimes K}$ was already computed in Example 24.77 and has square 1. Thus $t_R^2 \otimes 1 = (t_R \otimes 1)^2 = 1 \otimes 1$, implying $t_R^2 = 1$. □

Proof of Theorem 24.74. We can extend the Goldman element to an arbitrary tensor power $\widehat{R} = R_1 \otimes \cdots \otimes R_t$ of R, where each $R_i = R$. Namely, we define an action of S_t on $\widehat{R}$ by

$$\pi(r_1 \otimes \cdots \otimes r_t) = r_{\pi 1} \otimes \cdots \otimes r_{\pi t}.$$

For any transposition $\tau = (i\,j)$ we have the corresponding Goldman element in

$$1 \otimes \cdots \otimes 1 \otimes R_i \otimes 1 \otimes \cdots \otimes 1 \otimes R_j \otimes 1 \otimes \cdots \otimes 1,$$

which now we denote as t_τ; then $t_\tau \hat{r} = \tau(\hat{r}) t_\tau$ for $\hat{r} \in \widehat{R}$. Thus, for any permutation written as a product of permutations $\pi = \tau_1 \tau_2 \cdots \tau_m$, we can define $t_\pi = t_{\tau_1} t_{\tau_2} \cdots t_{\tau_m}$ and get

$$t_\pi(\hat{r}) = \pi(\hat{r}) t_\pi.$$

Furthermore, since each t_τ is invertible (and in fact has square 1), t_π is also invertible, so $\pi(\hat{r}) = t_\pi \hat{r} t_\pi^{-1}$.

But the action π was just the twist that was obstructing the proof of Theorem 24.74, as explained in Remark 24.74′. Explicitly, for any σ in G, we associate the permutation π_σ of (24.9). Then define a new action of G on $\widehat{R} \otimes_F E$ by

$$\sigma(\hat{r}) \otimes \beta = \hat{r} t_{\pi_\sigma} \otimes \sigma^{-1}\beta.$$

The isomorphism of (24.11) sends the original action of G to this new action, which is untwisted on $R^{\otimes t}$ and acts like the Galois group on E, so the fixed subalgebra is just $R^{\otimes t}$, as desired. □

Note that Theorem 24.74 (more precisely, its version for separable field extensions) gives a new proof of Theorem 24.62. Namely, if $\operatorname{ind} R = n$, we take a separable splitting field L of R having dimension n over F; then $R^{\otimes n} \sim \operatorname{cor}_{L/F}(R \otimes_F L) \sim \operatorname{cor}_{L/F}(L) \sim 1$. In fact, the generalization of this result to Azumaya algebras is the motivation for [**Sal2**]. The corestriction is a special case of the transfer map from cohomology theory, to be discussed in Chapter 25.

Introduction to the Merkurjev-Suslin Theorem.

Despite Amitsur's construction of a noncrossed product division algebra, Merkurjev and Suslin [**MeS**] proved the major theorem that $\operatorname{Br}(F)$ is generated by symbols when F has "enough" roots of 1; more precisely, if F has a primitive m-th root of 1, then every csa of exponent m is similar to a tensor product of symbol algebras of index dividing m. (In characteristic 0, even if F lacks primitive roots of 1, Merkurjev obtained the result

for $m = 2$ and $m = 3$, and Matzri for $m = 5$, but very little is known for $m > 5$.) The Merkurjev-Suslin Theorem actually is a more general result linking K-theory to cohomology, which has been generalized by Voevodsky (unpublished) and is a focus of recent research. The only known proofs rely heavily on K-theoretic techniques and the Severi-Brauer-Chatelet-Amitsur varieties described in Appendix 24A; the theorem has been generalized quite far in K-theory. However, a few more elementary comments are in order here.

The characteristic $p > 0$ analog had been proved earlier. Combining facts about inseparable field extensions with the Frobenius automorphism and Exercise 4.70 of Volume 1, Teichmüller and Albert proved that every p-algebra of exponent m is similar to a tensor product of cyclic p-algebras of exponent m. The arguments are quite ingenious, but elementary; some of them are given in Exercises 26–31, which indicate the role of purely inseparable field extensions. This raises the tantalizing prospect that perhaps the Merkurjev-Suslin Theorem (as stated above) could be proved using reduction modulo p, but so far to no avail.

Also Rosset had proved the special case of the Merkurjev-Suslin Theorem for an F-csa R of prime index p. Indeed, by Exercise 49, R is similar to the corestriction of a symbol algebra over a suitable separable field extension L of dimension dividing $(p-1)!$ over F, and Rosset-Tate proved that this corestriction is similar to a product of at most $[L:F]$ symbol algebras of index p.

If we could extend Rosset's result to the case for $\exp(R)$ prime, then we would have the Merkurjev-Suslin Theorem by an elementary inductive procedure; cf. Exercise 50. For $p = 2$, there are accessible proofs drawing on the theory of algebras with involution since, by a theorem of Albert, any csa of exponent 2 has an involution of the first kind. Actually, this special case, proved by Merkurjev, motivated the theorem. The case for p an odd prime is apparently much harder.

One immediate consequence of the Merkurjev-Suslin Theorem (for a field with enough roots of unity) is the divisibility of the Brauer group; cf. Exercise 51.

Central simple algebras over local fields

So far we have focused on the properties of a cda or csa over an arbitrary field. But, as in commutative algebra, we also need to consider the arithmetic aspect of the theory: What is the structure of a cda over a specific field and, in particular, how close is the cda to being cyclic? This area has been active in recent years, involving a fascinating interplay of arithmetic

and geometric techniques; here we present a relatively straightforward case when F is a local field, aiming for Hasse's Theorem that every cda over a local field F is cyclic (leading to a description of the Brauer group of any local field).

Valuations on division rings.

Of course, local field theory relies heavily on the use of valuations, described in Chapter 12 of Volume 1, and our trick is to generalize these notions to division algebras over local fields.

Definition 24.79 (cf. Definition 12.15). A **valuation** on a division ring D is a group homomorphism $v\colon D^\times \to \Gamma$, where $\Gamma = (\Gamma, +)$ is an ordered group, satisfying

$$v(a+b) \geq \min\{v(a), v(b)\} \quad (\text{whenever} \quad a \neq -b).$$

Define its **value ring** $V_D = \{0\} \cup \{a \in D^\times : v(a) \geq 0\}$, **value ideal** $P_D = \{0\} \cup \{a \in D^\times : v(a) > 0\}$, and **value group** $\Gamma_D = \{v(a) : a \in D^\times\}$.

Clearly, $P_D \lhd V_D$, and its **residue ring** $\bar{D} = V_D/P_D$ is a division ring. Thus, P_D is maximal as a left ideal. When D is a cda, we can prove that Γ_D is Abelian, cf. Exercise 52, so we assume this throughout. (This is why we use additive notation for Γ.) Thus, we have

Remark 24.80. If a division ring D has a valuation v, then any two conjugates a and dad^{-1} in D have the same value, since

$$v(a) = v(d) + v(a) - v(d).$$

In particular, conjugation by any element of $D^\times$ leaves V_D and P_D invariant, and thus produces an automorphism of $\bar{D}$. Let $\mathrm{N}_{\mathrm{red}}$ denote the reduced norm from D to F; cf. Definition 24.67′. Also recall that any ordered Abelian group Γ is torsion-free.

Proposition 24.81. *If* $\deg D = n$ *and* D *has a valuation* v*, then*

$$nv(a) = v(\mathrm{N}_{\mathrm{red}}(a)) \tag{24.12}$$

for all a *in* D*.*

Proof. By Corollary 24.69, there are conjugates a_i of a such that

$$v(\mathrm{N}_{\mathrm{red}}(a)) = v(\pm a_n \cdots a_1) = v(a_n) + \cdots + v(a_1) = nv(a). \qquad \square$$

It follows that v is uniquely determined by its restriction to F, and the group Γ_D is cyclic iff Γ_F is cyclic.

Theorem 24.82 (Cohn-Wadsworth). *A cda* D *has a valuation extending a given valuation* v *on* F*, iff* v *extends uniquely to a valuation of each maximal subfield* K *of* D*.*

Proof. ($\Leftarrow$) We use Equation (24.12) to define the extension of the valuation v to a function $\hat{v}\colon D \to \mathbb{Q} \otimes_{\mathbb{Z}} \Gamma$, given by $\hat{v}(a) = \frac{1}{n} v(N_{\mathrm{red}}(a))$, which then must restrict to the unique extension of v to K. We must check that $\hat{v}$ indeed defines a valuation on D. Clearly for any a, b in D,

$$\hat{v}(ab) = \frac{1}{n} v(\mathrm{N}_{\mathrm{red}}(ab)) = \frac{1}{n} \left(v(\mathrm{N}_{\mathrm{red}}(a)) + v(\mathrm{N}_{\mathrm{red}}(b))\right) = \hat{v}(a) + \hat{v}(b).$$

It remains to consider $\hat{v}(a+b)$. Taking a maximal subfield K containing $a^{-1}b$ and noting by hypothesis that $\hat{v}$ acts as a valuation on K, we have

$$\hat{v}(1 + a^{-1}b) \geq \min\{\hat{v}(1), \hat{v}(a^{-1}b)\};$$

hence,

$$\begin{aligned} \hat{v}(a+b) &= \hat{v}(a) + \hat{v}(1 + a^{-1}b) \geq \hat{v}(a) + \min\{\hat{v}(1), \hat{v}(a^{-1}b)\} \\ &= \min\{\hat{v}(a1), \hat{v}(aa^{-1}b)\} = \min\{\hat{v}(a), \hat{v}(b)\}. \end{aligned}$$

($\Rightarrow$) Clearly, the valuation on D restricts to a valuation on K, so we want to prove that this is unique; by Corollary 12.49 of Volume 1, it suffices to prove that the valuation ring V_K is integral over V_F. Take arbitrary $a \in V_K$ together with its minimal monic polynomial f over F. Again by Wedderburn's factorization theorem, with notation as in (24.6), each $a_i \in V_K$; hence each coefficient of f, being a sum of products of the a_i, is in $V_K \cap F = V_F$. Thus, each element of V_K is integral over V_F. $\square$

This theorem shows that any division algebra D over a Henselian field F has a valuation extending the given valuation on F.

Definition 24.83. As in valuation theory, for any division algebra D with valuation $\hat{v}$ extending a valuation v of F, define the **ramification index** $e = e(D/F) = [\Gamma_D : \Gamma_F]$ and **residue degree** $f = f(D/F) = [\bar{D} : \bar{F}]$.

Remark 24.84. We appeal freely to results concerning the ramification index and residue degree for valuations for a finite field extension E of F; cf. Chapter 12 of Volume 1. Some of these proofs do not rely on the commutativity of E, and work just as well for the valuation $\hat{v}$ of D. In particular, $ef \leq [D : F]$, equality holding when the valuation v of F is discrete (implying that $\hat{v}$ is discrete on D) and F is complete with respect to v; we

leave the standard proofs for Exercises 53 and 54. We also need two easy observations, where $n = \deg(D)$:

(i) If v is discrete, then $e(D/F) \leq n$. (Indeed, suppose $\hat{v}(\Pi)$ generates the value group Γ_D. Clearly the field $F(\Pi)$ contains all the powers of Π, implying that $e(D/F) = e(F(\Pi)/F) \leq [F(\Pi) : F] \leq n$.)

(ii) If the residue field $\bar{F}$ is finite, then D has some subfield K with $f(D/F) = f(K/f) \leq n$. (Indeed, $\bar{D}$ is also finite and thus is a finite field, by Theorem 24.42; hence, the group $\bar{D}^\times$ is cyclic, generated by some element $\bar{\zeta}$. Then the field $\bar{D} = \bar{F}(\bar{\zeta})$, implying that

$$f(D/F) = f(\bar{D}/\bar{F}) = f(F(\zeta)/F) \leq n.)$$

We are ready for Hasse's Theorem.

THEOREM 24.85. *Suppose F is a local field and D is a cda of degree n. Then D is a cyclic algebra, having a maximal subfield K isomorphic to the unramified extension of F of dimension n.*

Proof. (This elegant proof was suggested by Saltman.) By Remark 24.84(ii), D has a maximal subfield K with $f(K/F) = f(D/F) \leq n$. But also $e(D/F) \leq n$ and $e(D/F)f(D/F) = n^2$, yielding $e(D/F) = f(D/F) = n$. Hence $f(K/F) = n$ but $[K{:}F] \leq n$, implying that K/F is unramified and $[K{:}F] = n$; i.e., K is a maximal subfield of D. □

Remark 24.86. In the above proof, $\bar{K} = \bar{D}$.

Hasse went on to show that if $\deg D = n$, then $D \cong (K, \sigma, \pi^m)$ for some m, cf. Exercise 58, and we can identify (K, σ, π^m) with $\frac{m}{n}$. This gives an isomorphism $\nu\colon \mathrm{Br}(F) \to (\mathbb{Q}/\mathbb{Z}, +)$ for any local field F.

Since valuations are incompatible with zero divisors, the above results are restricted to cda's. Research is underway to broaden the theory to cda's by considering "quasi-valuations."

One can get information about $\mathrm{Br}(F)$ for arbitrary fields F by taking completions of valuations. Indeed, suppose v is a nonarchimedean valuation on F. Writing F_v for the completion of F with respect to v, we have the restriction maps $\mathrm{Br}(F) \to \mathrm{Br}(F_v)$, yielding a map $\mathrm{Br}(F) \to \Pi_v \mathrm{Br}(F_v)$, the product taken over all nonarchimedean valuations. Of course, when $F \subset \mathbb{R}$, we also should throw in the restriction map $\mathrm{Br}(F) \to \mathrm{Br}(\mathbb{R}) \cong C_2$, so in this case we get the map

$$\Phi\colon \mathrm{Br}(F) \to \Pi_v \mathrm{Br}(F_v),$$

where we view $\mathbb{R}$ as one of the F_v. A major early achievement (outside the scope of this book) due to Albert, Brauer, Hasse, and Noether was

that every cda over an algebraic number field is cyclic with index equal to exponent. This idea of studying a field in terms of "local information" has been carried much further and probably is best understood in terms of cohomology theory.

One major problem is to describe the structure of any cda D whose center F has transcendence degree d over an algebraically closed field. If $d = 0$ or $d = 1$, then $D = F$ by Exercise 16, and for $d = 2$, De Jong showed that $\exp(D) = \deg(D)$. A plausible conjecture is that for $\exp(D) = p$ prime, $\deg(D)$ divides p^{d-1}. Saltman [**Sal4**] has shown that when F is a finite extension of $\mathbb{Q}_p(\lambda)$, $\deg(D)$ divides $\exp(D)^2$, and (for this F) when $\deg(D)$ is prime, D must be cyclic.

Appendix 24A: Csa's and geometry

In this appendix, we see how the lack of algebraic closure of the base field leads to some interesting geometry that in turn casts further light on csa's. One could define algebraic varieties over an arbitrary field, not necessarily algebraically closed. Unfortunately, some of the most basic properties of algebraic varieties fail; for example, the zero set of $\lambda^2 + 1$ over $\mathbb{R}$ is empty. Accordingly, given a variety V over a field F whose algebraic closure is $\bar{F}$, one considers the variety $\bar{V} = V \otimes_F \bar{F}$; one already has the theory in hand for $\bar{V}$, which one then wants to transfer to V. For convenience, we suppose $V \subset F^{(n)}$ is affine; then $\bar{V} \subset \bar{F}^{(n)}$. One can pass back to V by thinking of $\bar{V}$ as the zero set of a set of polynomials defined over F; V is the set of zeroes in $F^{(n)}$.

Going in the other direction, starting with a variety $\bar{V}$ defined over $\bar{F}$, can we cut down to a variety V defined over F, which then tensors up back to $\bar{V}$?

Definition 24A.1. For fields $F \subseteq \bar{F}$, an $\bar{F}$-variety V is **defined over** F if the ideal of polynomials defining V is generated by a set of polynomials whose coefficients are all in F.

Similarly, given a field extension $K \supset F$, an F-algebra A defined over F is called a **form** of a K-algebra B iff $B \cong A \otimes_F K$. By Remark 24.19 and Proposition 18.33, the forms of $M_n(\bar{F})$ are just the F-central simple algebras; furthermore, the forms of $M_n(K)$ are the csa's having K as a splitting field. Thus, csa's play a critical role in the study of algebraic geometry over non-algebraically closed fields, and we briefly consider one major instance.

Severi-Brauer-Chatelet-Amitsur varieties.

The splitting fields of R play a paramount role in the theory, and have geometric significance by means of the so-called "Severi-Brauer varieties" $\mathbb{SB}_R$, although Chatelet [**Ch**] and Amitsur [**Am1**] were their actual discoverers. Let us take a short excursion to describe $\mathbb{SB}_R$. The underlying question to tackle is, "How can we describe the splitting fields of a csa R in geometric terms?"

Let $n = \deg(R)$. We start by recalling that R is split, iff R has a left ideal L of dimension n (viewed as a vector space over F); cf. Remark 24.20′. But a left ideal is just a subspace closed under left multiplication by each element of R. In other words, defining $\ell_r \colon R \to R$ as the left multiplication map, we see that R is split iff R has a subspace L of dimension n that is invariant under all left multiplication maps $\ell_r : r \in R$.

Thus, we consider R as an n^2-dimensional vector space over F, which embeds into $R \otimes_F \bar{F}$, an n^2-dimensional vector space over $\bar{F}$; we define $\mathbb{SB}_R$ to be the set of n-dimensional subspaces closed under all left multiplication maps of elements of R. We claim that $\mathbb{SB}_R$ is the set of F-points of a projective variety over $\bar{F}$.

Our main task is to find a way of picking out the set of n-dimensional subspaces of $V = F^{(n^2)}$ by means of algebraic equations. This is done by means of the Grassmann algebras $E(V)$; cf. Example 18.40(iv). Grading $E(V)$ in the usual way, we write $E^m(V)$ for the m-component. In particular, $v_1 v_2 \cdots v_m \in E^m(V)$ for any $v_1, \dots, v_m \in V$.

Remark 24A.12. Suppose W is a d-dimensional subspace (of $V = F^{(n^2)}$) having some base $\{b_1, \dots, b_d\}$. Put $\omega_W = b_1 \cdots b_d \in E^d(V)$. We claim that up to scalar multiple, ω_W is independent of the choice of base. Indeed, if $w_1, \dots, w_d \in W$, then writing $w_i = \sum_{j=1}^d \alpha_{ij} b_j$, we have

$$w_1 \cdots w_d = \sum_{j_1, \dots, j_d} \alpha_{1,j_1} \alpha_{2,j_2} \cdots \alpha_{d,j_d} b_{j_1} \cdots b_{j_d}.$$

But $b_{j_1} \cdots b_{j_d} = 0$ unless $j_1, \dots, j_t$ are distinct, in which case $j_i = \pi(i)$ for some $\pi \in S_d$ and $b_{j_1} \cdots b_{j_d} = (\operatorname{sg} \pi) b_1 \cdots b_d$. Hence, $w_1 \cdots w_d = \det(\alpha_{ij}) b_1 \cdots b_d$ is a multiple of ω_W, and is 0 unless the w_i are linearly independent.

In this way, we may identify W with the 1-dimensional subspace $F\omega_W$ of $E^d(V)$ or, in other words, as an F-point in the projective space $\mathbb{P}(E^d(\bar{V}))$ defined over $\bar{F}$. Plücker managed to define this in terms of equations, which thus enables one to display $\mathbb{SB}_R$ as an irreducible subvariety of $\mathbb{P}(E^d(\bar{V}))$; cf. Exercises A1–A7. Exercise A8 shows how to describe those points corresponding to n-dimensional left ideals. (Full details are given in [**Jac7**].)

Thus, in view of Remark 24.20′, $\mathbb{SB}_R$ has a K-rational point iff K is a splitting field for R. (But note in this case that $\mathbb{SB}_R$ has many K-rational points, one for each minimal left ideal of $R \otimes_F K$.)

Once we see that $\mathbb{SB}_R$ is an irreducible projective variety, we can take the field $F(\mathbb{SB}_R)_0$ of grade-0 rational functions. Almost by definition, $F(\mathbb{SB}_R)_0$ is a splitting field of R and is "generic" in a sense that we do not go into here.

This is just the beginning of a fascinating interplay between the Brauer group and algebraic geometry, that arises in most of the current research in division algebras. Much of the theory of a csa R can be encoded into a generic splitting field; this is still an active area of research.

Appendix 24B: Infinite-dimensional division algebras

Throughout this chapter, we have studied only f.d. division algebras. More generally, one could consider infinite-dimensional division algebras. By Goldie's Theorem, any Ore domain has a ring of fractions that is a division ring, thereby giving rise to a wealth of examples, some of which are very important.

Example 24B.1. (i) The Weyl algebras $\mathcal{A}_n(F)$ of Appendix 13A are simple Noetherian domains that are not Artinian, and thus (in view of Kaplansky's PI-theorem) cannot be finite-dimensional over their centers. The division ring of fractions of $\mathcal{A}_n(F)$ is denoted as $\mathcal{D}_n(F)$, and would be expected to play a key role in the theory.

(ii) The enveloping algebra $U(L)$ of a f.d. Lie algebra L is a Noetherian domain, by Remark 21C.4. (This is also true for quantum enveloping algebras; cf. Exercises 21C.17 and 21C.20.) Gel′fand and Kirillov conjectured that the division ring of fractions of $U(L)$ must be isomorphic to the division ring of fractions of a Weyl algebra over a purely transcendental field extension of F. This was proved independently by A. Joseph, J. McConnell, and Borho-Gabriel-Rentschler in the case that L is a solvable Lie algebra, but does not hold in general.

(iii) Ore extensions also give other examples of Noetherian domains, and thus of division rings.

(iv) Any domain not containing a free algebra is Ore, by Proposition 17.12, and thus has a division ring of fractions.

Similar considerations also lead us to study arbitrary prime Goldie rings in terms of their simple rings of fractions and the underlying division rings.

Chapter 25

Homological Algebra and Categories of Modules

One of the main trends of modern algebra has been to move from explicit computations to coordinate-free arguments. The most striking outcome of this trend has been the emergence of category theory, which provides the environment in which much research presently is undertaken. Although Grothendieck invented categories to investigate sheaves, category theory has become a major tool in algebra. In this chapter, we survey some of the highlights of this approach, starting with a brief overview, including a description of two important functors in algebra, Hom and $\otimes$, which lead to the basic notions of **projective**, **injective**, and **flat** modules. The failure of the functors Hom and $\otimes$ to be exact leads us to homology and cohomology, and more generally to "derived functors." In the process, we introduce projective and injective resolutions, and connect them to the free resolutions of Chapter 17. More recently, derived functors have been recast in the context of derived categories.

These concepts enable us to unify and advance much of the material treated earlier in this text, including exact sequences of modules, PLID's, Wedderburn's Principal Theorem, Levi's Theorems for Lie algebras, Hilbert's Theorem 90, and crossed products (central simple algebras). Clearly such powerful theories should be the focus of intensive investigation, and we cannot hope to cover the material adequately in such a small space. We do try to give the flavor of the basic concepts, especially in their relation to

the structure of algebras. For a more thorough treatment, the reader can find excellent accounts in the literature, such as Weibel [**Wei**] or Brown [**Broks**] (for a superb exposition of group cohomology).

In Appendix 25A, we turn to Morita's Theorem, which determines when two rings have "equivalent" categories of modules. This leads us in Appendix 25B to the notions of separable algebras and Azumaya algebras, far-reaching generalizations both of separable field extensions and of central simple algebras. Finally, in Appendix 25C, we tie together several strands by outlining a beautiful theory about representations of graphs, describing the representation type of f.d. algebras in terms of Dynkin diagrams.

The downside of category theory is that a rigorous development requires many verifications. Unlike the prior chapters, this chapter draws on all previous material, including the appendices and even a few easy exercises. In particular, we refer to Appendix 1A of Volume 1 for basic material about categories, such as morphisms (including monics and epics) and functors. In order to progress at a reasonable pace, we leave the purely categorical-theoretic verifications as exercises, and focus here on the algebraic concepts.

Often we focus on the category R-**Mod** of modules over a ring R. Although this has the advantage of familiarity, some confusion could arise because sometimes modules are described more naturally in category theory in terms of morphisms. One example of such ambiguity is the familiar kernel of a module map, which in module theory is viewed as a submodule. Similarly, the kernel of a group homomorphism is a normal subgroup, and the kernel of a ring homomorphism is an ideal. On the other hand, in category theory, the kernel is defined as a morphism. Namely, if K is the module-theoretic kernel of the map $f\colon A \to B$, then the categorical-theoretic kernel of f is the natural monic $k\colon K \to A$ (or, more precisely, its equivalence class; cf. Definition 1A.8 of Volume 1). To avoid confusion with the module-theoretic usage of "kernel," we call this the **categorical kernel**.

Likewise, in module theory, the **cokernel** of a map $f\colon A \to B$, written $\operatorname{coker} f$, is defined as the module $B/f(A)$. But the categorical cokernel is defined dually to the categorical kernel, which in the category R-**Mod** is the canonical epic $B \mapsto B/f(A)$; cf. Definition 1A.11 of Volume 1.

Throughout this chapter, we work over a special kind of category. Recall from Definition 1A.7 that a category $\mathcal{C}$ is called **pre-additive** when the composition $\operatorname{Hom}(B,C) \times \operatorname{Hom}(A,B) \to \operatorname{Hom}(A,C)$ is a bilinear map of Abelian groups, for all objects A, B and C of $\mathcal{C}$. (Weibel [**Wei**] calls this an **Ab**-category.) We assume that $\mathcal{C}$ is pre-additive and has a zero object 0, such that $\operatorname{Hom}(A,0) = \operatorname{Hom}(0,A) = 0$ for all $A \in \mathcal{C}$.

Recall that a **covariant functor** F satisfies $F(fg) = FfFg$ for all morphisms f, g, whereas a **contravariant functor** F satisfies $F(fg) = FgFf$.

We only consider functors F that are **additive**, in the sense that $F(f+g) = Ff + Fg$ for all morphisms f and g.

Definition 25.1. A category $\mathcal{C}$ is **Abelian** if $\mathcal{C}$ is a pre-additive category closed under taking finite (direct) products, such that every morphism f can be written as the composite hg, where h is a categorical kernel and g is a categorical cokernel. (Thus, in an Abelian category, every monic is a categorical kernel, and every epic is a categorical cokernel.)

A **short exact sequence** $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ can be defined in an arbitrary Abelian category, where we stipulate that f is monic, g is epic, and $\operatorname{im} f = \ker \operatorname{coker} f$ is also a categorical kernel of g.

The main example of an Abelian category is R-**Mod**. (Indeed, any map $f\colon A \to B$ factors as hg, where $g\colon A \to A/\ker f$ is the canonical epic and $h\colon A/\ker f \to B$ is the monic obtained from Noether's first isomorphism theorem.) R-**Mod** is in a sense the prototypical example of an Abelian category, since the Freyd-Mitchell Theorem [**Wei, Theorem 1.6.1**] shows that any small Abelian category can be viewed as a full subcategory of R-**Mod** for a suitable ring R. ¿From this point of view, we do not lose much by working directly with R-**Mod**, and our proofs will be formulated for R-**Mod**.

Unfortunately, we thereby lose some of the guidance provided by categorical formalism, as well as other advantages of working directly with Abelian categories. Besides the obvious benefit of having the theory available for other applications, the class of Abelian categories has one major advantage over the category R-**Mod**: The axioms defining Abelian categories are self-dual, in the sense that if $\mathcal{C}$ is an Abelian category, then so is its dual category $\mathcal{C}^{\mathrm{op}}$; cf. Definition 1A.9 and Exercises 1A.18 and 1A.19 of Volume 1. Thus, when working abstractly in Abelian categories, one often gets two theorems for the price of one — both the original theorem and the one obtained by reversing all arrows. Although the category R-**Mod** is not self-dual, we indicate where duality would come into play.

Exact and half-exact functors

If $0 \to K \xrightarrow{f} M \xrightarrow{\pi} N \to 0$ is a split exact sequence of modules, then so is

$$0 \to FK \xrightarrow{Ff} FM \xrightarrow{F\pi} FN \to 0$$

for any functor F. (Indeed, if $g\colon N \to M$ with $\pi g = 1_N$, then $F\pi\, Fg = 1_{FN}$.) However, as we shall see, functors need not preserve arbitrary exact sequences, so we prepare ourselves with a definition.

Definition 25.2. A covariant functor F is **exact** if F preserves all exact sequences; i.e., whenever $A \xrightarrow{f} B \xrightarrow{g} C$ is exact, then so is $FA \xrightarrow{Ff} FB \xrightarrow{Fg} FC$.

The functor F is **left exact** if it preserves the left part of each short exact sequence; i.e., whenever $0 \to A \xrightarrow{f} B \xrightarrow{g} C$ is exact, then so is the sequence $0 \to FA \xrightarrow{Ff} FB \xrightarrow{Fg} FC$.

Dually, the functor F is **right exact** if it preserves the right part of each short exact sequence; i.e., whenever $A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is exact, then so is the sequence $FA \xrightarrow{Ff} FB \xrightarrow{Fg} FC \to 0$.

A contravariant functor $F\colon \mathcal{C} \to \mathcal{D}$ is **left exact** if the corresponding covariant functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$ is left exact; i.e., whenever $A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is exact, then so is $0 \to FC \xrightarrow{Fg} FB \xrightarrow{Ff} FA$.

Left exact and right exact functors are sometimes called **half-exact**.

Remark 25.3. (i) To check that a covariant functor F is left exact, it suffices to show that whenever $0 \to A \xrightarrow{f} B \xrightarrow{g} C$ is exact, then so is $FA \xrightarrow{Ff} FB \xrightarrow{Fg} FC$. Indeed, what remains to show in the definition is that $0 \to FA \xrightarrow{Ff} FB$ is exact, which one gets by applying our criterion to $0 \to 0 \to A \xrightarrow{f} B$.

(ii) Any exact functor is left exact and right exact, by (i).

Here are our basic examples.

Example 25.4 (The Hom functors). We fix an object C of $\mathcal{C}$. (Although formulated here for $\mathcal{C} = R$-**Mod**, this example can be verified without difficulty for any Abelian category $\mathcal{C}$.)

(i) $F = \operatorname{Hom}(C, \_)\colon \mathcal{C} \to$ **Ab** is defined by $FA = \operatorname{Hom}(C, A)$ for each object A of $\mathcal{C}$; for each morphism $f\colon A \to B$, we define

$$Ff = f_{\#}\colon \operatorname{Hom}(C, A) \to \operatorname{Hom}(C, B)$$

by $f_{\#}(h) = fh$. Note that F is a covariant functor since $(gf)_{\#} = g_{\#}f_{\#}$.

We claim that F is left exact. Suppose $0 \to A \xrightarrow{f} A' \xrightarrow{g} A''$ is exact. By Remark 25.3, we need to show that

$$\operatorname{Hom}(C, A) \xrightarrow{Ff} \operatorname{Hom}(C, A') \xrightarrow{Fg} \operatorname{Hom}(C, A'')$$

also is exact. Suppose $h' \in \ker Fg$. Thus, $h'\colon C \to A'$ with $gh' = 0$. This means for any $c \in C$ we have $h'(c) \in \ker g = f(A)$, so $h'(c) = f(a)$ for

some $a \in A$; we define $h\colon C \to A$ by $h(c) = a$. The map h is well-defined since f is monic, and clearly $fh = h'$, as desired.

(ii) $G = \mathrm{Hom}(\_, C)\colon \mathcal{C} \to \mathbf{Ab}$ is defined by $GA = \mathrm{Hom}(A, C)$ for each object A of $\mathcal{C}$; for each morphism $f\colon A \to B$, we define

$$Gf = f^{\#}\colon \mathrm{Hom}(B, C) \to \mathrm{Hom}(A, C)$$

by $f^{\#}(h) = hf$. Now $(gf)^{\#} = f^{\#}g^{\#}$, so G is a contravariant functor.

We claim that g is left exact. Indeed, suppose $A \xrightarrow{f} A' \xrightarrow{g} A'' \to 0$ is exact. We need to show that

$$\mathrm{Hom}(A'', C) \xrightarrow{Gg} \mathrm{Hom}(A', C) \xrightarrow{Gf} \mathrm{Hom}(A, C)$$

also is exact. Suppose $h \in \ker Gf$. Thus, $h\colon A' \to C$ with $h'f = 0$. Define $h''\colon A'' \to C$ by $h(a'') = h'(a)$ where $g(a) = a''$. If also $g(a_1) = a'' = g(a)$, then $a_1 - a \in \ker g = f(A)$, so $h'(a_1) = h'(a)$, implying that h is a well-defined map, and clearly $hg = h'$, as desired.

(This proof foreshadows a technique called **diagram chasing**, to be utilized later in Lemma 25.38.) Although both $\mathrm{Hom}(C, \_)$ and $\mathrm{Hom}(\_, C)$ are left exact, they are not necessarily exact. Indeed, soon we shall obtain precise conditions for these functors to be exact in R-**Mod**.

Remark 25.4′. When $\mathcal{C} = R$-**Mod** for a commutative ring R, then for any modules M and N, $\mathrm{Hom}_R(M, N)$ also has the structure of an R-module by Remark 18.43; thus $\mathrm{Hom}_R(M, \_)$ and $\mathrm{Hom}_R(\_, N)$ are functors from $\mathcal{C}$ to itself.

More generally, consider the categories R-**Mod**-W of R, W-bimodules and T-**Mod**-W of T, W-bimodules for rings R, T, and W. Suppose M is a T, R-bimodule. For any $N \in T$-**Mod**-W, we view $\mathrm{Hom}_T(M, N)$ as an R, W-bimodule, where for $f \in \mathrm{Hom}_T(M, N)$, multiplication on the left by $r \in R$ is defined by $(rf)(a) = f(ar)$, and likewise multiplication on the right by $w \in W$ is defined by $(fw)(a) = f(a)w$. $\mathrm{Hom}_T(M, \_)$ is easily seen to be a left exact covariant functor from T-**Mod**-W to R-**Mod**-W.

Example 25.5 (The tensor functor). As in Example 18.9, given a T, R-bimodule M, we have the covariant functor

$$F = M \otimes_R \_ : R\text{-}\mathbf{Mod}\text{-}W \to T\text{-}\mathbf{Mod}\text{-}W$$

given by $FN = M \otimes_R N$ and $Ff = 1_M \otimes f$ for each morphism $f\colon N_1 \to N_2$. This functor is right exact, in view of Exercise 18.3. Likewise, $\_ \otimes_T M$ is a right exact, covariant functor from W-**Mod**-T to W-**Mod**-R.

Perhaps the tensor functors are most easy to conceptualize when $T = R$ is a commutative ring, since then $M \otimes_R \_$ sends R-**Mod** to itself, and we can readily study the properties of this functor. Even so, in contrast to the Hom functor, the tensor construction does not arise in arbitrary Abelian categories, so to carry out the theory categorically, one needs extra categorical structure, which we leave for Chapter 26.

The functors $\mathrm{Hom}_R(M, \_)$ and $M \otimes_R \_$ are tied together in an interesting way.

Definition 25.6. An **adjoint pair** (F, G) is a pair of functors $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$ such that, for any objects C of $\mathcal{C}$ and D of $\mathcal{D}$, there is a bijection $\eta_{C,D}\colon \mathrm{Hom}_{\mathcal{D}}(FC, D) \to \mathrm{Hom}_{\mathcal{C}}(C, GD)$ that is natural in each component (i.e., $\eta_{C,\_}$ and $\eta_{\_,D}$ are natural isomorphisms for any objects C and D, in the sense of Definition 1A.16 of Volume 1).

In this setup, the functor F is also called a **left adjoint** of G, and G is called a **right adjoint** of F.

The property of being an adjoint pair has significant consequences, as illustrated in Exercise 45. Here is the motivating example:

Example 25.6′. In the notation of Proposition 18.44, $(A \otimes_S \_,\ \mathrm{Hom}_R(A, \_))$ is an adjoint pair.

Here is another important example. Recall that $\mathrm{Unit}(R)$ denotes the multiplicative group of units of a ring R.

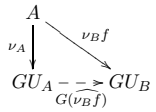
Example 25.7. (i) Lemma 19.15 says for any group G that

$$\mathrm{Hom}(C[G], R) \cong \mathrm{Hom}(G, \mathrm{Unit}(R)), \tag{25.1}$$

where the left side is in the category of C-algebras and the right side is in the category of groups. Thus, the group algebra construction is adjoint to the functor taking an algebra to its group of units. We shall have occasion to use (25.1) for $C = \mathbb{Z}$.

(ii) More generally adjoint pairs can be obtained from the construction of "universals" that has pervaded this volume: Suppose $G\colon \mathcal{C} \to \mathcal{D}$ is some functor such that, for each object D of $\mathcal{D}$, there is a corresponding universal (U_D, ν_D) to G; cf. Definition 8.5 of Volume 1.

Define a functor $F\colon \mathcal{D} \to \mathcal{C}$ by putting $FA = U_A$ for any object A of $\mathcal{D}$, where for any morphism $f\colon A \to B$ of $\mathcal{D}$, Ff is taken to be the (unique) morphism such that $G(Ff)\nu_A = \nu_B f$. In other words, $Ff = \widehat{\nu_B f}$, according to the following commutative diagram from Definition 8.5 of Volume 1:



It is easy to verify that (F, G) is an adjoint pair; the correspondence from $\mathrm{Hom}_{\mathcal{D}}(FC, D) \to \mathrm{Hom}_{\mathcal{C}}(C, GD)$ is given by sending $h\colon U_C \to D$ to the composite $(Gh)\nu_C : C \to GU_C \to GD$, whereas the reverse direction follows from the definition of universal.

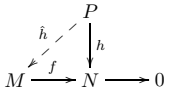
Conversely, one can also define universals by means of adjoint pairs; cf. Exercise 43. In particular, the construction of the tensor algebra can be viewed as the left adjoint of the forgetful functor from algebras to modules.

Projective modules

Perhaps the most appropriate question at this stage is, "For which objects C is the left exact functor $\mathrm{Hom}(C, \_)$ exact?" The answer leads to one of the most important classes of objects in category theory.

Definition 25.8. An object P of an Abelian category is **projective** if, for every epic $M \xrightarrow{f} N$, each morphism $h\colon P \to N$ can be **lifted** to a (not necessarily unique) morphism $\hat{h}\colon P \to M$ such that $h = f\hat{h}$; in other words, the following diagram is commutative:

(25.2)



Thus, in view of Example 25.4, an object P is projective iff the functor $\mathrm{Hom}(P, \_)$ is exact. So far we have no indication that projective objects exist! But we get a very satisfactory answer for the category R-**Mod**, using ideas from Chapter 2 of Volume 1.

Remark 25.9. Every free module P is projective, since we could define the lifting map $\hat{h}$ by means of a base $\{b_i : i \in I\}$ of P. Explicitly, given $h\colon P \to N$, pick $c_i \in M$ in the preimage (with respect to f) of $h(b_i)$ and define $\hat{h}\colon P \to M$ by $h(b_i) = c_i$, $\forall i \in I$.

Some easy examples of nonprojective modules are given in Exercises 1, 2, and 4.

PROPOSITION 25.10. *The following assertions are equivalent for an R-module P:*

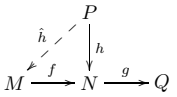
(1) *P is projective.*

(2) *P is a direct summand of a free module.*

(3) *There is a split epic $f\colon F \to P$ with F free. (Moreover, if P is generated by n elements, then we can take $F = R^{(n)}$.)*

(4) *Every epic $f\colon M \to P$ splits.*

(5) *If the sequence of modules $M \xrightarrow{f} N \xrightarrow{g} Q$ is exact and $h\colon P \to N$ with $gh = 0$, then h lifts to a map $\hat{h}\colon P \to M$ with $h = f\hat{h}$; in other words, the following diagram is commutative:*



Proof. (1) $\Rightarrow$ (5) Replace N by $f(M)$, noting that $h(P) \subseteq \ker g = f(M)$; then we obtain $\hat{h}$ from diagram (25.2).

(5) $\Rightarrow$ (1) Obvious; take $g = 0$.

(1) $\Rightarrow$ (4) Take $N = P$ and $h = 1_N$ in (25.2).

(4) $\Rightarrow$ (3) Take a free module F having a base of cardinality at least the number of generators as P; then there is an epic $F \to P$, which splits by hypothesis.

(3) $\Rightarrow$ (2) By Proposition 2.8 of Volume 1.

(2) $\Rightarrow$ (1) By Remark 25.9 and the following proposition. □

PROPOSITION 25.11. *A direct sum $P = \oplus P_i$ of modules is projective iff each of the P_i is projective.*

Proof. Take the projections $\pi_i\colon P \to P_i$ and the injections $\nu_i\colon P_i \to P$, satisfying $\pi_i\nu_i = 1_{P_i}$ and $\sum \nu_i\pi_i = 1_P$. Suppose $f\colon M \to N$ is epic.

($\Rightarrow$) Given $h_i\colon P_i \to N$, then $\bigoplus h_i\pi_i\colon P \to N$ lifts to a map $\hat{h}\colon P \to M$, and $\hat{h}\nu_i\colon P_i \to M$ is the desired map lifting $h_i\pi_i\nu_i = h_i$.

($\Leftarrow$) Given $h\colon P \to N$, we lift each $h\nu_i$ to $\hat{h}_i\colon P_i \to M$; then we have $\hat{h}\colon P \to M$ with $\hat{h}_i = \hat{h}\nu_i$ for each i. Thus, $h\nu_i = f\hat{h}\nu_i$ for each i, proving that $h = f\hat{h}$. □

Remark 25.12. Projective modules are fundamental in many aspects of mathematics. Condition (2) of Proposition 25.10 is the link to free modules, and provides the most common way of verifying projectivity in elementary

module theory. On the other hand, the definition of projectivity (as well as condition (4)) is categorical, so the language of category theory provides an impetus that is lacking for free modules.

Furthermore, condition (5) of Proposition 25.10 enables projective objects to play a fundamental role in homology theory, to be seen later in this chapter.

Remark 25.12′. By Theorem 15.13 and Proposition 25.10(4), a ring R is semisimple iff every short exact sequence of R-modules splits, iff every R-module is projective.

One key computational version of projectivity is the next result.

PROPOSITION 25.13 (DUAL BASIS LEMMA). *Suppose $P = \sum_{i\in I} Ra_i$ is a projective R-module. Then there are R-module maps $h_i\colon P \to R$ satisfying*

$$a = \sum_{i\in I} h_i(a)a_i \quad \forall a \in P, \tag{25.3}$$

where, for each a, $h_i(a) = 0$ for almost all i. Conversely, the existence of such h_i implies that P is a projective module.

Proof. For each direction, we take a free module $F = R^{(I)}$ with base $\{e_i : i \in I\}$ and the epic $f\colon F \to P$ given by $f(e_i) = a_i$, $\forall i \in I$. Also we define the canonical projections $\pi_j\colon F \to R$ by $\pi_j(e_i) = \delta_{ij}$. Any $c \in F$ satisfies $c = \sum \pi_i(c)e_i$.

($\Rightarrow$) In view of Proposition 25.10, f is split, so we take $g\colon P \to F$ with $fg = 1_P$. Put $h_i = \pi_i g$. Then any $a \in P$ satisfies

$$\begin{aligned} a = fg(a) &= f\Big(\sum_i \pi_i(g(a))e_i\Big) \\ &= \sum_i f(h_i(a)e_i) = \sum_i h_i(a)f(e_i) = \sum_i h_i(a)a_i, \end{aligned}$$

as desired.

($\Leftarrow$) Defining $g\colon P \to F$ by $g(a) = \sum h_i(a)e_i$, we have

$$fg(a) = \sum h_i(a)f(e_i) = \sum h_i(a)a_i = a.$$

Thus $fg = 1_p$, so P is projective, by Proposition 25.10(3). □

Examples of projective modules. Let us turn to specific examples of projective modules, referring the reader to Exercises 1, 2, and 4 for some well-known nonprojective modules.

Example 25.14. (i) For any idempotent e of a ring R, the module Re is projective, since $R = Re \oplus R(1-e)$. In particular, if $R = M_n(F)$, then the minimal left ideal $L = Re_{11}$ is projective but not free, since $\dim_F L = n < n^2 = \dim_F R$. (This was the cause of our difficulty in Example 17.22.)

(ii) If R is any commutative ring and $e \in R$ is a nontrivial idempotent, then the projective module Re is not free, since $0 \neq 1-e \in \operatorname{Ann}_R Re$. For example, $\mathbb{Z}/2$ is projective as a $\mathbb{Z}/6$-module, where $\mathbb{Z}/2$ is identified with $3\mathbb{Z}/6$, and $[3]^2 = [9] = [3]$ is idempotent. (Alternatively, one could use the Chinese Remainder Theorem.)

Before trying to extend results from free modules to projective modules, one might ask whether a projective module over a given ring R is necessarily free. The examples of non-free projective modules in Example 25.14 are "obvious" since they arise from a nontrivial idempotent of the ring R. What if R has no nontrivial idempotents — for example, if R is an integral domain?

Example 25.15. If R is a local ring with maximal ideal J, then every f.g. projective R-module P is free. Indeed, take elements $a_1, \dots, a_n$ generating P, with n minimal possible, and let $F = R^{(n)}$. By Proposition 25.10, we have a split epic $\pi\colon F \to P$ given by $\pi(e_i) = a_i$ for $1 \le i \le n$, by means of which we identify P as a direct summand of F.

First note that $\ker\pi \le JF$. (For otherwise say $(r_1, \dots, r_n) \in \ker\pi$ with some $r_m \notin J$. Then r_m is invertible but $\sum r_i a_i = 0$, so $a_m = -r_m^{-1}\sum_{i\neq m} r_i a_i$ is spanned by the other a_i, contrary to minimality of n.)

Thus $F = P \oplus \ker\pi \le P + JF$, so $F = P$ (cf. the version of Nakayama's Lemma given in Remark 8.25(i) of Volume 1).

(Kaplansky showed that the f.g. hypothesis is superfluous, by means of an ingenious counting argument.) This example becomes significant in conjunction with localization, especially since local rings arise often in algebraic geometry. Here is another easy application of the commutative theory.

Example 25.16. Suppose R is a PID (principal ideal domain). Any projective module, being a summand of a free module, is torsion-free; hence, by Theorem 2.54 of Volume 1, any f.g. projective module is a direct sum of copies of R and thus is free.

In Exercise 13.23, we obtained the same conclusion more generally for f.g. modules over a PLID (principal left ideal domain). Hence, any f.g. projective module over a PLID is free. One can also remove the hypotheses "f.g.," using Exercise 6.

Nevertheless, there are examples of nonfree projectives over Noetherian integral domains, and in fact the arithmetic theory of Dedekind domains can be interpreted in terms of their projective modules. The reason is quite easy:

Example 25.17. Suppose C is an integral domain.

(i) Any invertible ideal P of C (in the sense of Definition 12A.2 of Volume 1) is f.g. projective as a module. (Indeed, if Q is a fractional ideal such that $QP = C$, then we write $1 = \sum q_i a_i$, for $a_i \in P$, $q_i \in Q$, and define $f_i\colon P \to C$ by $f_i(a) = q_i a$ for all $a \in P$. By definition,

$$\sum f_i(a)a_i = \sum q_i a a_i = \big(\sum q_i a_i\big)a = a;$$

hence, P is f.g. projective by the Dual Basis Lemma.) The converse also holds, by Exercise 7.

(ii) If an ideal P of an integral domain is free as a module, then P is principal, since any two elements b_1, b_2 are linearly dependent, satisfying $b_1 b_2 - b_2 b_1 = 0$.

(iii) By (i), every ideal of a Dedekind domain is f.g. projective, but any free ideal is principal, by (ii). It follows that any Dedekind domain of class number > 1 must have ideals that are nonfree f.g. projective.

Related examples are given in Exercises 10 and 11. These considerations elevate projective modules to a position of prominence in the theory of Dedekind domains, and also can be generalized to modules over arbitrary commutative rings.

Projective resolutions of modules.

We introduced free resolution of modules in Definition 17.20ff, partly as a tool to see how "far" a module is from being free, but quickly noted their limitations in Example 17.22. The proper setting is in the projective theory.

Definition 25.18. A **resolution** $\mathbb{P}$ of a module M is a chain

$$\cdots \to P_n \xrightarrow{f_n} P_{n-1} \to \cdots \to P_1 \xrightarrow{f_1} P_0$$

that is part of an exact sequence

$$\cdots \to P_n \xrightarrow{f_n} P_{n-1} \to \cdots \to P_1 \xrightarrow{f_1} P_0 \xrightarrow{\varepsilon} M \to 0. \tag{25.4}$$

The map $\varepsilon\colon P_0 \to M$ is called the **augmentation map**; $K_n = \ker f_n$ is called the n-th **syzygy**. $\mathbb{P}$ has **length** n if $P_j = 0$ for all $j > n$. The resolution $\mathbb{P}$ of M is **projective** if each module P_i is projective.

We say that the projective resolution $\mathbb{P}$ is f.g. if each P_i is also f.g. By Example 17.21′, any module M has a free, and thus projective, resolution. Likewise, any f.g. module over a Noetherian ring R has a f.g. projective resolution.

We define the **projective dimension** $\operatorname{pd}(M) = n$, whenever M has a projective resolution of length n (for n minimal). When M does not have any projective resolution of finite length, we say $\operatorname{pd}(M) = \infty$.

A module M is projective iff $\operatorname{pd}(M) = 0$ (since $0 \to \ldots \to 0 \to M$ is a projective resolution of M); we want to study arbitrary modules by means of pd as an inductive tool. Syzygies play a special role: If, in Definition 25.18, $K_{n-1} = \ker f_{n-1}$ is projective, we could take $P_n = K_{n-1}$ and then $P_j = 0$ for all $j > n$, so $\operatorname{pd}(M) \le n$. In fact, for any $n \ge 1$, it follows at once from Exercise 21 (or the results below about Ext) that $\operatorname{pd}(M) \le n$ iff the $(n-1)$-syzygy of some (and thus every) projective resolution is projective. This may be the best way of viewing pd.

Remark 25.19. The exact sequence (25.4) can be cut at any syzygy. For example, assuming $\operatorname{pd} M \ge 1$, cutting at $K_0 = \ker \varepsilon$ gives us the exact sequence

$$\cdots \to P_n \xrightarrow{f_n} P_{n-1} \to \cdots \xrightarrow{f_2} P_1 \xrightarrow{f_1} K_0 \to 0,$$

so $\cdots \to P_n \xrightarrow{f_n} P_{n-1} \to \cdots \xrightarrow{f_2} P_1$ is a projective resolution of K_0, and $\operatorname{pd}(K_0) = \operatorname{pd}(M) - 1$. In this manner, we apply induction to the syzygies.

Since Remark 25.19 is only useful for modules of finite pd, we are led to the next definition.

Definition 25.20. The (left) **global dimension** of the ring R, denoted $\operatorname{gl\,dim} R$, is $\sup\{\operatorname{pd} M : M \in R\text{-}\mathbf{Mod}\}$.

Example 25.21. (i) By Remark 25.12′, R is semisimple iff $\operatorname{gl\,dim} R = 0$.

(ii) If R is a PLID, then $\operatorname{gl\,dim} R \le 1$ because, by Exercise 6, the 0-syzygy of the projective resolution of any module is projective (in fact, free).

(iii) If $\operatorname{gl\,dim} R = m$, then $\operatorname{gl\,dim} R[\lambda] = m + 1$; cf. Exercise 24. Thus, $\operatorname{gl\,dim} F[\lambda_1, \ldots, \lambda_n] = n$.

From the point of view of f.g. projective modules, the class of rings of greatest interest is the class of Noetherian rings of finite global dimension. Here is an instance of their role.

Digression 25.22 (Serre's conjecture and K_0-theory). By Example 25.16, any projective module over the polynomial ring $F[\lambda]$ is free. Serre conjectured more generally that for all n, every f.g. projective module over the polynomial ring $F[\lambda_1, \dots, \lambda_n]$ is free. This query was motivated by the analogy between vector bundles in topology, which become trivial after (geometric) localization, and projective modules, which by Example 25.15 become free after algebraic localization. Serre's conjecture, a natural analogy of the fact that vector bundles over affine space are trivial, turned out to be a major challenge and was one of the spurs to the initial development of algebraic K-theory; finally, it was solved independently by Quillen [**Q**] and Suslin [**Su**]. (In fact, by a trick of Bass [**Ba1**], which says that projectives with "enough" generators are free, one can conclude from the Quillen-Suslin Theorem that every projective module over $F[\lambda_1, \dots, \lambda_n]$ is free.) We pause a moment to lay down some of the groundwork.

There is an Abelian monoid $\mathcal{M}$ consisting of the isomorphism classes $\bar{P}$ of f.g. projective R-modules whose monoid operation is $\oplus$.

We want to embed $\mathcal{M}$ into a group $\mathcal{F}$. There is a standard construction of embedding a cancellative monoid into a group, motivated by the construction of $\mathbb{Z}$ from $\mathbb{N}$. However, here we have the extra wrinkle that the monoid $\mathcal{M}$ need not be cancellative. For example, one might have $P \oplus R^{(n)} \cong R^{(m)}$ for suitable m, n, in which case we say P is **stably free**. (The first example of a non-free, stably free projective module also came from topology — the tangent bundle over a sphere.)

Thus, in parallel to Definition 8.1 of Volume 1, we define an equivalence relation on ordered pairs (P, Q) of f.g. projective modules, saying that $(P_1, Q_1) \sim (P_2, Q_2)$ iff

$$P_1 \oplus Q_2 \oplus P' \cong P_2 \oplus Q_1 \oplus P'$$

for a suitable f.g. projective module P'; taking Q' such that $P' \oplus Q' = F$ is f.g. free, we have $P_1 \oplus Q_2 \oplus F \cong P_2 \oplus Q_1 \oplus F$, so we could replace P' by the f.g. free module F.

We can addition on equivalence classes by

$$[(P_1, Q_1)] + [(P_2, Q_2)] = [(P_1 \oplus P_2, Q_1 \oplus Q_2)];$$

this is easily seen to be well-defined, and

$$[(P, Q)] + [(Q, P)] = [(P \oplus Q, P \oplus Q)] = [(0, 0)].$$

Hence, the set of equivalence classes $[(P, Q)]$ is indeed an Abelian group, which we call $K_0(R)$. Writing $[P]$ for $[(P, 0)]$, we have a natural monoid homomorphism $\mathcal{M} \to K_0(R)$, given by $\overline{P} \mapsto [P]$. Intuitively, $K_0(R)$ is generated as a group by the isomorphism classes of f.g. projective modules, but satisfies all extra relations of the form $[P] + [Q] = [P \oplus Q]$.

$\langle [R] \rangle$ denotes the subgroup of $K_0(R)$ generated by $[R]$. The image of a f.g. projective module P lies in $\langle [R] \rangle$ iff P is stably free. Thus, $\langle [R] \rangle = K_0(R)$ iff every f.g. projective module is stably free; in this case we say $K_0(R)$ is **trivial**.

Serre proved that if R is left Noetherian with $\operatorname{gl\,dim} R < \infty$ and $K_0(R)$ trivial, then the same holds for $R[\lambda]$. In particular, by induction, every projective module of the polynomial ring $R[\lambda_1, \dots, \lambda_n]$ over a PLID R is stably free.

In view of Serre's Theorem, Serre's conjecture boils down to showing, when F is a field, that every f.g. stably free module over $F[\lambda_1, \dots, \lambda_n]$ is free; this was proved by Quillen [**Q**] and Suslin [**Su**] using surprisingly elementary arguments. By the way, this assertion fails when F is a noncommutative division ring, despite Serre's Theorem.

Gubeladze proved the striking theorem for an arbitrary field F and a torsion-free, cancellative commutative monoid S: Every projective $F[S]$-module is free, iff S has the property that for any element a in the group of fractions of S, if $a^2 \in S$ and $a^3 \in S$, then $a \in S$. This property obviously holds for the monoid of monomials in several commuting indeterminates; hence, the Quillen-Suslin Theorem is a special case. Gubeladze's theorem has an algebraic proof given in Swan [**Swa**], along with similar results for other related rings.

The rank of a projective module over a commutative ring.

The theory is strengthened when we take a commutative base ring C.

Remark 25.23.

(i) The tensor product $P_1 \otimes_C P_2$ of projective C-modules is projective, for if $P_i \oplus Q_i \cong F_i$ is free for $i = 1, 2$, then

$$P_1 \otimes P_2 \oplus (P_1 \otimes Q_2 \oplus Q_1 \otimes P_2 \oplus Q_1 \otimes Q_2) \cong (P_1 \oplus P_2) \otimes (Q_1 \oplus Q_2) \cong F_1 \otimes F_2$$

is free, by Corollary 18.13.

(ii) If P is a projective C-module, then for any C-algebra R the extended module $R \otimes_C P$ (cf. Example 18.9) is projective over R; indeed if $P \oplus Q \cong F$ is free as a C-module, then $R \otimes P \oplus R \otimes Q \cong R \otimes F$ is free as an R-module (having the same base over R as F has over C). Thus, in the language of Digression 25.22, $[P] \mapsto [P \otimes_C R]$ yields a map $K_0(C) \to K_0(R)$.

Here is a partial converse to Remark 25.23(i).

PROPOSITION 25.24. *If P, Q are modules over a commutative ring C such that $P \otimes Q \cong C^{(n)}$, then P is projective.*

Proof. We take a base $\{\sum_{j=1}^{t_i} a_{ij} \otimes q_{ij} : 1 \le i \le n\}$ of $P \otimes Q$ over C. Letting $m = \sum_{i=1}^{n} t_i$, we take a standard base $\{e_{ij} : 1 \le i \le n,\ 1 \le j \le t_i\}$ of $C^{(m)}$, and we can define a map $h\colon C^{(m)} \to P$ sending $e_{ij} \to a_{ij}$. The map $h \otimes_C 1_Q : C^{(m)} \otimes Q \to P \otimes Q \cong C^{(n)}$ is epic and thus splits, by Proposition 25.10(4). Hence, $h \otimes_C 1_Q \otimes_C 1_P$ splits and can be identified with $(h, \dots, h)$ taken n times, since $Q \otimes P \cong P \otimes Q \cong C^{(n)}$. Hence each component splits; i.e., h is a split epic, so P is projective. □

(In the other direction, if P is faithful f.g. projective, one can always find a projective module Q such that $P \otimes Q \cong C^{(n)}$ for suitably large n, but the proof relies on [**Ba1**], which is beyond the scope of this text.)

Suppose P is a projective C-module. Remark 25.23(ii) shows that for any localization $C_{\mathfrak{p}}$ of C at a prime ideal $\mathfrak{p}$, the module $P_{\mathfrak{p}} = C_{\mathfrak{p}} \otimes_C P$ is projective and thus free, since any projective module over a local ring is free. (We now use $\mathfrak{p}$ to denote a prime ideal of C, since the letter P is reserved here for a projective module.)

Definition 25.25. (i) Notation as above for any projective module P, we define $\operatorname{rank}_{\mathfrak{p}} P$ to be the rank of the free module $P_{\mathfrak{p}}$ over the local ring $C_{\mathfrak{p}}$. Fixing P yields a function $\operatorname{Spec} C \to \mathbb{N}$; we write $\operatorname{rank}(P)$ for $\{\operatorname{rank}_{\mathfrak{p}} P : \mathfrak{p} \in \operatorname{Spec} C\}$. We say P has **(constant) rank** n if $\operatorname{rank}_{\mathfrak{p}} P = n$ for every $\mathfrak{p} \in \operatorname{Spec} C$.

(ii) (Compare with Example 25.17(i).) A module P over a commutative ring C is **invertible** if it is faithful projective of rank 1.

Remark 25.26. If P is a f.g. projective module, then the **dual module** $P^* = \operatorname{Hom}_R(P, R)$ is a f.g. projective right R-module, according to the module multiplication of Remark 25.4′, and for any right module M there is an isomorphism of Abelian groups

$$P^* \otimes_R M \to \operatorname{Hom}_R(P, M),$$

via the balanced map sending (f, a) to the map $p \mapsto f(p)a$. (The verifications are easy for $P = R$, but also pass easily to finite direct sums and summands, so we appeal to Proposition 25.10(ii).)

By Proposition 25.24, any invertible module is f.g. projective and is seen to be of rank 1; cf. Exercise 17. The **Picard group** is defined as the set of isomorphism classes of invertible modules, with multiplication induced by the tensor product. The identity element is $[C]$; the inverse is given by $[P]^{-1} = [P^*]$ by Exercise 17.

Remark 25.27. A f.g. projective C-module P is faithful iff $\operatorname{rank}_{\mathfrak{p}} P > 0$ for each $\mathfrak{p} \in \operatorname{Spec} C$. (Indeed, if $cP = 0$ for $c \neq 0$, then, by Proposition 8.22 of Volume 1, there is $\mathfrak{p} \in \operatorname{Spec} C$ for which the canonical image $\frac{c}{1}$ of c in $C_{\mathfrak{p}}$ is not 0; it follows that $P_{\mathfrak{p}} = 0$. Conversely, if $P_{\mathfrak{p}} = 0$, then writing $P = \sum_{i=1}^{t} Ca_i$ and taking $s_i \in C \setminus \mathfrak{p}$ with $s_i a_i = 0$, we see that $s_1 \cdots s_t a_i = 0$ for $1 \le i \le t$, implying that $s_1 \cdots s_t \in \operatorname{Ann} P$.)

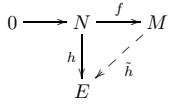
More properties of the rank are found in Exercises 12–18.

Injective modules

Since Definition 25.8 is a paradigm of a categorical definition, we are led to the categorical dual to projective, called *injective*, which also turns out to be a very important concept in module theory and sheaf theory.

Definition 25.28. An object E of an Abelian category is **injective** if, for every monic $f\colon N \to M$ and every morphism $h\colon N \to E$, there is a morphism $\tilde{h}\colon M \to E$ such that $h = \tilde{h} f$, i.e., the following diagram is commutative:

(25.5)



The dual assertions to Proposition 25.10 would be expected to hold, namely:

1. The direct product of injective modules is injective.
2. E is injective iff every monic $E \to M$ splits for each module M.
3. Each module M can be injected into a suitable injective module E.

Unfortunately, having left the general theory of Abelian varieties to facilitate our study of projective modules, we cannot automatically obtain the injective theory by dualizing. In fact, (3) is false in general Abelian categories! Although these assertions are true for injective modules, their proofs (given in Exercises 26–38) are more difficult than their analogs for projective modules, since free modules no longer are available; one introduces a new notion, "divisible," to facilitate our results. On the other hand, the extra effort is worthwhile since we get the stronger property that any module M is a large submodule of a suitable injective module $E(M)$, called the **injective hull** of M; cf. Exercise 39. (A parallel situation holds in algebraic geometry for sheaves.) This can be stated categorically, but its categorical dual is false for projective modules (cf. Exercise 41)!

Injective hulls play a crucial role in general localization theory; for example, the ring of quotients $Q(R)$ of Exercise 16.23 turns out to be $\mathrm{Hom}_R(E(R), E(R))$, and Gabriel carried this out much more generally, in a categorical setting.

Injective resolutions of modules.

We are ready for the dual concept to "projective resolution."

Definition 25.29. An **injective resolution** $\mathbb{E}$ of a module M is a chain

$$E_0 \to E_1 \to \cdots \to E_n \to \cdots,$$

with each E_i injective, together with a map $f\colon M \to E_0$ such that the sequence

$$0 \to M \xrightarrow{f} E_0 \to E_1 \to \cdots \to E_n \to \cdots$$

is exact.

Remark 25.30. By the dual argument to Example 17.21′, every module M has an injective resolution. Namely, take E_0 to be an injective hull of M, E_1 to be an injective hull of E_0/M, and, inductively, E_n to be an injective hull of E_{n-1}/E_{n-2}, for each n.

We define the **injective dimension** of a module to be the minimal length of an injective resolution; a module has injective dimension 0 iff it is injective. But every R-module is injective iff R is semisimple, which is the case iff $\operatorname{gl\,dim} R = 0$. More generally, $\operatorname{gl\,dim} R$ also is the supremum of the injective dimensions of the R-modules; cf. Exercise 61.

Homology and cohomology

Sometimes the failure of a property opens the way to a far richer mathematical theory than if the property had been found to hold. This is the case with exactness. The attempt to measure how far certain chains of morphisms fall from exactness leads to the powerful theory of homology and cohomology whose exposition occupies the remainder of this chapter.

Chain complexes.

We start by introducing the category that supports the theory. We continue to assume throughout that $\mathcal{C}$ is R-**Mod**, although the theory can be formulated for an arbitrary Abelian category.

Definition 25.31. A **chain complex** $(\mathbf{A}, d)$ is a collection of modules $\{A_n\colon n \in \mathbb{Z}\}$, together with maps $d_{n+1}\colon A_{n+1} \to A_n$, $n \in \mathbb{Z}$, called **boundary maps**, which satisfy $d_n d_{n+1} = 0$ for all n.

Although the definition is stated for infinite chains, any finite chain satisfying $d_n d_{n+1} = 0$ for all n can be extended to an infinite chain complex by tacking on the zero morphisms at each end. To unify notation, we define the (infinite) chain complex $(\mathbf{A}, d)$ to be **bounded** if almost all $A_n = 0$.

It is convenient to think of any object A itself as a chain complex **concentrated in dimension 0**, taking $A_0 = A$ and all other $A_n = 0$.

Remark 25.32. We get a new category $\mathbf{Ch} = \mathbf{Ch}(\mathcal{C})$, whose objects are chain complexes over $\mathcal{C}$ and whose morphisms $f\colon (\mathbf{A}, d) \to (\mathbf{A}', d')$, called **chain maps**, are $\mathbb{Z}$-tuples of morphisms $f_n\colon A_n \to A'_n$ for which the following diagram is commutative for all n:

$$\begin{array}{ccc} A_{n+1} & \xrightarrow{d_{n+1}} & A_n \\ {\scriptstyle f_{n+1}}\downarrow & & \downarrow{\scriptstyle f_n}; \\ A'_{n+1} & \xrightarrow{d'_{n+1}} & A'_n \end{array} \tag{25.6}$$

thus, $f_n d_{n+1} = d'_{n+1} f_{n+1}$ for each n. We write $\mathbf{Ch}_+$ for the subcategory of chains $\mathbf{A}$ for which $A_n = 0$ for all $n < 0$, and we write $\mathbf{Ch}_-$ for the subcategory of chains $\mathbf{A}$ for which $A_n = 0$ for all $n > 0$.

Remark 25.33. Given a chain complex $(\mathbf{A}, d)$ of modules, it is convenient to consider $\mathbf{A}$ as an infinite direct sum, graded by $\mathbb{Z}$, and to view all of the d_n together as a graded map $d\colon \mathbf{A} \to \mathbf{A}$ of degree -1 satisfying $d^2 = 0$. Then (25.6) becomes $fd = d'f$. From this point of view, a chain map is just a graded map of degree 0; moreover, we can take direct sums or products of chain complexes, and also form quotient complexes of two chain complexes, all defined by means of matching components. The category **Ch** is seen at once to be an Abelian category with zero object $\mathbf{0}$, where $\mathbf{0}_n = 0$ and $d_n = 0$ for each n; all verifications are done componentwise.

¿From the opposite point of view, any graded direct sum $\mathbf{A} = \oplus_{n\in\mathbb{Z}} A_n$ can be viewed as a chain complex with **trivial boundary map**, where each $d_n\colon A_{n+1} \to A_n$ is taken to be the 0 morphism. Thus, we see that a chain complex has two basic ingredients: a chain complex with trivial boundary (i.e., a graded direct sum) and the original boundary map. It is convenient to consider these two ingredients separately.

As indicated by the terminology, chain complexes originated in topology.

Example 25.34. Any directed graph $\Gamma = (E, V)$ (notation as in Definition 17.23) can be viewed as a bounded complex, where A_0 is the free module with base V, A_1 is the free module with base E, $A_n = 0$ for all other n, and $d_1 \colon A_1 \to A_0$ is given by $d_1(v, w) = v - w$, for any edge (v, w).

Example 25.34′. More generally, given a simplicial complex $\mathcal{K}$, let A_n be the free module whose base is the set of simplices of dimension n. Notation as in Definition 17A.3, given a simplex $\mathcal{S} = \mathcal{P}(S)$, where

$$S = \{v_0, \dots, v_n\},$$

we define the **face map** $\partial_i \colon S \to S_i$, where $S_i = \{v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_n\}$ is obtained by deleting v_i. We define $d_n \colon \mathcal{S} \to A_{n-1}$ to be $\sum_{i=0}^{n} (-1)^i \partial_i$, which extends naturally to a map $d_n \colon A_n \to A_{n-1}$. To see that $d^2 = 0$, it suffices to note that in the computation of $d_{n-1} d_n$, the terms of $\partial_i \partial_j$ and $\partial_j \partial_i$ are paired for $i < j$. However, the corresponding coefficients are, respectively, $(-1)^i (-1)^j$ and $(-1)^{j-1} (-1)^i$, since v_j is the j–1 vertex in S_i, so their sum is

$$(-1)^{i+j} + (-1)^{i+j-1} = 0.$$

Remark 25.35. We can reverse the direction of the arrows in a chain complex $\mathbf{A}$ by writing A^n instead of A_{-n} throughout, and δ^n instead of d_n; this gives us graded maps $\delta \colon \mathbf{A} \to \mathbf{A}$ of degree $+1$, called **coboundary maps**, such that $\delta^n \delta^{n-1} = 0$ for all n. This setup is called a **cochain complex**, since it is a complex for the dual category. Thus, the theories of chain complexes and cochain complexes are the same, except for a change of notation. We define $\mathbf{Ch}^+$ to be the dual category of $\mathbf{Ch}_-$; the objects are the cochain complexes with $A^n = 0$ for each $n < 0$.

The functors B, Z, and H.

Definition 25.36. Given a chain complex $(\mathbf{A}, d)$, define the **boundary** $B_n(\mathbf{A}) = \operatorname{im} d_{n+1}$, the **cycle** $Z_n(\mathbf{A}) = \ker d_n$, and the **homology**

$$H_n(\mathbf{A}) = Z_n(\mathbf{A}) / B_n(\mathbf{A}).$$

As before, we write $\mathbf{B}(\mathbf{A})$ (resp. $\mathbf{Z}(\mathbf{A})$) for the aggregate of the $B_n(\mathbf{A})$ (resp. $Z_n(\mathbf{A})$). Each of these is a chain complex with trivial boundary, as defined in Remark 25.33.

$\mathbf{H}_n(A) = 0$ iff the chain complex $\mathbf{A}$ is exact at A_n. From that point of view, $\mathbf{H}$ measures how far a chain complex is from being exact. A chain complex is called **acyclic** if its homology is 0 everywhere.

Remark 25.36′. (i) As with Remark 25.35, we can pass to cochains, our notation becoming Z^n, B^n, and H^n, respectively, for the **cocycle, coboundary**, and **cohomology**, which are even more relevant for our applications than homology. From this point of view, we can identify the homology of $\mathbf{Ch}_-$ with the cohomology of $\mathbf{Ch}^+$.

(ii) For any chain complex $(\mathbf{A}, d)$, we could define the **shifted chain complex** $(\mathbf{S}(\mathbf{A}), \mathbf{S}(d))$ by $\mathbf{S}(A)_n = A_{n-1}$ and $\mathbf{S}(d)_n = d_{n-1}$. This gives rise to the **shift functor** $\mathbf{S} \colon \mathbf{Ch} \to \mathbf{Ch}$ defined by $S(\mathbf{A}, d) = (\mathbf{S}(\mathbf{A}), \mathbf{S}(d))$, The shift functor is a useful bookkeeping device since the map $d \colon \mathbf{A} \to \mathbf{S}(\mathbf{A})$ is homogeneous and $H_n(\mathbf{A}) = H_{n+1}(\mathbf{S}(\mathbf{A}))$.

The theory becomes richer when we bring in some category theory.

Remark 25.37. Any commutative square

$$\begin{array}{ccc} A_1 & \xrightarrow{f_1} & A_1' \\ {\scriptstyle d}\downarrow & & \downarrow{\scriptstyle d'} \\ A_2 & \xrightarrow{f_2} & A_2' \end{array}$$

yields maps $\widetilde{f}_1 \colon \ker d \to \ker d'$, $\overline{f}_2 \colon \operatorname{coker} d \to \operatorname{coker} d'$, and $\hat{f}_2 \colon d(A_1) \to d'(A_1')$, and extends naturally to a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker d & \longrightarrow & A_1 & \xrightarrow{d} & A_2 & \longrightarrow & \operatorname{coker} d & \longrightarrow & 0 \\ & & {\scriptstyle \widetilde{f}_1}\downarrow & & {\scriptstyle f_1}\downarrow & & {\scriptstyle f_2}\downarrow & & \downarrow{\scriptstyle \overline{f}_2} & & \\ 0 & \longrightarrow & \ker d' & \longrightarrow & A_1' & \xrightarrow{d'} & A_2' & \longrightarrow & \operatorname{coker} d' & \longrightarrow & 0 \end{array}$$

(Indeed, $\widetilde{f}_1$ can be taken to be the restriction of f_1 to $\ker d$, for if $a_1 \in \ker d$, then $d' f_1(a_1) = f_2 d(a_1) = 0$. We can define $\hat{f}_2$ to be the restriction of f_2 to $d(A_1)$, since $f_2(d(A_1)) = d' f_1(A_1) \subseteq d'(A_1')$. Likewise, the composite

$$A_2 \xrightarrow{f_2} A_2' \to \operatorname{coker} d' = A_2' / d'(A_1')$$

sends $d(A_1) \to 0$, so Noether I (Theorem 1.13 of Volume 1) gives the map $\overline{f}_2$. The rest of the assertion is clear.)

Lemma 25.38 (The Snake Lemma). *Notation as in Remark 25.37.*

(i) *Any commutative diagram*

$$\begin{array}{ccccccc} & & A_1'' & \xrightarrow{f_1} & A_1 & \xrightarrow{g_1} & A_1' \\ & & {\scriptstyle d''}\downarrow & & {\scriptstyle d}\downarrow & & \downarrow{\scriptstyle d'} \\ 0 & \longrightarrow & A_2'' & \xrightarrow{f_2} & A_2 & \xrightarrow{g_2} & A_2' \end{array}$$

gives rise to an exact sequence $\ker d'' \xrightarrow{\tilde{f}_1} \ker d \xrightarrow{\tilde{g}_1} \ker d'$.

(ii) *Any commutative diagram*

$$\begin{array}{ccccccc} A_1'' & \xrightarrow{f_1} & A_1 & \xrightarrow{g_1} & A_1' & \longrightarrow & 0 \\ {\scriptstyle d''}\downarrow & & {\scriptstyle d}\downarrow & & \downarrow{\scriptstyle d'} & & \\ A_2'' & \xrightarrow{f_2} & A_2 & \xrightarrow{g_2} & A_2' & & \end{array}$$

gives rise to an exact sequence $\operatorname{coker} d'' \xrightarrow{\overline{f_2}} \operatorname{coker} d \xrightarrow{\overline{g_2}} \operatorname{coker} d'$.

(iii) *If, in* (i), g_1 *is epic, then there is an exact sequence*

$$\ker d'' \xrightarrow{\tilde{f}_1} \ker d \xrightarrow{\tilde{g}_1} \ker d' \xrightarrow{\partial} \operatorname{coker} d'' \xrightarrow{\overline{f_2}} \operatorname{coker} d \xrightarrow{\overline{g_2}} \operatorname{coker} d', \tag{25.7}$$

where ∂ *is defined in the proof and the other maps are from* (i) *and* (ii).

Proof. The maps $\tilde{f}_1, \tilde{g}_1, \overline{f_2}$, and $\overline{g_2}$ come from Remark 25.37. The technique of this proof is called **diagram chasing**.

(i) Clearly $f_1(\ker d'') \le \ker \tilde{g}_1$, and we need to show equality. Suppose $g_1(a) = 0$ for $a \in \ker d$. Then $a = f_1(a'')$ for $a'' \in A_1''$, but $f_2(d''(a'')) = d(a) = 0$, implying $d''(a'') = 0$; hence $a \in f_1(\ker d'')$.

(ii) The proof is dual to that of (i).

(iii) The exactness of the first two arrows in (25.7) is by part (i), and of the last two arrows by part (ii). Let us define ∂. Given $a' \in \ker d'$ write $a' = g_1(a)$; then $0 = d'g_1(a) = g_2 d(a)$, so $d(a) = f_2(a_2'')$ for some $a_2'' \in A_2''$; define $\partial(a') = a_2'' + d''(A_1'')$.

If we choose some other $\hat{a} \in A_1$ with $a' = g_1(\hat{a})$, then $\hat{a} - a \in f_1(A_1'')$, implying that $d(\hat{a}) - d(a) \in f_2 d''(A_1'')$; writing $d(\hat{a}) = f_2(\hat{a}_2'')$, we have nevertheless $f_2(\hat{a}_2'' - a_2'') \in f_2 d''(A_1'')$, implying that $\hat{a}_2'' - a_2'' \in d''(A_1'')$. Hence, ∂ is well-defined.

It remains to prove exactness of (25.7) at $\ker d'$ and at $\operatorname{coker} d''$. We check it at $\ker d'$. In the notation we have been using, we need to show that $a_2'' \in d''(A_1'')$ iff $a \in \ker g_1$. But clearly $g_1(a) = 0$ iff $a = f_1(a'')$ for some $a'' \in A_1''$, in which case $f_2 d''(a'') = d f_1(a'') = d(a) = f_2(a_2'')$, so we conclude that $a_2'' = d''(a'') \in d''(A_1'')$, as desired. (This argument is reversible.) The proof for exactness at $\operatorname{coker} d''$ is analogous, and is left for the reader. □

Remark 25.39. Given a chain map $f\colon (\mathbf{A}, d) \to (\mathbf{A}', d')$, we see from Remark 25.37 that $f_{n+1}(\ker d_n) \le \ker d_n'$ and $f_n(\operatorname{im} d_{n+1}) \le \operatorname{im} d_{n+1}'$; thus, f induces a map from $\ker d_n / \operatorname{im} d_{n+1} \to \ker d_n' / \operatorname{im} d_{n+1}'$. In other words, f induces maps

$$\mathbf{B}f_n\colon B_n(\mathbf{A}) \to B_n(\mathbf{A}'), \qquad \mathbf{Z}f_n\colon Z_n(\mathbf{A}) \to Z_n(\mathbf{A}'),$$

which in turn yields a map $\mathbf{H}f_n\colon H_n(\mathbf{A}) \to H_n(\mathbf{A}')$. In this manner $\mathbf{B}, \mathbf{Z}$, and $\mathbf{H}$ are functors, where $(\mathbf{BA})_n = B_n(\mathbf{A})$ and $(\mathbf{B}f)_n = \mathbf{B}f_n$, and likewise for $\mathbf{Z}$ and $\mathbf{H}$. $\mathbf{H}f$ is often denoted as f_*.

Remark 25.40. We can dualize in a different way from Remark 25.35, by taking the categorical duals of the cycle and boundary; namely, we define

$$Z_n' = \operatorname{coker} d_{n+1} = A_n / d_{n+1}(A_{n+1}) = A_n / B_n, \quad B_n' = A_n / \ker d_n = A_n / Z_n.$$

By Proposition 1.12 of Volume 1, the natural surjection $A_n \to B_n'$ induces a natural surjection $Z_n' \to B_n'$ that has kernel isomorphic to $Z_n / B_n = H_n$. $\mathbf{Z}'$ is useful, because of the following relationship to $\mathbf{Z}$.

Remark 25.41. The map d_{n+1} induces a map $\widehat{d}_{n+1}\colon Z_{n+1}'(\mathbf{A}) \to Z_n(\mathbf{A})$.

PROPOSITION 25.42. *The functor* $\mathbf{Z}$ *is left exact, and the functor* $\mathbf{Z}'$ *is right exact.*

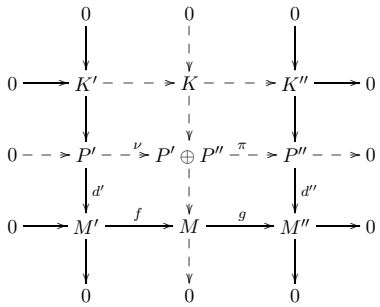
Proof. An immediate application of the Snake Lemma to Remark 25.3(i) shows that $\mathbf{Z}$ is left exact, and the other assertion is the dual (i.e., has the dual proof). □

Exact sequences of chain complexes.

Since we are studying resolutions of modules, our next objective is to lift exact sequences of modules to exact sequences of resolutions.

LEMMA 25.43 (THE HORSESHOE LEMMA). *Suppose one has the following "horseshoe" diagram of solid lines, where the rows and columns are exact, with* P' *and* P'' *projective:*

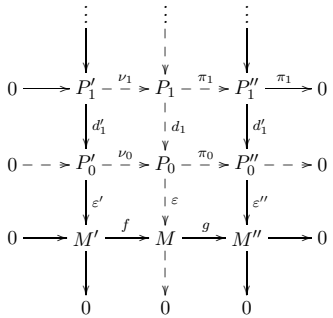
(25.8)



Then diagram (25.8) can be filled out to a commutative diagram (with all the dotted lines filled in), where all rows and columns are exact, and specifically, taking $P = P' \oplus P''$, $\nu\colon P' \to P$ is the canonical monic and $\pi\colon P \to P''$ is the canonical epic.

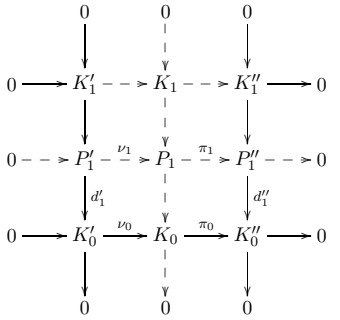
Proof. First, since P'' is projective, d'' can be lifted to a map $h\colon P'' \to M$; i.e., $d'' = gh$. Define $d\colon P \to M$ by $d(p', p'') = fd'(p') + h(p'')$. Now the bottom two rows make a commutative diagram, which we fill in with the kernels by means of Lemma 25.38(i). □

THEOREM 25.44. *For any exact sequence $0 \to M' \to M \to M'' \to 0$ of modules and respective projective resolutions $(\mathbb{P}', d')$ and $(\mathbb{P}'', d'')$ of M' and M'', there exists a projective resolution $(\mathbb{P}, d)$ of M, such that $P_n = P'_n \oplus P''_n$ for each n, for which the following diagram is commutative:*



Proof. Let $\nu_n\colon P'_n \to P_n$ be the canonical monic and $\pi\colon P_n \to P''_n$ the canonical projection. Also let $K'_0 = \ker \varepsilon'$ and $K''_0 = \ker \varepsilon''$. One applies the Horseshoe Lemma, writing P_0 instead of P, to get $\varepsilon\colon P_0 \to M$. But then

letting $K_0 = \ker \varepsilon$ and applying the Horseshoe Lemma to the diagram



we get the next stage, and iterating yields the entire theorem. □

The long exact sequence.

One major goal in the theory is to compute the derived functors Tor and Ext, to be defined below. To do this, we need to build up homology (resp. cohomology) of chain complexes in terms of homology (resp. cohomology) of more basic chain complexes. Specifically, given an exact sequence of chain complexes

$$0 \to (\mathbf{A}'', d'') \to (\mathbf{A}, d) \to (\mathbf{A}', d') \to 0,$$

we need to be able to describe $H(\mathbf{A})$ in terms of $H(\mathbf{A}'')$ and $H(\mathbf{A}')$. The key is contained in the following subtle result.

THEOREM 25.45. *If $0 \to (\mathbf{A}'', d'') \xrightarrow{f} (\mathbf{A}, d) \xrightarrow{g} (\mathbf{A}', d') \to 0$ is a short exact sequence of complexes, then there is a* **long exact sequence**

(25.9)
$$\cdots \to H_{n+1}(\mathbf{A}'') \xrightarrow{f_*} H_{n+1}(\mathbf{A}) \xrightarrow{g_*} H_{n+1}(\mathbf{A}') \xrightarrow{\partial_*} H_n(\mathbf{A}'') \xrightarrow{f_*} H_n(\mathbf{A}) \xrightarrow{g_*} \cdots$$

where $(\partial_*)_{n+1}\colon H_{n+1}(\mathbf{A}') \to H_n(\mathbf{A}'')$ is obtained by applying the Snake Lemma to the commutative diagram

(25.10)
$$\begin{array}{ccccccc} & & Z'_{n+1}(\mathbf{A}'') & \xrightarrow{f_{n+1}} & Z'_{n+1}(\mathbf{A}) & \xrightarrow{g_{n+1}} & Z'_{n+1}(\mathbf{A}') \longrightarrow 0 \\ & & \downarrow \widehat{d}''_{n+1} & & \downarrow \widehat{d}_{n+1} & & \downarrow \widehat{d}'_{n+1} \\ 0 & \longrightarrow & Z_n(\mathbf{A}'') & \xrightarrow{\widetilde{f}_n} & Z_n(\mathbf{A}) & \xrightarrow{\widetilde{g}_n} & Z_n(\mathbf{A}') \end{array},$$

where the columns are obtained from Remark 25.37.

Proof. This is the Snake Lemma applied at each n. □

As we see below, Theorem 25.45 often is utilized by showing that certain key positions of the long exact sequence are 0, thereby yielding isomorphisms along the way.

δ-functors and derived functors

In studying homology and cohomology, we may want some flexibility to move up and down various resolutions. As preparation, we encapsulate the basic abstract properties of the homology functor **H** by means of a definition.

Definition 25.46. A **homological δ-functor T** is a collection of additive functors $T_n\colon \mathcal{D} \to \mathcal{C}$ such that for every short exact sequence

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

in $\mathcal{D}$, we have morphisms $\delta_n\colon T_n(A') \to T_{n-1}(A'')$ and a long exact sequence

(25.11)
$$\begin{aligned} \dots \xrightarrow{\delta} T_{n+1}(A'') \xrightarrow{Tf_{n+1}} T_{n+1}(A) \xrightarrow{Tg_{n+1}} T_{n+1}(A') \xrightarrow{\delta} T_n(A'') \\ \xrightarrow{Tf_n} T_n(A) \xrightarrow{Tg_n} T_n(A') \xrightarrow{\delta} \dots \end{aligned}$$

such that T_0 is right exact, and the δ are natural in the sense that they commute with morphisms of short exact sequences in $\mathcal{D}$; in other words, given a commutative diagram of exact sequences

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A'' & \longrightarrow & A & \longrightarrow & A' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f''} & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle f'} & & \\ 0 & \longrightarrow & \widetilde{A}'' & \longrightarrow & \widetilde{A} & \longrightarrow & \widetilde{A}' & \longrightarrow & 0 \end{array}$$

we have commutative diagrams

$$\begin{array}{ccc} T_n(A') & \xrightarrow{\delta} & T_{n-1}(A'') \\ \downarrow & & \downarrow \\ T_n(\widetilde{A}') & \xrightarrow{\delta} & T_{n-1}(\widetilde{A}'') \end{array}.$$

A cohomological δ-functor is defined analogously (where now T^0 is left exact and we have $\delta^n\colon T^n(A') \to T^{n+1}(A'')$).

A **morphism T → S** of δ-functors is a system of natural transformations $T_n \to S_n$ that commute with δ; i.e., the diagram

$$\begin{array}{ccccccccccc} \cdots & T_{n+1}(A'') & \xrightarrow{Tf_{n+1}} & T_{n+1}(A) & \xrightarrow{Tg_{n+1}} & T_{n+1}(A') & \xrightarrow{\delta} & T_n(A'') & \cdots \\ & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\ \cdots & S_{n+1}(A'') & \xrightarrow{Sf_{n+1}} & S_{n+1}(A) & \xrightarrow{Sg_{n+1}} & S_{n+1}(A') & \xrightarrow{\delta} & S_n(A'') & \cdots \end{array}$$

commutes for every short exact sequence $0 \to A'' \to A \to A' \to 0$ in $\mathcal{D}$.

In many applications, $\mathcal{D} = \mathbf{Ch}(\mathcal{C})$; the homology functor **H** is an example of a homological δ-functor, in view of the long exact sequence (25.9). Cohomology on $\mathbf{Ch}(\mathcal{C})^+$, viewed dually as homology on $\mathbf{Ch}(\mathcal{C})_-$, translates to a homological δ-functor.

However, the main point of this exposition has been to lay the groundwork for another very important example of a homological δ-functor. We want to define the n-th **derived functor T** of a half-exact functor F whose action on a module M is to be defined as $T_n(M) = H_n(F(\mathbb{P}))$ for a suitable resolution $\mathbb{P}$ of M.

Remark 25.47. Our main objective is to show that **T** is a homological δ-functor. This raises the following issues:

1. (Well-definedness) Since there is ambiguity as to the choice of resolution of M, do we get the same result for $T_n(M)$?
2. How do we lift a morphism $f\colon M \to M'$ to a morphism $\mathbf{f}\colon \mathbb{P} \to \mathbb{P}'$, so that we can define $T_n(f) = H_n(F(\mathbf{f}))$?
3. How can we verify the important property (25.11) of Definition 25.46? (Furthermore, we want the long exact sequence to start or end with 0.)
4. (The first step in computing $T_n(M)$): When is $T_n(M) = 0$?

As the reader may have guessed from the notation, $\mathbb{P}$ usually is taken to be a projective resolution of A. To deal with (1)–(4), we must examine projective resolutions more closely, especially in conjunction with homology.

Homology on projective resolutions.

There is a general instance when two chain maps induce the same maps on homology.

Definition 25.48. Chain maps $f, g\colon (\mathbf{A}, d) \to (\mathbf{A}', d')$ are called **homotopic**, written $f \sim g$, if there is a graded map $s\colon \mathbf{A} \to \mathbf{A}'$ of degree 1 such that $f - g = d's + sd$. Two chain complexes $(\mathbf{A}, d)$ and $(\mathbf{A}', d')$ are **homotopy equivalent** if there are chain maps $f\colon (\mathbf{A}, d) \to (\mathbf{A}', d')$ and $f'\colon (\mathbf{A}', d') \to (\mathbf{A}, d)$ such that $f'f \sim 1_{\mathbf{A}}$ and $ff' \sim 1_{\mathbf{A}'}$.

Remark 25.49. (i) If f and g are homotopic and $F\colon \mathbf{Ch} \to \mathbf{Ch}$ is a functor, then Ff and Fg also are homotopy equivalent; indeed, if $f - g = d's + sd$, then $Ff - Fg = Fd'Fs + FsFd$.

(ii) Homotopic chain maps induce the same maps on the homology. Indeed, if $z \in Z_n$, then writing $f - g = d's + sd$, we have

$$f_n(z) - g_n(z) = d'_n s(z) + sd_n(z) = d'_n s(z) \in B'_n,$$

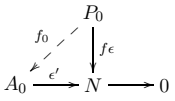
so $f_*(z) = g_*(z)$. In particular, any chain complex which is homotopy equivalent to 0 is acyclic.

(iii) Any functor preserves homotopy equivalence by (i), and thus defines the same homology for homotopy equivalent chain complexes.

Now we bring in projective resolutions.

PROPOSITION 25.50. *Given a map $f\colon M \to N$ of modules, a resolution $\mathbf{A}$ of N, and a projective resolution $\mathbf{P}$ of M, one can lift f to a chain map $\mathbf{f}\colon \mathbf{P} \to \mathbf{A}$ that is unique up to homotopy equivalence.*

Proof. Write the resolution $\mathbf{P}$ as $\cdots \to P_n \xrightarrow{d_n} P_{n-1} \xrightarrow{d_{n-1}} \dots \xrightarrow{d_1} P_0$ and $\mathbf{A}$ as $\cdots \to A_n \xrightarrow{d'_n} A_{n-1} \xrightarrow{d'_{n-1}} \dots \xrightarrow{d'_1} A_0$, together with the respective epics $\epsilon\colon P_0 \to M$ and $\varepsilon'\colon A_0 \to N$. Define $f_0\colon P_0 \to A_0$ via the diagram



(cf. Definition 25.8). Inductively, given $f_{n-1}\colon P_{n-1} \to A_{n-1}$, we can define $f_n\colon P_n \to A_n$ as follows:

The map $h = f_{n-1}d_n\colon P_n \to A_{n-1}$ satisfies

$$d'_{n-1}h = d'_{n-1}f_{n-1}d_n = f_{n-2}d_{n-1}d_n = 0;$$

by Proposition 25.10(5), h lifts to a map $f_n\colon P_n \to A_n$ with $h = d'_n f_n$, as desired; clearly $\mathbf{f} = (f_n)$ is a chain map.

For the last assertion, suppose we have some other lifting of f to a chain map $(g_n\colon P_n \to A_n)$, and let $q_n = g_n - f_n$. Note that

$$d'_n q_n = d'_n g_n - d'_n f_n = g_{n-1}d_n - f_{n-1}d_n = q_{n-1}d_n.$$

Since $\epsilon' q_0 = (f - f)\epsilon = 0$, Proposition 25.10(5) yields a map $s_0\colon P_0 \to A_1$ with $q_0 = d'_1 s_0$. Put $s_{-1} = 0$. Thus $q_0 = d'_1 s_0 - s_{-1}d_0$. Inductively, suppose we have $s_{n-1}\colon P_{n-1} \to A_n$ such that $q_{n-1} = d'_n s_{n-1} + s_{n-2}d_{n-1}$; then

$$d'_n(q_n - s_{n-1}d_n) = q_{n-1}d_n - (q_{n-1} - s_{n-2}d_{n-1})d_n = q_{n-1}d_n - q_{n-1}d_n = 0.$$

Using Proposition 25.10(5), we can find $s_n\colon P_n \to A_{n+1}$ such that $d'_{n+1}s_n = q_n - s_{n-1}d_n$, so $q_n = d'_{n+1}s_n + s_{n-1}d_n$. □

One can interpret Proposition 25.50 as a universal, leading us at once to the following consequence.

COROLLARY 25.51. *Any two projective resolutions of a module M are homotopy equivalent.*

Proof. By abstract nonsense. Namely, we take $\mathbf{A}$ to be a projective resolution of M, and taking $M = N$ in Proposition 25.50, we have chain maps $f\colon \mathbf{P} \to \mathbf{A}$ and $g\colon \mathbf{A} \to \mathbf{P}$ lifting the identity map on M. But the composition in each direction must be homotopic to the identity, by the uniqueness in Proposition 25.50, proving that $\mathbf{P}$ and $\mathbf{A}$ are homotopy equivalent. □

The derived homological functors.

Example 25.52. (i) Suppose F is a right exact covariant functor. Then we can define $L_n(M) = H_n(F(\mathbb{P}))$ (with $L_0(M) = F(M)$), where $\mathbb{P}$ is a projective resolution of M. This definition is well-defined, in view of Remark 25.49 and Corollary 25.51. We define the action of L_n on maps $f\colon M \to M'$ by lifting f to a chain map of projective resolutions, via Proposition 25.50, and then take the corresponding maps of the homology modules. The exact sequence (25.11) comes from the long exact sequence in Theorem 25.44; since F is left exact, the long exact sequence can be truncated to yield

(25.12)
$$\cdots \to L_1F(A'') \to L_1F(A) \xrightarrow{L_1Fg} L_1F(A') \to FA'' \to FA \xrightarrow{Fg} FA' \to 0.$$

Thus the L_nF comprise a homological δ-functor, called the **left derived (homological) functor** of F.

(ii) Suppose G is a left exact contravariant functor. Then we can define $R^n(M) = H^n(G(\mathbb{P}))$ (with $R_0(M) = F(M)$), where $\mathbb{P}$ is a projective resolution of M. As in (i), this definition is well-defined, and defines the action of L_n on maps $f\colon M \to M'$ as in (i). The property (25.11) comes from the long exact sequence in Theorem 25.44. Since G is contravariant and left exact, the long exact sequence can be truncated at

$$0 \to GA'' \to GA \to GA' \to R^1G(A'') \to R^1G(A) \to R^1G(A') \to \cdots.$$

Thus the R^nF comprise a cohomological δ-functor, called the **right derived (cohomological) functor** of G.

(iii) Suppose G is a left exact covariant functor. Ironically, working backwards, we see that we must start with 0 and use cohomology; projective resolutions do not work here, and instead we must use injective resolutions. The dual versions of Remark 25.49 and Corollary 25.51 hold, so we can define the cohomological δ-functor $R^n(M) = H^n(G(\mathbb{E}))$, where $\mathbb{E}$ is an injective resolution of M; this is also called the **right derived (cohomological) functor** of G. (Thus we distinguish between (ii) and (iii) depending on whether G is covariant or contravariant.)

We still have not addressed the question of which modules become trivial under the derived functors.

Remark 25.53. If K_n is the n-th syzygy of a projective resolution $\mathbb{P}$ of M, and $K_0 = \ker \varepsilon$, then $L_nF(M) = L_1F(K_n)$ for each n, as seen by cutting the projective resolution (cf. Remark 25.19).

PROPOSITION 25.54. *The following conditions are equivalent for a right exact covariant functor F:*

(1) *F is exact.*
(2) *$L_1F = 0$.*
(3) *$L_nF = 0$ for all n.*

Proof. (2) implies (3) by Remark 25.53; the other two implications are by definition. □

The two major examples, $\otimes$ and Hom deserve separate billing.

Definition 25.55. When F is the functor $M \otimes \_$, for some right module M, the left derived functor L_nF is called $\mathrm{Tor}_n(M, \_)$.

When $F = \_ \otimes N$ for some (left) module N, L_nF is called $\mathrm{Tor}_n(\_, N)$.

For $G = \mathrm{Hom}(M, \_)$, the right derived functor R_nG is called $\mathrm{Ext}^n(M, \_)$. (The superscripts are used because we are in the cohomological setting.)

For $G = \mathrm{Hom}(\_, M)$, the right derived functor R_nG is called $\mathrm{Ext}^n(\_, M)$.

Tor and flat modules.

We turn to Tor, the derived functors of $\otimes$. Let us start by fixing a right module M. Given an exact sequence of modules $0 \to N' \to N \to N'' \to 0$, we get the long exact sequence

(25.13)
$$\cdots \to \mathrm{Tor}_{n+1}(M, N') \to \mathrm{Tor}_n(M, N'') \to \mathrm{Tor}_n(M, N) \to \mathrm{Tor}_n(M, N') \to$$
$$\cdots \to M \otimes N'' \to M \otimes N \to M \otimes N' \to 0.$$

In one sense the theory should run smoothly since the functor $M \otimes \_$ is covariant and right exact, enabling us to utilize projective resolutions in our computations. The hitch is that we have not yet characterized those right modules M for which the functor $M \otimes \_$ is exact.

Definition 25.56. A right R-module M is **flat** if, whenever $f\colon N_1 \to N_2$ is monic, then $1 \otimes f : M \otimes_R N_1 \to M \otimes_R N_2$ is also monic.

Remark 25.57. The right module M is flat iff the functor $M \otimes_R \_$ is exact, since right exactness is seen in Example 25.5. Consequently, by Proposition 25.54, the following assertions are equivalent for a module P:

1. P is flat.
2. $\mathrm{Tor}_1(M, P) = 0$ for all modules M.
3. $\mathrm{Tor}_n(M, P) = 0$ for all modules M and all n.

Having catapulted flat modules into the center of homology theory, let us consider a few of their properties.

PROPOSITION 25.58. *The direct sum $\bigoplus M_i$ of right modules is flat iff each M_i is flat.*

Proof. By Exercise 18.2, $\ker(f \otimes 1_M) = \bigoplus \ker(f \otimes 1_{M_i})$, which is 0 iff each direct summand is 0. □

COROLLARY 25.59. *Every projective module P is flat.*

Proof. First we take P free, and then apply the proposition. □

This is a fortunate result, since now our next quest, the existence of a "flat resolution," is superfluous; we already have free resolutions, constructed rather easily, which are projective resolutions and thus flat resolutions!

Remark 25.60. Suppose $0 \to N'' \to P \to N' \to 0$ is an exact sequence of modules, with P flat. Then in the long exact sequence (25.13) we have $\mathrm{Tor}_n(M, P) = 0$ for each $n \geq 1$, implying that $\mathrm{Tor}_{n+1}(M, N') = \mathrm{Tor}_n(M, N'')$ for each $n \geq 0$. This kind of identification is called **dimension shifting**, and often enables us to compute Tor by means of exact sequences.

The notation for Tor raises another question: Does $\mathrm{Tor}_n(M, N)$ denote $\mathrm{Tor}_n(M, \_)(N)$ or $\mathrm{Tor}_n(\_, N)(M)$? Fortunately, these are both the same. This can be seen via an ad hoc argument given in Exercise 47. A more natural approach is to treat both sides in the same **double complex**, or

bicomplex, defined in Exercise 49. An elegant proof (shown to me by S. Shnider) that the two definitions of Tor are the same is outlined in Exercise 51; see [**Wei,** §**2.7**] for a detailed treatment.

Ext.

We recall that, being the derived functor of Hom, which is covariant on one side and contravariant on the other, Ext is defined in terms of projective resolutions on one side and injective resolutions on the other.

The same wrinkle has developed: Does $\mathrm{Ext}^n(M,N)$ mean $\mathrm{Ext}^n(M,\_)$ applied to N, or $\mathrm{Ext}^n(\_,N)$ applied to M? The first way would be to take an injective resolution $\mathbb{E}$ of N and compute the cohomology of $\mathrm{Hom}(M,E_n)$; the second way would be to take a projective resolution $\mathbb{P}$ of M and compute the cohomology of $\mathrm{Hom}(P_n,N)$. Again, these are the same; cf. Exercise 48. Thus, we can compute $\mathrm{Ext}^n(M,N)$ using only projective resolutions, which often are more accessible than injective resolutions.

Ironically, in algebraic geometry one usually uses the version of Ext involving injective resolutions, since coherent sheaves always have injective resolutions (but may not have projective resolutions)!

Remark 25.61. By Proposition 25.54, the following assertions are equivalent for a module P:

1. P is projective.
2. $\mathrm{Ext}^n(P,M)=0$ for all n and all modules M.
3. $\mathrm{Ext}^1(P,M)=0$ for all modules M.

Likewise, the following assertions are equivalent for a module E:

1. E is injective.
2. $\mathrm{Ext}^n(N,E)=0$ for all n and all modules N.
3. $\mathrm{Ext}^1(N,E)=0$ for all modules N.

Remark 25.62. A derived homological or cohomological δ-functor $\mathbf{T}$ is called **universal** if, for any other δ-functor S, any natural transformation $\eta_0\colon T_0 \to S_0$ can be extended to unique morphisms of functors $\eta\colon \mathbf{T} \to \mathbf{S}$. Thus, by abstract nonsense, different constructions providing the same universal property lead to the same δ-functor.

As we have seen, Tor and Ext are derived δ-functors. Moreover, they are universal in this sense. The demonstration is long, given in [**Wei**], but not too difficult, accomplished by means of the mapping cone of Exercise 52. This enables us to construct Tor and Ext in various ways, and is a key feature of the theory; in particular, it provides an alternative proof that $\mathrm{Ext}^n(M,N)$ does not depend on which of the two components we use to build the resolution.

Digression 25.62′. One might try to use Remark 25.26 to pass from the first definition of Ext to the second via the modules $E_n = \mathrm{Hom}(P_n,N)$. However, the module E_n need not necessarily be injective. Cartan and Eilenberg [**CarE, X.8**] deal with this by weakening the definitions of projectivity and injectivity.

Another way is to make the E_n injective by fiat — a ring R is called **quasi-Frobenius**, or **QF**, if R is injective as an R-module. For example, any semisimple ring is QF, as is any group algebra over a field; cf. Exercise 58. (More generally, f.d. Hopf algebras are QF; cf. Remark 26.29′, and the cohomology theory of QF rings is often applied to f.d. Hopf algebras.)

Examples of homology and cohomology

Let us consider some basic examples, illustrating the wide range of the theory. The main idea is to interpret algebraic structures in terms of module theory and then take the appropriate projective (or injective) resolution of the ensuing module. There is a slight complication in Definition 25.55, insofar as M must be a right module. However, in all of our examples, M will be naturally a bimodule.

Since projective resolutions (especially free resolutions) are usually easier to construct than injective resolutions, we usually start with free resolutions, passing at times to cohomology by dualizing via the dual module of Remark 25.26.

There is a very well-developed theory of homology and cohomology of modules over a commutative ring C, which we do not consider in this text. One aspect of particular historical importance is the problem of finding **extensions of modules** — given modules M,N we want to find a module E together with an exact sequence $0 \to N \to E \to M \to 0$. The possible such E (up to a certain equivalence) are in 1:1 correspondence with $\mathrm{Ext}^1(M,N)$; cf. Exercise 65.

Example 25.63 (The standard resolution of a group, and group homology).

(i) The group homology of a group G is defined as the homology of $\mathbb{Z}$ as a trivial G-module (and thus as a $\mathbb{Z}[G]$-module). A free resolution of $\mathbb{Z}$ is usually described directly for a finite group, in terms of simplices.

Any finite group G can be viewed as a simplex whose vertices are the elements of G; the face of dimension n is identified as the set of simplices. Explicitly, we define a free resolution $\mathbf{F}$ of $\mathbb{Z}$ as follows:

F_n is the free $\mathbb{Z}[G]$-module generated by $(n+1)$-tuples $(g_0, \dots, g_n)$ from G, with the natural action

$$g(g_0, \dots, g_n) = (gg_0, \dots, gg_n);$$

F_n is seen to be free, isomorphic to $\mathbb{Z}[G]^{[n]}$, for we obtain a base when we normalize to assume $g_0 = 1$.

As usual, one defines the boundary map $d_n\colon F_n \to F_{n-1}$, via

$$d_n(g_0, \dots, g_n) = \sum (-1)^i (g_0, \dots, g_{i-1}, g_{i+1}, \dots, g_n);$$

clearly $d^2 = 0$.

The ensuing chain $\mathbf{F} = (F_n, d)$ of free modules is a free resolution of $\mathbb{Z}$ as the trivial $\mathbb{Z}[G]$-module (i.e., $g1 = 1$, $\forall g \in G$), with respect to the **augmentation map** $\epsilon\colon F_0 \to \mathbb{Z}$ given by $g_0 \mapsto 1$, $\forall g_0 \in G$. We write $[g_1|g_2|\cdots|g_n]$ for $(1, g_1, g_1g_2, \dots, g_1g_2\cdots g_n)$. In this "bar" notation, we have

(25.14)
$$\begin{aligned} d_n\big([g_1|g_2|\cdots|g_n]\big) =& g_1[g_2|\cdots|g_n] \\ &+ \sum_{1\le i<n} (-1)^i [g_1|\cdots|g_{i-1}|g_ig_{i+1}|g_{i+2}|\cdots|g_n] \\ &+ (-1)^n [g_1|\cdots|g_{n-1}], \end{aligned}$$

called the **bar resolution** of G. In this manner, we can calculate $H_*(G)$, defined to be the homology of $\mathbb{Z}$ as a $\mathbb{Z}[G]$-module.

(ii) Now any G-module N has the resolution $\mathbf{F} \otimes_{\mathbb{Z}[G]} N$, where by definition $(\mathbf{F} \otimes_{\mathbb{Z}[G]} N)_n = F_n \otimes_{\mathbb{Z}[G]} N$, the tensor product of G-modules taken as in Definition 20.27, the boundary maps being $d_n \otimes 1$. The ensuing homology $\operatorname{Tor}(\mathbb{Z}, N)$ is called $H_*(G, N)$, the **homology of G with coefficients in N**. Note that $H_0(G, N) = \mathbb{Z} \otimes_{\mathbb{Z}[G]} N \approx N/\mathcal{I}N$, where $\mathcal{I}$ is the augmentation ideal of $\mathbb{Z}[G]$; cf. Example 19.17.

Remark 25.64. Although we derived (25.13) for groups, it makes sense also for monoids, and thus for the multiplicative monoid of an algebra; this point of view is taken in Exercise 25B.5.

Example 25.65 (Group cohomology). To compute the cohomology of a group, we need to reverse arrows in Example 25.63. For any G-module N, we define $H^n(G, N)$ to be $\operatorname{Ext}^n(\mathbb{Z}, N)$, which we view as $\operatorname{Ext}^n(\_, N)$ applied to $\mathbb{Z}$. Thus, we construct a free resolution of $\mathbb{Z}$ as follows: Observing that $\mathbb{Z}[G]^{\otimes n} \cong \mathbb{Z}[G^{(n)}]$ is free as a $\mathbb{Z}$-module with base $G^{(n)}$, take

$$C^n = C^n(G, M) = \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}[G]^{\otimes n}, M).$$

Of course, any $f \in C^{(n)}$ is determined by its action on the base; i.e., f is determined by $\{f(g_1, \dots, g_n) : g_i \in G\}$. We go through a procedure reversing that of Example 25.63; given $f \in C^n$ we define $\widetilde{f}\colon G^{n+1} \to M$ by

$$\widetilde{f}(g_0, g_1, \dots, g_n) = g_0 f(g_0^{-1}g_1, \dots, g_{n-1}^{-1}g_n).$$

Then $\widetilde{f}(gg_0, gg_1, \dots, gg_n) = g\widetilde{f}(g_0, g_1, \dots, g_n)$ for all $g \in G$.; thus, $\tilde{f} \in \operatorname{Hom}_G(\mathbb{Z}[G]^{\otimes n}, M)$, which is consistent with Example 25.6(i).

Letting $h_0 = 1$ and $h_i = g_1g_2\cdots g_i$ for $i \ge 1$, note that

$$f(g_1, \dots, g_n) = \widetilde{f}(h_0, h_1, \dots, h_n).$$

Now, given $f \in C^n$, define

$$\delta^n f(g_1, \dots, g_{n+1}) = \sum_{i=0}^{n+1} (-1)^i \widetilde{f}(h_0, h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_{n+1});$$

explicitly, this is

(25.15)
$$\begin{aligned} g_1 f(g_2, \dots, g_{n+1}) + \sum_{i=1}^{n} (-1)^i f(g_1, \dots, g_{i-1}, g_ig_{i+1}, g_{i+2}, \dots, g_{n+1}) \\ + (-1)^{n+1} f(g_1, \dots, g_n). \end{aligned}$$

$C^0(G, M)$ is the set of maps from the set $\{e\}$ to M; this can be identified with M. For $f \in C^0(G, M)$, $\delta f(g_1) = g_1 f - f$.

$C^1(G, M) = \operatorname{Hom}(\mathbb{Z}[G], M)$. $\delta f(g_1, g_2) = g_1 f(g_2) - f(g_1g_2) + f(g_1)$.

For $f \in C^2(G, M)$,

$$\delta f(g_1, g_2, g_3) = g_1 f(g_2g_3) - f(g_1g_2, g_3) + f(g_1, g_2g_3) - f(g_1, g_2).$$

$H^0(G, N) = \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}, N)$.

Here are a few examples of group cohomology; also cf. Exercise 68.

Example 25.66. The multiplicative group $K^\times$ of a field K is an Abelian group, and thus a $\mathbb{Z}$-module. Suppose a group G acts as automorphisms on K; then $K^\times$ is a G-module, and thus a $\mathbb{Z}[G]$-module.

In Example 25.65 we identified $C^0(G, K^\times)$ with $K^\times$. An element $f \in K^\times$ is in Z^0 iff $\sigma(f) = f$, $\forall \sigma \in G$. Thus, $Z^0 = (K^G)^\times$.

Now consider $f \in C^1(G, K^\times)$; i.e., $f\colon G \to K^\times$. To say $f \in Z^1$ means $f(\sigma_1\sigma_2) = \sigma_1 f(\sigma_2) + f(\sigma_1)$. In the case when $G = \langle\sigma\rangle$ cyclic of order m,

written multiplicatively, this is the same as $f(\sigma^i) = \sigma(f(\sigma^{i-1}))f(\sigma)$ for each i (seen by taking $\sigma_1 = \sigma$ and $\sigma_2 = \sigma^{i-1}$), so $f(1) = 1$ and, by induction,

$$f(\sigma^i) = f(1)f(\sigma)\sigma^2(f(\sigma)) \cdots \sigma^{i-1}(f(\sigma)) = c\sigma(c)\sigma^2(c) \cdots \sigma^{i-1}(c), \tag{25.16}$$

where $c = f(\sigma)$. We could use (25.16) to define f, which is well-defined iff the norm $N_\sigma(c) = 1$, since then

$$f(1) = 1 = c\sigma(c)\sigma^2(c) \cdots \sigma^{m-1}(c) = f(\sigma^m).$$

On the other hand, $f \in B^1$ iff $f = \sigma(a)a^{-1}$ for some $a \in K^\times$. Thus Hilbert's Theorem 90 (Lemma 4.95 of Volume 1) is equivalent to the assertion that $H^1(G, K^\times) = 1$ whenever G is a finite cyclic group of automorphisms of a field K.

(ii) For any finite group G, $f \in Z^2(G, K^\times)$ says

$$f(\sigma, \tau)f(\sigma\tau, \rho) = \sigma(f(\tau\rho))f(\sigma, \tau\rho).$$

Thus, defining $c_{\sigma,\tau} = f(\sigma, \tau)$ for σ, τ in G, we see that the factor set condition of Equation (24.2) is satisfied iff $f \in Z^2(G, K^\times)$. On the other hand, by Exercise 24.23, the corresponding crossed product algebra $(K, G, (c_{\sigma,\tau}))$ is trivial iff $f \in B^2(G, K^\times)$. Letting $F = K^G$, we see that $H^2(G, K^\times)$ can be identified with the subgroup of $\mathrm{Br}(F)$ corresponding to central simple algebras (csa's) having a maximal subfield isomorphic to K. In other words, $H^2(G, K^\times)$ is identified with the part of the Brauer group split by K. In view of the Koethe-Noether-Jacobson Theorem (24.51), each csa is split by a finite separable extension of F, so we identify $H^2(G, K^\times)$ with a subgroup of $H^2(G, F_s)$, where F_s denotes the separable algebraic closure of F.

This fundamental observation is the basis of much of the study of the Brauer group, and in fact has led to the generalization of Example 25.65 to profinite groups G, where $C^n(G, M)$ is taken to be the set of continuous maps from G^n to M. Then we can identify $H^2(G, F_s)$ with the direct limit of the $H^2(G, K^\times)$. This is the starting point for Serre [**Ser3**].

Shapiro's Lemma and the transfer map.

We turn to cohomology of induced and coinduced modules, as described in Remark 20.41. Given a group G, a subgroup L, and an L-module M, write M_L^G for the induced module $\mathbb{Z}[G] \otimes_{\mathbb{Z}[L]} M$ and $\mathrm{Coind}_L^G(M)$ for the coinduced module $\mathrm{Hom}_{\mathbb{Z}[L]}(\mathbb{Z}[G], M)$.

Lemma 25.67 (Shapiro's Lemma).
(i) *$H_n(G, M_L^G) \cong H_n(L, M)$ for each L-module M and all n;*
(ii) *$H^n(G, \mathrm{Coind}_L^G(M)) \cong H^n(L, M)$ for all n.*

Proof. (i) We tensor up a projective resolution $\mathbb{P}$ of $\mathbb{Z}$ to get

$$\mathbb{P} \otimes_{\mathbb{Z}[G]} (\mathbb{Z}[G] \otimes_{\mathbb{Z}[L]} M) \cong \mathbb{P} \otimes_{\mathbb{Z}[L]} M.$$

(ii) We take the cohomology of

$$\mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{P}, \mathrm{Hom}_{\mathbb{Z}[L]}(\mathbb{Z}[G], M)) \approx \mathrm{Hom}_{\mathbb{Z}[L]}(\mathbb{P}, M). \qquad \square$$

Remark 25.68. Suppose the subgroup L has finite index in G. We apply Shapiro's Lemma to the isomorphism $\mathrm{Hom}_{\mathbb{Z}[L]}(\mathbb{Z}[G], M) \cong \mathrm{Coind}_L^G(M) \cong M_L^G$ (cf. Remark 20.41) to see that the natural map $M_L^G \to M$ induces maps $H^n(L, M) \to H^n(G, M_L^G) \to H^n(G, M)$ for any G-module M. These maps, called the **transfer maps**, play a fundamental role in cohomology theory. For example, the corestriction map of Chapter 24 is a special case; cf. Exercise 69.

Ext and extensions.

Ext^1 is also used to calculate extensions in exact sequences of groups.

Example 25.70. Suppose A is a normal Abelian subgroup of E, and $G = E/A$. Then G acts on A via conjugation; writing $g = wA$ for $w \in E$, one defines $ga = waw^{-1}$, which is independent of the choice of representative w since A is Abelian. It is customary to write A in additive notation, although G is written in multiplicative notation.

An exact sequence $0 \to A \xrightarrow{i} E \xrightarrow{\pi} G \to 1$ is called an **extension of G by A**. We say the extension $0 \to A \xrightarrow{i} E \xrightarrow{\pi} G \to 1$ **splits** if there is a group homomorphism $s\colon G \to E$ such that $\pi s = 1_G$; we call s a **splitting** of the group extension. In this case, there is a bijection $A \times G \to E$, given by $(a, g) \mapsto i(a)s(g)$, and the group operation on E is given by

$$(a, g)(b, h) = i(a)s(g)i(b)s(h) = i(a)i(b)^{s(g)}s(g)s(h) = (a + gb, gh).$$

The reader might recognize this as the **semidirect product** of G and A.

Conversely, how can we classify those group homomorphisms $s\colon G \to E$ that split E? Clearly $s(g) = (\delta(g), g)$ for some function $\delta\colon G \to A$. Then

$$s(gh) = s(g)s(h) = (\delta(g), g)(\delta(h), h) = (\delta(g) + g\delta(h), gh),$$

implying $\delta(gh) = \delta(g) + g\delta(h)$. But this is $Z^1(G, A)$. Thus, we get a map from $Z^1(G, A)$ to the semidirect product of G and A.

On the other hand, we can identify two splittings s_1 and s_2, calling them A-**conjugate** if there is $a \in A$ such that $s_1(g) = s_2(g)^{i(a)}$ for all $g \in G$. Taking δ_i corresponding to s_i, we then see

$$(\delta_1(g), g) = (a, 1)(\delta_2(g), g)(a, 1)^{-1} = (a + \delta_2(g) - ga, g),$$

so $(\delta_2 - \delta_1)(g) = ga - a \in B^1(G, A)$.

We thereby identify $H^1(G, A)$ with the A-conjugacy classes of splittings $0 \to A \to E \to G \to 1$. (We wrote 0 at the left because A is written in additive notation.)

Example 25.71. Continuing Example 25.70, let us consider when an arbitrary function $s\colon G \to E$ is a homomorphism. We assume the necessary condition $s(1) = 1$. Since

$$\pi s(gh) = gh = \pi s(g)s(h),$$

we conclude that $s(g)s(h)s(gh)^{-1} \in i(A)$; in other words, we have some function $f\colon G \times G \to A$ satisfying

$$s(g)s(h) = i(f(g, h))s(gh).$$

Note that $f(g, 1) = 0 = f(1, g)$ since $s(1) = 1$; such f is called **normalized**. Moreover, a computation parallel to that of Example 24.45 shows that associativity of multiplication in G is equivalent to the formula

$$f(g, h) + f(gh, k) = gf(h, k) + f(g, hk);$$

i.e., $f \in Z^2(G, A)$. From this point of view, we get a map taking elements of $Z^2(G, A)$ to the extensions of G by A.

We say two extensions $0 \to A \to E \to G \to 1$ and $0 \to A \to E' \to G \to 1$ are **equivalent** if there is an isomorphism $\varphi\colon E \to E'$ such that the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{i} & E & \xrightarrow{\pi} & G & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle 1_A} & & \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle 1_G} & & \\ 0 & \longrightarrow & A & \xrightarrow{i'} & E' & \xrightarrow{\pi'} & G & \longrightarrow & 1 \end{array} \tag{25.17}$$

commutes. Elements of $B^2(G, A)$ yield equivalent extensions, so we get a map from $H^2(G, A)$ to the equivalence classes of extensions of G by A, which is easily seen to be 1:1.

Group extensions with nonabelian kernel are described in terms of H^3 in [**Broks, IV.6**].

Cohomology of Lie algebras.

Example 25.69 (The Chevalley-Eilenberg complex). Taking a Lie algebra L (defined over a field F) instead of a group, we need to start with a projective resolution of F as a module over the enveloping algebra $U(L)$ of L. It is harder here than in the group case to construct free modules because of the skew-symmetry in Lie multiplication. One way is to work with the Grassman algebra $E(L)$. Since L acts on itself via the adjoint action, L also acts naturally on $E(L)$, via

$$a(a_1 \wedge \cdots \wedge a_n) = \sum_{i=1}^{n} a_1 \wedge \cdots \wedge [aa_i] \wedge \cdots \wedge a_n.$$

(We wrote the $\wedge$ to emphasize the wedge multiplication in $E(L)$.) Thus $E(L)$ is a $U(L)$-module. Writing $E(L)_n$ for the $U(L)$-submodule of $E(L)$ spanned by (wedge) products of length n, we define $V_n(L) = U(L) \otimes_F E(L)_n$, which by Remark 18.27 are free $U(L)$-modules; the $V_n(L)$ provide a free resolution of F with augmentation map $V_0(L) = U(L) \to F$ and boundary maps $d_n\colon V_n(L) \to V_{n-1}(L)$ given by

$$\begin{aligned} &d_n(r \otimes (a_1 \wedge \cdots \wedge a_n)) \\ &\quad = \sum_{i=1}^{n} (-1)^i r a_i \otimes (a_1 \wedge \cdots \wedge \hat{a}_i \wedge \cdots \wedge a_n) \\ &\qquad + \sum_{i<j}^{n} (-1)^{i+j} r \otimes ([a_i, a_j] \wedge a_1 \wedge \cdots \wedge \hat{a}_i \wedge \cdots \wedge \hat{a}_j \wedge \cdots \wedge a_n), \end{aligned}$$

where $\hat{a}_i$ indicates that a_i does not occur in the wedge product.

The verification requires some work; details can be found in [**Jac1, pp. 174–185**] or [**Wei, pp. 239–242**]. Then one gets the following analogous situation to Example 25.66:

For a Lie algebra L and Lie module M, $C^n(L, M)$ is the set of skew-symmetric n-linear mappings from L^n to M, where $\delta_n f(a_1, \ldots, a_n)$ is defined to be

$$\begin{aligned} &\sum_{i=1}^{n+1} (-1)^{n+1-i} a_i f(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{n+1}) \\ &\quad + \sum_{i,j} (-1)^{i+j+1} f(a_1, \ldots, a_{i-1}, a_{i+1} \ldots, a_{j-1}, a_{j+1}, \ldots, a_{n+1}, [a_i a_j]). \end{aligned}$$

Dimension 1 is interpreted in Exercise 71. Whitehead's lemmas (Exercises 21.44 and 21.47), and thus Levi's Theorem (Exercise 21.48), are really

results about the Lie cohomology groups; their proofs become much clearer when we bring in this extra structure, because we can rely on properties of the Casimir element (cf. Exercise 70).

The cohomological theories of groups and Lie algebras are quite similar, and several verifications can be united in the framework of Hopf algebras, to be discussed in Remark 26.26.

Example 25.70. Ironically, there is an extra complication in dealing with associative algebras. For homology and cohomology of algebras, we need to consider bimodules in order to use $\otimes$ and **Hom**; cf. Proposition 18.7 and Remark 25.4′, respectively. By Remark 18.45, we could view R as an $R \otimes_C R^{\mathrm{op}}$-module; this approach is studied in Appendix 25B, so we leave this application for Exercise 25B.5ff.

Appendix 25A: Morita's theory of categorical equivalence

As one might expect, projective modules play a key role in the categorical theory of modules, but even so, the concepts we have discussed above meld together in an amazingly sublime way, which we consider step by step. If we wish to study all the categorical properties of R-**Mod** for a given ring R, we should start by determining when the categories R-**Mod** and R'-**Mod** provide the same theories, i.e., are categorically equivalent; cf. Definition 1A.17 of Volume 1. Honoring Morita for his celebrated theorem, we say in this case that the two rings R and R' are **Morita equivalent**. We start with some examples.

Example 25A.1. (i) Suppose R and R' are fields. Since any vector space is just a direct sum of copies of the base field, we could correspond the R-modules with R'-modules, via $R^{(I)} \mapsto R'^{(I)}$. (In particular, $R \mapsto R'$.) But this is not a category equivalence unless the natural 1:1 correspondence also acts on morphisms; since $\mathrm{Hom}_R(R,R) \cong R$ and $\mathrm{Hom}_{R'}(R',R') \cong R'$, our categorical equivalence would have to identify R with R' (as sets). In fact, as the theory develops, we see by Exercise 25A.6 that two commutative rings are Morita equivalent iff they are isomorphic.

(ii) In Exercises A1 through A3, we see that the rings R and $M_n(R)$ always are Morita equivalent, via the categorical equivalence sending an R-module M to the $M_n(R)$-module $M^{(n)}$. It follows that any categorical property of rings (and modules) holding for R must hold automatically for $M_n(R)$, and conversely. Note that whereas "projective" is categorical, "free" is not; indeed, R itself as an R-module is sent to the non-free module $R^{(n)}$ over $M_n(R)$.

Some care is required in recasting familiar definitions in categorical terms. A module M is simple iff every morphism from M is monic or 0. Thus, simplicity of modules is categorical; however, cyclicity of modules is not categorical (cf. Exercise A4). Some subtler examples: f.g. and Noetherian are each categorical; cf. Exercise A5.

Since $M_n(R) \cong \mathrm{End}_R R^{(n)}$ by Proposition 13.42(ii), we see that the endomorphism ring of a f.g. free R-module is Morita equivalent to R. The goal of this discussion is to show that all rings Morita equivalent to R "almost" have this form. But our result should be expressed in categorical language, and "free" is not categorically defined. Toward this end we can try endomorphism rings of projective modules instead of free modules, but we still need another categorical property involving free modules, reflecting the fact that every R-module is a homomorphic image of a direct sum of copies of R.

Definition 25A.2. A **generator** of an Abelian category $\mathcal{C}$ is an object C such that for every nonzero morphism $A \xrightarrow{f} B$ there is a morphism $C \xrightarrow{h} A$ with $fh \neq 0$.

Here is a down-to-earth criterion for being a generator in R-**Mod**.

Remark 25A.3. If there is an epic $g\colon M^{(n)} \to R$, then M is a generator of R-**Mod**. (Indeed, write $g_i\colon M \to R$ for the restriction of g to the i component of $M^{(n)}$. Given $f\colon A \to B$ with $f(a) \neq 0$, define $h_i\colon M \to A$ by $h_i(y) = g_i(y)a$. We claim that $fh_i \neq 0$ for some i. Indeed, write $g(y_1,\dots,y_n) = 1$ for suitable $(y_1,\dots,y_n) \in M^{(n)}$; i.e., $\sum g_i(y_i) = 1$. Then

$$\sum fh_i(y_i) = \sum f(g_i(y_i)a) = f\left(\left(\sum g_i(y_i)\right)a\right) = f(a) \neq 0,$$

as desired.)

To prove the converse, we bring in some more notation. Given an R-module M, let $R' = (\mathrm{End}_R M)^{\mathrm{op}}$ and $M^* = \mathrm{Hom}_R(M,R)$.

Remark 25A.5. It was noted in Appendix 15A that M is an R,R'-bimodule, where $ar' = r'(a)$ for $a \in M$, $r' \in R'$. Likewise, M^* is an R',R-bimodule — M^* is a right R-module under the product $fr\colon M \to R$ given by

$$fr(a) = f(a)r, \tag{25A.1}$$

and a (left) R'-module under the product $r'f$ given by

$$r'f(a) = f(r'(a)). \tag{25A.2}$$

(In our provisional notation, f, g, f_i, g_i, and h denote elements of M^*.) Hence, $M \otimes_{R'} M^*$ is an R,R-bimodule.

Definition 25A.6. Define $T(M)$ to be $\{\sum_{\text{finite}} f_i(a_i) : f_i \in M^*,\ a_i \in M\}$.

$T(M)$ is called the **trace ideal**, in view of the following result.

LEMMA 25A.7. *There is an R, R-bimodule map*

$$\tau\colon M \otimes_{R'} M^* \to R$$

given by $a \otimes f \mapsto f(a)$; its image is $T(M)$, which thus is an ideal of R, and equals R iff M is a generator.

Proof. The map $(a, f) \mapsto f(a)$ is balanced in view of (25A.2), thereby yielding the desired map τ, which is an R, R-bimodule map by (25A.1). Hence $T(M)$ is an R, R sub-bimodule of R; i.e., $T(M) \lhd R$. $T(M) = R$ iff $1 \in T(M)$; i.e., there are $f_i \in M^*$ and $a_i \in R$ such that $\sum_{i=1}^n f_i(a_i) = 1$ (for some n). In this case, we have an onto map $M^{(n)} \to R$ given by $(a_1, \dots, a_n) \mapsto \sum f_i(a_i)$, which implies that M is a generator in view of Remark 25A.3.

On the other hand, if $T(M) \neq R$, then the natural map $\pi\colon R \mapsto R/T(M)$ satisfies $\pi h = 0$ for all $h\colon M \to R$, so M is not a generator. □

Remark 25A.7′. In the first paragraph of the proof, we actually showed that if $T(M) = R$, then R is a homomorphic image of the module $M^{(n)}$ for suitable $n \in \mathbb{N}$.

PROPOSITION 25A.8. *The following properties are equivalent for an R-module M:*

(i) *M is a generator in R-**Mod**.*
(ii) *$T(M) = R$.*
(iii) *R is a homomorphic image of $M^{(n)}$ for some n.*

Proof. We have already proved that (iii) ⇒ (i) ⇒ (ii) ⇒ (iii), in view of Remark 25A.7′. □

We also have the following analog of Lemma 25A.7.

LEMMA 25A.9. *There is an R', R'-bimodule map*

$$\tau'\colon M^* \otimes_R M \to R'$$

given by $f \otimes a \mapsto r'$, where $r'\colon M \to M$ is defined by $r'(b) = f(b)a$. τ' is onto iff M is projective as an R-module.

Proof. This time we define the balanced map $M^* \times M \to R'$ by $(f, a) \mapsto r'$, noting that

$$(fr)(b)a = f(b)ra$$

by (25A.1); this balanced map yields the desired map τ', an R', R'-bimodule map by (25A.2). To say that τ' is onto is to say that 1 is in the image of τ', an ideal of R'; this is precisely the assertion of the Dual Basis Lemma (Proposition 25.13). □

Inspired by Lemmas 25A.7 and 25A.9, we are ready for the main concept in this discussion.

Definition 25A.10. An R-**progenerator** is a f.g. projective R-module that is also a generator in R-**Mod**.

For example, any f.g. free module is an R-progenerator. On the other hand, if $R = R_1 \times R_2$ as rings, then R_1 is a direct summand of R as an R-module and thus is projective, but is not a generator.

Here is another nice criterion for a module to be a progenerator, which works particularly well in the commutative case.

Definition 25A.11. A projective module P over a ring R is called **faithfully projective** if $AP \neq P$ for every $A \lhd R$.

Remark 25A.12. (i) Any f.g. faithfully projective module P is a progenerator. (Indeed, by Equation (25.3) from the Dual Basis Lemma, we can write any element a of P as $\sum h_i(a)a_i \in T(P)P$, implying that $P = T(P)P$ so, by hypothesis, $T(P) = R$.)

(ii) Any faithful f.g. projective module P over a commutative ring C is faithfully projective, and thus is a progenerator. Indeed otherwise there is a maximal ideal $\mathfrak{p}$ of C with $\mathfrak{p}P = P$, implying by "Nakayama's Lemma" (Proposition 8.24 of Volume 1) that $P_{\mathfrak{p}} = 0$, contrary to P faithful.

Example 25A.13. Over any commutative ring C, a f.g. module M is a progenerator iff it is faithfully projective. (Indeed, (⇐) is by Remark 25A.12. Conversely, in view of Proposition 25A.8, if M is a progenerator, then C is a homomorphic image of $M^{(n)}$ for some n; hence, each localization $C_{\mathfrak{p}}$ is a homomorphic image of $M_{\mathfrak{p}}$, implying that $\operatorname{rank} M_{\mathfrak{p}} > 0$, so we are done by Remark 25.27.)

Having some feel for progenerators, let us see how they fit into the Morita theory.

PROPOSITION 25A.14. *If R and R' are Morita equivalent rings, then there is an R-progenerator P such that $R' \cong (\mathrm{End}_R P)^{\mathrm{op}}$.*

Proof. Take the category equivalence $F\colon R'\text{-}\mathbf{Mod} \to R\text{-}\mathbf{Mod}$. Since F sends progenerator to progenerator and R' is a progenerator of R'-**Mod**, we see that $P = FR'$ is a progenerator of R-**Mod**, and in view of Example 13.41,

$$R' \cong (\mathrm{End}_{R'} R')^{\mathrm{op}} \cong (\mathrm{End}_R(FR'))^{\mathrm{op}} = (\mathrm{End}_R P)^{\mathrm{op}}. \qquad \square$$

We recover Example 25A.1(ii) by taking $P = R^{(n)}$; then we can identify $R' \cong (\mathrm{End}_R R^{(n)})^{\mathrm{op}} \cong M_n(R)^{\mathrm{op}} \cong M_n(R)$, in view of Remark 13.21.

Our main objective is to prove the converse. Given an R-module M, we need a mechanism to demonstrate that R and $R' = (\mathrm{End}_R M)^{\mathrm{op}}$ are Morita equivalent when M is a progenerator. Let us fix notation; r and r_i always denote elements of R, whereas r' and r'_i always denote elements of R'. Also, a, b, a_i, and b_i denote elements of M, and f, g, f_i, and g_i denote elements of M'.

To highlight the inherent duality between R and R', we need a more formal definition, stated for arbitrary rings R and R'.

Definition 25A.15. A **Morita context** is a six-tuple $(R, R', M, M', \tau, \tau')$, where R, R' are rings, M is an R, R'-bimodule, M' is an R', R-bimodule, $\tau\colon M \otimes_{R'} M' \to R$ is an R, R-bimodule map, and $\tau'\colon M' \otimes_R M \to R'$ is an R', R'-bimodule map, such that the following two extra associative laws hold:

$$a\tau'(f \otimes b) = \tau(a \otimes f)b, \qquad f\tau(a \otimes g) = \tau'(f \otimes a)g \tag{25A.3}$$

for all $a, b \in M$ and $f, g \in M'$.

Write (b, g) for $\tau(b \otimes g)$ and $[g, b]$ for $\tau'(g \otimes b)$. Then Equations (25A.3) say that $a[f, b] = (a, f)b$ and $f(a, g) = [f, a]g$ for all $a, b \in M$ and $f, g \in M'$. See Exercise A8 for a more concise way of writing the Morita context.

Given a Morita context $(R, R', M, M', \tau, \tau')$, we also have the **dual Morita context** $(R', R, M', M, \tau', \tau)$, in which the roles of R and R' are reversed; any general result for Morita contexts also holds in this dual context. This observation, called **Morita duality**, enables us to shorten subsequent proofs.

Sometimes in the literature, the definition of Morita context requires τ and τ' to be onto, because of the following important information that comes as a consequence of these assumptions:

Remark 25A.16. Suppose $(R, R', M, M', \tau, \tau')$ is a Morita context.

(i) If τ is onto, then τ is an isomorphism. (Indeed, by hypothesis, $\sum_j (b_j, g_j) = 1$ for suitable $b_j \in M$ and $g_j \in M'$. It remains to show that $\ker \tau = 0$. Suppose $\sum a_i \otimes f_i \in \ker \tau$, i.e., $\sum (a_i, f_i) = 0$. Then

$$\begin{aligned}\sum a_i \otimes f_i &= \sum_i a_i \otimes f_i \sum_j (b_j, g_j) = \sum_{i,j} a_i \otimes f_i(b_j, g_j) \\ &= \sum_{i,j} a_i \otimes [f_i, b_j] g_j = \sum_{i,j} a_i [f_i, b_j] \otimes g_j \\ &= \sum_i (a_i, f_i) \sum_j b_j \otimes g_j = 0,\end{aligned}$$

proving that $\ker \tau = 0$.)

(ii) Applying Morita duality to (i) shows that when τ' is onto, τ' is an isomorphism.

LEMMA 25A.17. *Suppose $(R, R', M, M', \tau, \tau')$ is a Morita context, with τ' onto, and let $M^* = \mathrm{Hom}_R(M, R)$.*

(i) *There is an isomorphism $\Phi\colon M' \to M^* = \mathrm{Hom}_R(M, R)$ given by $f \mapsto (\ \ , f)$. Here $(\ \ , f)$ denotes the map $a \mapsto (a, f)$ in M^*.*

(ii) *$R' \cong \mathrm{End}_R M$ under the right regular representation $\rho\colon R' \to \mathrm{End}_R M$, defined by taking $\rho(r')$ to be right multiplication by r'; in other words, $\rho(r')(a) = ar'$.*

(iii) *M is a projective right R'-module.*

Proof. Take $\sum [f_j, a_j] = 1$ for $f_j \in M'$ and $a_j \in M$. Then, for any map $h\colon M \to N$ of R-modules and any $a \in M$, we have

$$h(a) = h\left(a \sum [f_j, a_j]\right) = h\left(\sum (a, f_j) a_j\right) = \sum (a, f_j) h(a_j). \tag{25A.4}$$

To prove (i), note that Φ is monic since $(\ \ , f) = 0$ implies

$$f = \sum [f_j, a_j] f = \sum f_j (a_j, f) = 0.$$

On the other hand, Φ is onto, since for any $h \in M'$ and $a \in M$, Equation (25A.4) yields

$$h(a) = \sum (a, f_j) h(a_j) = \left(a, \sum f_j h(a_j)\right),$$

implying $h = \left(\ \ , \sum f_j h(a_j)\right)$.

To prove (ii), first note that $\ker \rho = 0$, for if $\rho(r') = 0$, then

$$r' = \sum [f_j, a_j]r' = \sum [f_j, a_j r'] = 0.$$

But ρ also is onto since, for any $h \in \mathrm{End}_R M$, we apply Equation (25A.4) to get

$$h(a) = \sum (a, f_j)h(a_j) = \sum a[f_j, h(a_j)] = a \sum [f_j, h(a_j)],$$

implying $h = \rho\big(\sum [f_j, h(a_j)]\big)$.

(iii) is immediate from the Dual Basis Lemma (Proposition 25.13), taking $h_i = (\_\_, f_i)$, since

$$a = a \sum_i [f_i, a_i] = \sum (a, f_i)a_i = \sum h_i(a)a_i$$

for all $a \in M$. □

Remark 25A.17′. If τ is onto, then $R \cong \mathrm{End}\, M'_{R'}$, seen by applying Morita duality to Lemma 25A.17(ii).

Example 25A.18. Given an R-module M, we consider the Morita context $(R, R', M, M', \tau, \tau')$, where $R' = (\mathrm{End}_R M)^{\mathrm{op}}$ and $M' = M^*$. If M is an R-progenerator, then, by Lemmas 25A.7 and 25A.9 respectively, τ and τ' are onto and thus isomorphisms. Hence, $R \cong \mathrm{End}\, M'_{R'}$, by Remark 25A.17.

The more general definition of Morita context presented here permits other interesting examples such as Exercise A8 and Exercise 26.38.

Theorem 25A.19 (Morita's Theorem). *Two rings R, R' are Morita equivalent iff there is an R-progenerator M such that $R' \cong (\mathrm{End}_R M)^{\mathrm{op}}$; in this case the categorical equivalence R-**Mod** $\to$ R'-**Mod** is given by $M^* \otimes_R \_\_$, and the equivalence in the other direction is given by $M \otimes_{R'} \_\_$.*

Proof. ($\Rightarrow$) is by Proposition 25A.14, so we need only prove ($\Leftarrow$). We have the Morita context $(R, R', M, M^*, \tau, \tau')$ from Example 25A.18; τ is onto by Lemma 25A.7, and τ' is onto by Lemma 25A.9. Thus M and M' are Morita equivalent.

Thus, $M \otimes_{R'} M^* \cong R$ and, for any R-module N,

$$N \cong R \otimes_R N \cong (M \otimes_{R'} M^*) \otimes N \cong M \otimes_{R'} (M^* \otimes N),$$

showing that the composition of the functors $M \otimes_{R'} \_\_$ and $M^* \otimes_R \_\_$ is naturally isomorphic to the identity. By Morita duality, the composition in the other direction is also naturally isomorphic to the identity. □

The proof of Morita's Theorem takes on a life of its own, drawing us to many other fascinating properties.

Corollary 25A.19′. *Notation as in Theorem 25A.19, M is also a progenerator in* **Mod**-R'.

Proof. M is a projective R'-module by Lemma 25A.17(iii), and is a generator in R'-**Mod** by the dual argument of Lemma 25A.7. □

Also see Exercises A9 and A10. Also Morita's Theorem yields connections between Morita equivalent rings which at first blush would seem to have nothing to do with category theory; cf. Exercise A11.

Example 25A.20. Morita's Theorem "explains" the Wedderburn-Artin Theorem in categorical terms. For R a simple Artinian ring, all modules are projective. Furthermore, any simple module is isomorphic to a given minimal left ideal M, so all modules are direct sums of copies of M, implying that M is a progenerator. Letting $D = (\mathrm{End}_R M)^{\mathrm{op}}$, a division ring by Schur's Lemma, we see that R is Morita equivalent to D. Furthermore, there is a f.g. D-module M^* such that $R \cong (\mathrm{End}\, M^*_D)^{\mathrm{op}}$. Since $M^* \cong D^{(n)}$ for some n, we conclude that $R \cong M_n(D)$. Picking out the essential parts of this argument leads to the quick direct proof of the Wedderburn-Artin Theorem given in Exercise 14.6.

More generally, suppose R is a direct product $R_1 \times \cdots \times R_t$ of simple Artinian rings. Take L_i a minimal left ideal of R_i, and let $D_i = \mathrm{End}_R L_i$, a division ring. Then $P = L_1 \oplus \cdots \oplus L_t$ is a progenerator of R-**Mod**, and $\mathrm{End}_R P \cong D_1 \times \cdots \times D_t$ is Morita equivalent to R.

These results have encouraged mathematicians to study a given arbitrary ring in terms of the Morita equivalent ring with the "best" structure, but this project has stalled in general, because of examples like Exercise A7.

Appendix 25B: Separable algebras

The algebras used in the building blocks of algebraic geometry are often local rings and not fields. Azumaya succeeded in extending the theory of csa's to algebras over local (commutative) rings, generalized further by Auslander and Goldman to algebras over arbitrary commutative rings. The outcome is a lovely subject that subsumes separable field extensions, and unifies and explains many fundamental concepts used in algebra. Turning to Proposition 18.29′ for intuition, we consider an algebra R over a given commutative ring C. Preceding Remark 18.45, we defined $R^e = R \otimes_C R^{\mathrm{op}}$; this gives a new perspective to projectivity.

Definition 25B.1. A C-algebra R is **separable** if R is projective as an R^e-module.

Example 25B.2. Any finite separable field extension $K \supseteq F$ is a separable F-algebra in the sense of Definition 25B.1. Indeed, by Proposition 18.29′, the ring $K^e = K \otimes_F K$ is semisimple, so every module (in particular, K) is projective.

To understand Definition 25B.1, we consider the multiplication map $p\colon R^e \to R$ given by $a \otimes b \mapsto ab$. (One easily checks that p is an R^e-module map.) It is standard to let J denote $\ker p$.

PROPOSITION 25B.3. *J is generated as an R-module by*

$$\{r\otimes 1 - 1\otimes r \,:\, r \in R\}.$$

Proof. $p(r\otimes 1 - 1\otimes r) = r - r = 0$, proving that $r\otimes 1 - 1\otimes r \in J$. Conversely, if $y = \sum a_i \otimes b_i \in J$, then $\sum a_i b_i = p(y) = 0$, implying that

$$y = \sum a_i \otimes b_i - \sum a_i b_i \otimes 1 = -\sum a_i(b_i \otimes 1 - 1 \otimes b_i). \qquad \square$$

Remark 25B.4. By Proposition 25.10, R is separable iff p splits, i.e., iff there is a map $f\colon R \to R^e$ such that $pf = 1_R$. In this case, f is monic and $f(R) \cong R$ is a direct summand of R^e; in fact, $f(R) = f(R^e \cdot 1) = R^e f(1)$.

Definition 25B.5. For as in Remark 25B.4, the element $e = f(1) \in R^e$ is called the **separability idempotent** for R.

THEOREM 25B.6. *The separability idempotent e is indeed an idempotent, and also $Je = 0$; i.e.,*

$$(r \otimes 1)e = (1 \otimes r)e, \quad \forall r \in R. \tag{25B.1}$$

Conversely, if there exists an idempotent $e \in R^e$ satisfying $Je = 0$, then R is separable over C, and e is a separability idempotent of R.

Proof. For all $r \in R$,

$$(1\otimes r)e = (1\otimes r)f(1) = f(11r) = f(r11) = (r\otimes 1)f(1) = (r\otimes 1)e,$$

yielding (25B.1); $Je = 0$ by Proposition 25B.3.

Now writing $e = \sum a_i \otimes b_i$ we have $1 = p(e) = \sum a_i b_i$ and

$$\begin{aligned} e^2 &= \Big(\sum a_i \otimes b_i\Big)\, e = \sum (a_i \otimes 1)(1 \otimes b_i)e \\ &= \sum (a_i \otimes 1)(b_i \otimes 1)e = \Big(\sum a_i b_i \otimes 1\Big)\, e = (1 \otimes 1)e = e. \end{aligned} \qquad \square$$

Remark 25B.7. When we write the separability idempotent $e = \sum a_i \otimes b_i$ for $a_i, b_i \in R$, the conditions of Theorem 25B.6 translate to:

$$\sum a_i b_i = 1; \qquad \sum_i r a_i \otimes b_i = \sum_i a_i \otimes b_i r, \quad \forall r \in R.$$

Conversely, any element e satisfying these conditions is a separability idempotent by the same verifications as given in Theorem 25B.6.

Formulated in this way, the conditions are quite easy to verify. Here are some applications of the separability idempotent.

PROPOSITION 25B.8.

(i) *Any homomorphic image R/B of a separable algebra R is separable over $(C + B)/B \cong C/(C \cap B)$, the separability idempotent being the homomorphic image of e.*
(ii) *If R is separable over C and H is any commutative C-algebra, then $R \otimes_C H$ is separable over H.*
(iii) *If R is an H-algebra and separable over C, where H is a commutative C-algebra, then R is separable over H.*
(iv) *If R_i are separable algebras over commutative C-algebras C_i, then $R_1 \times R_2$ is separable over $C_1 \times C_2$, and $R_1 \otimes_C R_2$ is separable over $C_1 \otimes_C C_2$.*

Proof. (i) The image of e is clearly a separability idempotent, in view of Remark 25B.7.

(ii) $e \otimes 1$ is a separability idempotent.

(iii) The separability idempotent e of R over C is also a separability idempotent over H. (Alternatively, one could combine (i) and (ii).)

(iv) The pair (e_1, e_2) and the tensor product $e_1 \otimes e_2$ of separability idempotents respectively provide the required separability idempotent. $\square$

The last assertion smooths over a rough spot that we encountered in the theory of csa's; although $M_n(\mathbb{C})$ does not have a subfield of dimension n over its center $\mathbb{C}$, it does have a maximal commutative separable subalgebra $\mathbb{C}^{(n)}$ of dimension n. In fact, any maximal commutative separable subalgebra of a csa of degree n has dimension n over the center; cf. Exercise B18.

The separability idempotent also gives us a generalization of Maschke's Theorem, pointed out to me by Saltman:

THEOREM 25B.9. *Suppose P is a module over a separable C-algebra R. If P is projective as a C-module, then P is projective as an R-module.*

Proof. Take a separability idempotent $e = \sum a_i \otimes b_i$ of R. In view of Proposition 25.10(4), we need to split any epic $\pi: M \to P$ of R-modules. By hypothesis, there is some C-module map $g: P \to M$ such that $\pi g = 1_M$; we want to modify g by means of an averaging process to get an R-module map.

For any $w \in P$, there is a balanced map $\psi_w: R \times R \to M$ given by $\psi_w(a,b) = ag(bw)$, which thus induces a C-module map $\bar{\psi}_w: R \otimes_C R \to M$ given by $a \otimes b \mapsto ag(bw)$. By Remark 25B.7,

$$\sum a_i g(b_i r w) = \bar{\psi}_w\Big(\sum a_i \otimes b_i r\Big) = \bar{\psi}_w\Big(\sum r a_i \otimes b_i\Big) = \sum r a_i g(b_i w).$$

Consequently, defining $\bar{g}: P \to M$ by

$$\bar{g}(w) = \sum a_i g(b_i w), \quad w \in P,$$

we have

$$\bar{g}(rw) = \sum a_i g(b_i r w) = \sum r a_i g(b_i w) = r\bar{g}(w),$$

implying that $\bar{g}$ is an R-module map. Moreover $\pi\bar{g}(w) = \sum a_i \pi g(b_i w) = \sum a_i b_i w = w$, proving that $\bar{g}$ splits π. □

Corollary 25B.10. *If R is separable over a field F, then R is separable in the classical sense; i.e., R is semisimple and $R \otimes_F \bar{F}$ is semisimple where $\bar{F}$ is the algebraic closure of F.*

Proof. Every R-module is projective as an F-module, and thus is projective as an R-module. Hence, R is semisimple. Likewise, by Proposition 25B.8(ii) $R \otimes_F \bar{F}$ is separable and thus semisimple. □

Remark 25B.10′. For R as in Corollary 25B.10, the center of R is the direct sum of the centers of its simple components, and is a field iff R itself is simple.

We also have considerable control over the center of a separable algebra R and its homomorphic images.

Definition 25B.11. For any R, R-bimodule M, M^R denotes the centralizer of R in M, defined as $\{a \in M : ra = ra, \forall r \in R\}$.

Thus $R^R = \mathrm{Cent}(R)$.

Remark 25B.12. The following conditions are equivalent for $a \in M$, viewing the R, R-bimodule M also as an R^e-module:

(i) $a \in M^R$.

(ii) $(r \otimes 1 - 1 \otimes r)a = 0$ for all r in R.

(iii) $Ja = 0$ (by Proposition 25B.3).

Remark 25B.13. Suppose R is separable over C, and $Z = \mathrm{Cent}(R)$. Then, notation as in Theorem 25B.6:

(i) $(R^e)^R$ is a right ideal of R^e by Remark 25B.12.

(ii) $e \in (R^e)^R$ by Equation (25B.1), implying that $eR^e \subseteq (R^e)^R$ by (i).

(iii) $p(eR^e) = Z$. Indeed, ($\subseteq$) is by (ii), and ($\supseteq$) holds because $z = p((z \otimes 1)e) = p(e(z \otimes 1)))$ for any $z \in Z$.

(iv) Z is a summand of R as a C-module, since $r \mapsto p(e(r \otimes 1))$ defines a projection $R \to Z$, in view of (iii).

(v) R is separable over C and thus over Z; by (iv), Z is a summand of R as a Z-module. If $A \lhd Z$, then $A = Z \cap AR$, seen by matching components.

Proposition 25B.14. *If R is separable with center Z and $B \lhd R$, then*

$$\mathrm{Cent}(R/B) = (Z + B)/B \cong Z/(Z \cap B).$$

Proof. Write $\overline{R} = R/B$. For any $\bar{z} \in \mathrm{Cent}(\overline{R})$, by Proposition 25B.8(i) and Remark 25B.13(iii),

$$\bar{z} \in p(\bar{e}\overline{R}^e) = \overline{p(eR^e)} = \bar{Z}. \qquad \square$$

Proposition 25B.15. *If R is separable over its center C, then any maximal ideal B of R has the form AR, where $A = B \cap C \lhd C$, and R/AR is central simple over the field C/A.*

Proof. By Proposition 25B.8, R/B is separable over C/A, which is its center by Proposition 25B.14. But R/B is simple, so C/A is a field. Likewise, R/AR is separable over the field $C/A = \mathrm{Cent}(R/AR)$. By Corollary 25B.10, R/AR is semisimple; hence, R/AR is simple by Remark 25B.10′, implying that $AR = B$. □

Separability also is studied by means of derivations; cf. Exercises 8–12.

Azumaya algebras

We are finally ready to introduce the desired theory generalizing csa's.

Definition 25B.16. R is an **Azumaya** algebra (over C) if $C = \mathrm{Cent}(R)$ and R is separable as a C-algebra.

Thus, every csa is Azumaya, but this definition does not relate well to the language of Chapter 24. The key is Remark 24.13, which introduces the important map $\Phi: R^e \to \mathrm{End}_C R$, which for csa's is an isomorphism by Theorem 24.14. The parallel isomorphism for Azumaya algebras follows from Lemma 25A.17(ii), when we utilize the power of the Morita theory.

THEOREM 25B.17. *For any C-algebra R, the following assertions are equivalent:*

(i) *R is Azumaya over C.*
(ii) *In the natural Morita context $(R^e, C, R, \mathrm{Hom}_{R^e}(R, R^e), \tau, \tau')$, τ and τ' are onto.*
(iii) *$\Phi\colon R^e \to \mathrm{End}_C R$ is an isomorphism, and R is f.g. faithful projective as a C-module.*
(iv) *Φ is an isomorphism, and R is a progenerator as an R^e-module.*
(v) *Φ is an isomorphism, and R is a progenerator as a C-module.*
(vi) *C-Mod and R^e-Mod are equivalent categories (via the categorical equivalence given by $R \otimes_C \_\_$).*

Proof. (i) $\Rightarrow$ (ii) $\mathrm{Hom}_{R^e}(R, R)$ is identified with $R^R = C$, so one gets the Morita context $(R^e, C, R, \mathrm{Hom}_{R^e}(R, R^e), \tau, \tau')$. To show that τ and τ' are onto, we show that R is an R^e-progenerator, or, by Remark 25A.12, that R is faithfully projective. Assume on the contrary that $BR = R$ for $B \lhd R^e$; we may assume that B is a maximal ideal. But R and thus R^{op} are separable with center C, so R^e is separable with center C; hence $B = AR^e$ for some maximal $A \lhd C$, by Proposition 25B.15. Thus, $AR = A(R^eR) = BR = R$, contradicting Remark 25A.12(ii).

(ii) $\Rightarrow$ (iii) Remark 25A.17′ provides the isomorphism $R^e \to \mathrm{End}_C R$. Furthermore, Proposition 25A.19′ says R is a progenerator as a C-module (actually as a right C-module, but this is irrelevant since C is commutative); we are done by Example 25A.13.

(ii),(iii) $\Rightarrow$ (iv) R is an R^e-progenerator by Lemmas 25A.7 and 25A.9.

(iv) $\Rightarrow$ (ii) is clear from Morita's Theorem.

(iii) $\Rightarrow$ (v) is by Remark 25A.12.

(ii) $\Rightarrow$ (vi) is by Morita's Theorem 25A.19.

(vi) $\Rightarrow$ (i) $R = R \otimes_C C$ is a progenerator of R^e-**Mod**. It remains to identify C with $\mathrm{Cent}(R)$. Morita's Theorem 25A.19 provides a category equivalence from R^e-**Mod** to C–**Mod**, given by the functor $\mathrm{Hom}_{R^e}(R, R^e) \otimes_{R^e} \_\_$, which by the adjoint isomorphism (Proposition 18.44) can be replaced by $\mathrm{Hom}_{R^e}(R, \_\_)$. Hence, $C \cong \mathrm{Hom}_{R^e}(R, R) = \mathrm{Cent}(R)$.

(v) $\Rightarrow$ (i) We build a Morita context $(C, \mathrm{End}_C R, R, R^*, \tau, \tau')$ with τ, τ' onto; since $\mathrm{End}_C R \cong R^e$ we can repeat the argument of the previous paragraph. □

Example 25B.18. Since local rings are so important in algebraic geometry, it is natural to inquire about the structure of an Azumaya algebra R over a local ring C. (This was the case studied by Azumaya.) R is a free module over C, by Example 25.15. Let P be the maximal ideal of C.

By Proposition 25B.15, PR is the unique maximal ideal of R, and, by Proposition 25B.14, $\overline{R} = R/PR$ is a csa of some degree n over the field $F = C/P$, i.e., $[\overline{R} : F] = n^2$. Using Nakayama's Lemma (Proposition 8.24 of Volume 1), we conclude that R has rank n^2 over C. We have proved that $(\deg \overline{R})^2 = \mathrm{rank}(R)$.

The tensor product of Azumaya algebras is Azumaya, in view of Proposition 25B.8; using Morita equivalence, we obtain a group generalizing Definition 24.16.

Remark 25B.19. The Morita equivalence classes of Azumaya algebras over C form a group $\mathrm{Br}(C)$ called the **Brauer group** of C; again we have $[R]^{-1} = [R]^{\mathrm{op}}$ since $R \otimes_C R^{\mathrm{op}} \cong \mathrm{End}_C R \sim C$.

Br can be viewed as a functor from the category of commutative rings to the category of groups in the following manner: Given a ring homomorphism $\psi\colon C_1 \to C_2$, we defines $\mathrm{Br}(\psi)\colon \mathrm{Br}(C_1) \to \mathrm{Br}(C_2)$ by sending an Azumaya C_1-algebra A to $A \otimes_{C_1} C_2$, viewing C_2 as a C_1-module via ψ.

In particular, if F is the field of fractions of an integral domain C, we have a (not necessarily 1:1) map $\mathrm{Br}(C) \to \mathrm{Br}(F)$, which is of considerable interest.

As with csa's, the group properties of $\mathrm{Br}(C)$ fuel the theory of Azumaya algebras, but this approach is beyond the scope of this book.

The Artin-Procesi Theorem.

Theorem 25B.17(iii) shows that any Azumaya algebra satisfies a polynomial identity, and the similarity between Proposition 25B.15 and Theorem 23.39 suggests a much deeper connection with PI-theory.

THEOREM 25B.20. *The following conditions are equivalent for any C-algebra R:*

(i) *R is Azumaya over C of rank n^2.*
(ii) *R satisfies all identities of $M_n(\mathbb{Z})$, and every simple homomorphic image of R has degree n (over its center).*
(iii) *R satisfies all identities of $M_n(\mathbb{Z})$, and no homomorphic image of R satisfies the standard identity s_{2n-2}.*
(iv) *R satisfies all identities of $M_n(\mathbb{Z})$, and no homomorphic image of R satisfies the identity h_n, where h_n is defined in Equation (23.4).*
(v) *R satisfies all identities of $M_n(\mathbb{Z})$, and the additive subgroup $\mathfrak{a}$ of C generated by the values of h_n contains* 1.

Proof. (i) $\Rightarrow$ (ii) By Proposition 25B.15, every simple homomorphic image has the form R/PR, where P is a maximal ideal of C. But every element

of $S = C \setminus P$ is invertible modulo P in the field C/P, so in view of the universal property of localization, there is an algebra surjection $R_P \to R/PR$ implying, by Example 25B.18, that the C/P-csa R/PR has degree n.

It remains to prove the first assertion. Since R can be embedded into the product of its localizations at prime ideals of C, one may assume C is local with maximal ideal P. Then PR is a maximal ideal of R. Let $\overline{R} = R/PR$, a csa of degree n. By Proposition 24.53, there are elements $r_j, a \in R$ such that $\{\bar{a}^{i-1}\bar{r}_j\}$ are a base of $\overline{R}$ over $\bar{C}$, so $\{a^{i-1}r_j\}$ span R over C by Nakayama's Lemma, implying that $r_1, \dots, r_n$ span R as a right module over the commutative ring $C[a]$. Hence the left regular representation embeds R into $\mathrm{End}_{C[a]} R$, which satisfies all identities of $n \times n$ matrices. (This is an easy consequence of Exercise 13.20.)

(ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) Obvious, since h_n is central for $M_n(\mathbb{Z})$ and thus an identity of $M_{n-1}(\mathbb{Z})$.

(iv) $\Rightarrow$ (v) $\mathfrak{a} \lhd C$ since h_n is multilinear, and we are done if $1 \in \mathfrak{a}$. So we assume on the contrary that $\mathfrak{a}$ is contained in a maximal ideal $\mathfrak{p}$ of C. Localizing at $S = C \setminus \mathfrak{p}$, we may assume that C is local with maximal ideal $\mathfrak{p}$. Take any maximal ideal P of R. Then $\overline{R} = R/P$ is simple, of degree n (by hypothesis), and, as above, $\bar{\mathfrak{a}}$ is an ideal of the field $\mathrm{Cent}(\overline{R})$. Hence $1 \in \bar{\mathfrak{a}}$; i.e., $1 - a \in P \cap C \subseteq \mathfrak{p}$ for suitable $a \in \mathfrak{a}$. Hence $1 \in \mathfrak{p}$, a contradiction.

(v) $\Rightarrow$ (i) This would follow from Braun's criterion (Exercise B14) if h_n took on the value 1, but here is a direct proof due to Schelter that works in general. We rely heavily on Lemma 23.38. Write

$$1 = \sum_{i=1}^{k} h_n(r_{i1}, \dots, r_{im}).$$

Define $f_{ij}\colon R \to C$ by

$$f_{ij}(r) = (-1)^{j+1} h_n(r, r_{i1}, \dots, r_{i,j-1}, r_{i,j+1}, \dots, r_{im})$$

for $1 \le i \le k$ and $1 \le j \le n^2$, and note that

$$r = \sum h_n(r_{i1}, \dots, r_{im}) r = \sum_i \sum_j f_{ij}(r) r_{ij};$$

hence R is projective over C by the Dual Basis Lemma.

We also use h_n to prove that $\Phi\colon R^e \to \mathrm{End}_C R$ is an isomorphism. Namely, write $h_n = \sum h_{u1} X_1 h_{u2}$ for suitable polynomials h_{u1}, h_{u2}. Let

$$a_{iju} = (-1)^{j+1} h_{u1}(r, r_{i1}, \dots, r_{i,j-1}, r_{i,j+1}, \dots, r_{im}),$$

$$b_{iju} = h_{u2}(r, r_{i1}, \dots, r_{i,j-1}, r_{i,j+1}, \dots, r_{im}).$$

For any $f \in \mathrm{End}_C R$ and $r \in R$, one has

$$\begin{aligned} f(r) = f(1r) &= f\Big(\sum_{i=1}^{k} h_n(r_{i1}, \dots, r_{im}) r\Big) \\ &= f\Big(\sum_{i,j} (-1)^{j+1} h_n(r, r_{i1}, \dots, r_{i,j-1}, r_{i,j+1}, \dots r_{im}) r_{ij}\Big) \\ &= \sum_{i,j,u} a_{iju} r b_{iju} f(r_{ij}), \end{aligned}$$

implying that $f = \Phi(\sum_{i,j,u} a_{iju} \otimes b_{iju} f(r_{ij}))$, so Φ is onto. Φ is seen to be 1:1 by a similar computation: Suppose $\sum_q r_{q1} \otimes r_{q2} \in \ker \Phi$. Then for all $r_1, \dots, r_m$ in R, we have

$$\begin{aligned} h_n(r_1, \dots, r_m) \sum_q r_{q1} \otimes r_{q2} &= \sum_q h_n(r_1, \dots, r_m) r_{q1} \otimes r_{q2} \\ &= \sum_q \sum_{i=1}^{n^2} (-1)^{i-1} h_n(r_{q1}, r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_m) r_i \otimes r_{q2} \\ &= \sum_q \sum_i (-1)^{i-1} r_i \otimes h_n(r_{q1}, r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_m) r_{q2} \\ &= \sum_i r_i \otimes \sum_{u,q} a_u r_{q1} b_u r_{q2} \\ &= \sum_i r_i \otimes \Big(\sum_u a_u \sum_q r_{q1} b_u r_{q2}\Big), \end{aligned}$$

for suitable a_u, b_u which are sums of values of the h_{u1} and h_{u2}. But the last term is 0 by hypothesis, so $\sum_q r_{q1} \otimes r_{q2} \in \mathrm{Ann}(h_n(R)R) = 0$. □

(The equivalence (i) $\Leftrightarrow$ (ii), due to M. Artin and Procesi, amazingly predates the discovery of central polynomials.)

COROLLARY 25B.27. *Suppose R is any PI-algebra of PI-class n, and take any $0 \ne c \in h_n(R)$. Let $C = \mathrm{Cent}(R)$ and $S = \{c^i : i \in \mathbb{N}\}$. Then $S^{-1}R$ is Azumaya and a free module over $S^{-1}C$ of rank n^2, in view of Theorem 23.39.*

Appendix 25C: Finite-dimensional algebras revisited

Having studied projective modules, we return to the representation theory of f.d. algebras. Unfortunately, it still is difficult to pin down a more precise description of a f.d. algebra than Wedderburn's Principle Theorem (Theorem 15.26), since the interplay of the semisimple part and the radical is very

intricate. Accordingly, we need a more sophisticated approach to the representation theory of f.d. algebras, which was already hinted in Chapter 16 with the definition of finite representation type (f.r.t.) in Definition 16.41. The theory relies heavily on some of the techniques developed in the last few chapters, so we take the opportunity here to give a brief indication of some fundamental research in this important topic over the last forty years. We refer to a few technical (but easy) module-theoretic results given in Part IV and its exercises.

As noted in Remark 16.37 as well as in the introduction to Part V, the representations of f.d. algebras correspond to their modules. In view of Lemma 16.40, the representation theory can be viewed as an attempt to determine all indecomposable modules up to isomorphism.

Recall from Example 25.14 that, for any idempotent e, the left ideal Re is a direct summand of R, and thus is projective as an R-module. This leads us to start our investigation with idempotents. We say that an idempotent e is **primitive** if e cannot be written as the sum of two nontrivial orthogonal idempotents.

Remark 25C.1. An idempotent e is primitive iff the module Re is indecomposable, in view of Proposition 13.5.

Remark 25C.2. Any 1-sum set $\{e_1, \dots, e_t\}$ of orthogonal primitive idempotents yields a direct sum decomposition $R = Re_1 \oplus \cdots \oplus Re_t$ of R into indecomposable (projective) modules. Conversely, if $R = M_1 \oplus \cdots \oplus M_t$ where M_i is indecomposable, then $M_i = Re_i$ for suitable primitive idempotents $e_1, \dots, e_t$, thus providing a 1-sum set of orthogonal primitive idempotents.

This decomposition is unique up to isomorphism (and permutation) of summands, by the Krull-Schmidt Theorem; cf. Exercise 16.29.

This observation enhances the role of idempotents and, in particular, the Peirce decomposition of Exercise 13.9, which we recall is the direct sum decomposition

$$R = \bigoplus_{i=1}^{t} e_i R e_j. \tag{25C.1}$$

By Exercise 13.12, e_iRe_j is isomorphic to $\mathrm{Hom}(Re_i, Re_j)$.

In the case when R is a f.d. algebra over a field F, each Peirce component e_iRe_j is an F-subspace of R, and e_iRe_i is an F-algebras with unit element e_i.

We can cut down on t by means of the equivalence relation on $\{1, \dots, t\}$ defined by $i \sim j$ when $Re_i \cong Re_j$. We call a subset of $\{e_1, \dots, e_t\}$ **basic** if it contains exactly one index from each equivalence class, and then we call their sum a **basic idempotent**. For example, e_{11} would be a basic idempotent of $R = M_n(F)$, since each $Re_{ii} \cong Re_{11}$.

Definition 25C.3. The **basic subring** of R is eRe (with multiplicative unit e), where e is a basic idempotent. A ring R is **basic** if its basic idempotent e is 1.

Remark 25C.3′. By the Krull-Schmidt Theorem (Exercise 16.29), the basic set of idempotents is unique up to permutation and isomorphism of the Re_i. In view of Proposition 13.43, the basic subring eRe is isomorphic to $(\mathrm{End}_R Re)^{\mathrm{op}}$, and thus is unique up to isomorphism. Likewise, since e is the multiplicative unit of eRe, the basic subring of any f.d. algebra is basic.

The basic subring of R clearly has the same indecomposable modules (up to isomorphism) as R and, in fact, is Morita equivalent to R by an easy instance of Morita's Theorem (Theorem 25A.19), since Re is a progenerator and $eRe \cong (\mathrm{End}_R Re)^{op}$.

Example 25C.4. If a basic ring R is semisimple, then, by Theorem 13.47, $R \cong \mathrm{End}_R(\oplus_{i=1}^t L_i) \cong \prod_{i=1}^t D_i$, a direct product of division rings, where $D_i = \mathrm{End}_R L_i$.

Remark 25C.5. In general, let $J = \mathrm{Jac}(R)$ and $\overline{R} = R/J$.

(i) Any map $f\colon Re_i \to Re_j$ induces a map $\overline{f}\colon \overline{Re_i} \to \overline{Re_j}$, given by $\overline{f}(\overline{re_i}) = \overline{f(re_i)}$. If f is an isomorphism, then $\overline{f^{-1}} = \overline{f}^{-1}$, implying that $\overline{f}$ is also an isomorphism.

(ii) Conversely, in view of Exercise 13.11, any isomorphism $\overline{Re_i} \to \overline{Re_j}$ is given by right multiplication by $\bar{a}$ for some element $a \in e_iRe_j$; the inverse map is given by right multiplication by some element $\bar{b} \in \overline{e_jRe_i}$. Thus, $\overline{ab}$ is invertible in $\overline{e_iRe_i} \cong e_iRe_i/e_iJe_i$, implying by Exercise 15.30 that ab is invertible in e_iRe_i. Hence right multiplication by a gives an isomorphism from Re_i to Re_j. We are interested in the contrapositive: If Re_i and Re_j are not isomorphic, then $\overline{Re_i}$ and $\overline{Re_j}$ are not isomorphic.

(iii) It follows from (i) and (ii) that if a f.d. algebra R is basic, then R/J also is basic and thus a direct product of division rings, by Example 25C.4. Any map $f\colon Re_i \to Re_j$ induces a map $\overline{R}\bar{e}_i \to \overline{R}\bar{e}_j$, which must be 0; thus $f(Re_i) \subseteq Je_j$, and we conclude as in Exercise 13.12 that $e_iRe_j = e_iJe_j$.

We focus on the F-space $e_iJe_j/e_iJ^2e_j$, which we write as $e_i(J/J^2)e_j$.

Definition 25C.6. Suppose a f.d. algebra R is basic with $J = \mathrm{Jac}(R)$, and $e_1, \dots, e_t$ is a 1-sum set of primitive idempotents of R. We define its

quiver to be the directed graph $\Gamma(R)$ with vertices $\gamma_0 = \{1, \dots, t\}$, with each vertex i corresponding to the idempotent e_i of R, where there is an edge (called an **arrow**) from i to j of **weight** $m_{ij} = [e_i(J/J^2)e_j : F]$ whenever $e_i(J/J^2)e_j \neq 0$. It is convenient for us to write an arrow of weight m_{ij} explicitly as m_{ij} distinct edges $\ell_{ij1}, \dots, \ell_{ijm_{ij}}$ from e_i to e_j.

In view of the Krull-Schmidt Theorem (Exercise 16.29), the quiver is independent of the particular choice of $\{e_1, \dots, e_t\}$ (which are actually unique up to conjugation by an element of R, in view of Exercise 16.30). The quiver is one of the fundamental tools in the representation of algebras.

We can go in the other direction, obtaining an algebra from any directed graph. (From this point of view, the material ties in to Chapter 17.)

Definition 25C.7. Given any quiver Γ, one can define the **path monoid** $\hat{\Gamma}$ to be the monoid whose generators are elements corresponding to the vertices and edges of Γ, modulo the following relations, letting $\hat{e}_i$ denote the element corresponding to the vertex i, and letting ℓ denote a typical arrow (identified with an element of $\hat{\Gamma}$):

The $\hat{e}_i$ are orthogonal idempotents; i.e., $\hat{e}_i\hat{e}_k = \delta_{i,k}\hat{e}_i$ for all i, k;

$\hat{e}_i\ell = 0$ unless i is an initial vertex of ℓ, in which case $\hat{e}_i\ell = \ell$;

$\ell\hat{e}_j = 0$ unless j is a terminal vertex of ℓ, in which case $\ell\hat{e}_j = \ell$.

Intuitively, the elements of the path monoid $\hat{\Gamma}$ correspond to paths in Γ, and the product in the path monoid corresponds to the composition of paths.

The **path algebra** $F[\Gamma]$ is the monoid algebra (cf. Exercise 17.1) of the path monoid $\hat{\Gamma}$. We write $\mathcal{P}(R)$ for the path algebra of the quiver of R.

By Exercise 17.3, the path algebra, like any other monoid algebra, is an example of a monomial algebra.

Proposition 25C.8. *If R is any basic f.d. algebra with $J^2 = 0$, then R is a homomorphic image of $\mathcal{P}(R)$.*

Proof. Let V_{ij} denote the subspace of $\mathcal{P}(R)$ spanned by the edges from i to j. Note that $V_{ij} \cong e_i(J/J^2)e_j$, since by definition they have the same dimension.

Define a vector space map $\varphi\colon \mathcal{P}(R) \to R$ by sending the element $\hat{e}_i$ corresponding to the vertex i of Γ to the idempotent e_i of R; this yields a vector space isomorphism $\varphi\colon V_{ij} \to e_iJe_j$. Extending φ via multiplication, we see that φ preserves the defining relations. Thus φ is an algebra homomorphism and is onto by Proposition 15.25′ (which says that R is generated by the images of the $\varphi(e_i)$ and $\varphi(V_{ij})$). □

Corollary 25C.9. *Every f.d. algebra R with $J^2 = 0$ is Morita equivalent to a homomorphic image of the path algebra of the quiver of the basic subring of R.*

Remark 25C.10. Of course, the homomorphism given in Proposition 25C.8 need not be 1:1. The path algebra itself need not be finite-dimensional; in fact, $\mathcal{P}(R)$ is finite-dimensional iff it has no cycles. Nevertheless, writing $R \cong \mathcal{P}(R)/I$ for suitable $I \triangleleft \mathcal{P}(R)$, we know that the R-modules are precisely the $\mathcal{P}(R)$-modules annihilated by I, so any indecomposable module for R is also an indecomposable module for $\mathcal{P}(R)$.

The first problem in the representation theory of R would be to determine when there are only finitely many indecomposables; this is the case when R has f.r.t. Indeed, the renaissance of the representation theory of f.d. algebras was brought about by the following remarkable theorem of Gabriel [**Ga2**].

Theorem 25C.11 (Gabriel). *Suppose R is a f.d. algebra over an algebraically closed field and $J^2 = 0$. Then R has f.r.t. iff its quiver (viewed as an undirected graph) is a disjoint union of Dynkin diagrams of types A_n, D_n, E_6, E_7, or E_8.*

This theorem is an instance in which the second proof, given in Bernstein, Gel′fand, and Ponomarev [**BernGP**], was so elegant that it completely revolutionized the subject. Here is the essence of the proof of one direction ($\Leftarrow$), actually due to Tits.

STEP I. The f.g. modules of the path algebra $F[\Gamma]$ correspond to the representations of the graph Γ in the sense of Remark 22.27. Indeed, any module M restricts to vector spaces e_iM together with the corresponding linear transformations between the vector spaces attached to the initial and terminal vertices of the edges.

STEP II. By Theorem 22.28 and Proposition 22.25, if $\mathcal{P}(R)$ has f.r.t. (in the sense of having only a finite number of indecomposable representations, up to isomorphism), then Γ is a disjoint union of Dynkin diagrams of the desired form.

STEP III. Applying Remark 25C.10 to Step II, we see that if R has f.r.t., then its quiver is a disjoint union of Dynkin diagrams of the desired form.

In the other direction, one can pass from the algebra R to the hereditary algebra $\left(\begin{smallmatrix} R/J & J \\ 0 & R/J \end{smallmatrix}\right)$ and then to the path algebra. Bernstein, Gel′fand, and Ponomarev defined **reflection functors** that enable one to generate indecomposable representations in an analogous way to the generation of all positive roots of a f.d. Lie algebra from the simple roots.

The argument sketched above is so elegant that it led researchers to study path algebras of arbitrary quivers and their indecomposable representations, via the **Tits quadratic form** (22.7). For example, the quiver of the free associative algebra on n indeterminates is a bouquet of n circles. We call its category of representations $\mathcal{C}$, and say that an arbitrary algebra R has **wild representation type** if $\mathcal{C}$ can be embedded into the category of representations of R; otherwise R has **tame representation type**.

Donovon-Freislich [**DonF**] and Nazarova [**Naz**] extended Gabriel's theorem to show that $\mathcal{P}(R)$ has tame (but not finite) representation type, iff its quiver is one of the extended Dynkin diagrams $\tilde{A}_m, \tilde{D}_m, \tilde{E}_6, \tilde{E}_7$, or $\tilde{E}_8$ that arise in the classification of affine Lie algebras.

Dimension vectors are defined to be the positive roots of the Tits quadratic form, where the m_{ij} are as in Definition 25C.6. Kac [**Kacv2**], [**Kacv3**] showed in general that the dimension vectors of a representation of a quiver correspond to the positive roots of the Cartan matrix of the Tits quadratic form. Thus, even in wild type, the representation theory can be described in terms of Lie algebras of infinite Gel′fand-Kirillov dimension. Details can also be found in the article by Kraft and Riedtmann [**KrafR**].

Categories play a fundamental role in the theory of representations of algebras, as indicated in Remark 22.27. Auslander picked up on this, noticing that Definition 1A.17 of Volume 1 is applicable in a more general setting, cf. Exercise C3, which yields the "correct" approach to the theory.

Explicit representations of finite-dimensional algebras.

Sometimes one wants a more explicit way of describing representations of algebras over an algebraically closed field F. The approach of Theorem 25C.11 is based on module theory and Morita equivalence. But Morita-equivalent algebras differ greatly in certain aspects of their combinatorics, such as the polynomial identities that they satisfy; the field F satisfies the PI $[x_1, x_2]$, whereas the minimal PI of the Morita-equivalent algebra $M_n(F)$ is the standard identity s_{2n}. In order to determine the PIs of a representable algebra R, we need some approach that supplements the module theory and describes R explicitly as a subalgebra of $M_n(F)$. Thus, we keep track of the given (faithful) representation $\rho\colon R \to M_n(F)$.

Together with Theorem 15.23, Wedderburn's Principal Theorem (Theorem 15.26) applied to $\rho(R)$ provides a good strategy. Intuitively, one first writes the semisimple part $S = \rho(R)/J$ as a direct product $\prod_{u=1}^{t} S_u$, where $S_u = M_{n_u}(D_u)$. Then $S_u = S\bar{e}_u$ for suitable orthogonal central idempotents $\bar{e}_1, \dots, \bar{e}_t$ of S, which lift to a 1-sum set of orthogonal (but not necessarily central) idempotents $e_1, \dots, e_t$ of R. Let us define the explicit form that we want to investigate.

Definition 25C.12. A representation $\rho\colon R \to M_n(F)$ has **block upper triangular form** if there is a 1-sum set of orthogonal idempotents $e_1, \dots, e_t$ of $M_n(F)$ for which the algebra $e_u\rho(R)e_u \cong M_{n_u}(F)$ (for suitable n_u) whereas $e_u\rho(R)e_v = 0$ for all $u > v$. When ρ is the identity, we say that R has **block upper triangular form** in $M_n(F)$.

Remark 25C.13. If R has block upper triangular form in $M_n(F)$, notation as in the definition, then $\sum_{u<v} e_u R e_v$ is a nilpotent ideal J of R, whereas $R/J \cong \prod_{u=1}^{t} e_u\rho(R)e_u$ is semisimple; thus, $J = \mathrm{Jac}(R)$. In this way, block upper triangular form provides us with an explicit description of the Wedderburn decomposition of R as a direct sum of the semisimple and radical parts, enabling us to compute with the Wedderburn components.

For example, any evaluation of the standard polynomial s_{2n_1} on R is 0 on the first diagonal block, and it is easy to see inductively that $s_{2n_1} \cdots s_{2n_t}$ is a PI of R.

Example 25C.14. (i)

$$R = \begin{pmatrix} F & F & F \\ F & F & F \\ 0 & 0 & F \end{pmatrix}.$$

Here the semisimple part is $\begin{pmatrix} F & F & 0 \\ F & F & 0 \\ 0 & 0 & F \end{pmatrix}$ and the radical is $\begin{pmatrix} 0 & 0 & F \\ 0 & 0 & F \\ 0 & 0 & 0 \end{pmatrix}$; $e_1 = e_{11} + e_{22}$ and $e_2 = e_{33}$. The polynomial $s_4(x_1, \dots, x_4)[x_5, x_6]$ is a PI of R.

(ii)

$$R = \begin{pmatrix} F & F & 0 \\ 0 & F & 0 \\ 0 & F & F \end{pmatrix}.$$

R does not seem to be in upper block triangular form, but changing the base by switching the second and third components yields the algebra $\begin{pmatrix} F & 0 & F \\ 0 & F & F \\ 0 & 0 & F \end{pmatrix}$, which has the desired form, the radical being $\begin{pmatrix} 0 & 0 & F \\ 0 & 0 & F \\ 0 & 0 & 0 \end{pmatrix}$. (Note that $[x_1, x_2][x_3, x_4]$ is a PI of R, so more care is required when distinguishing upper block triangular algebras in terms of their identities.)

Not all finite-dimensional algebras can be represented so well. In order to describe the procedure more easily, we assume that F is algebraically closed, in which case each $D_u = F$ by Theorem 14.27. Even in this case, one has to deal with subtleties of the following sorts:

$$R = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} : \alpha \in F \right\} \cong F; \tag{25C.2}$$

$$R = \left\{ \begin{pmatrix} \alpha & \beta \\ 0 & \alpha \end{pmatrix} : \alpha, \beta \in F \right\} \cong F[\lambda]/\langle \lambda^2 \rangle. \tag{25C.3}$$

In (25C.3), $\dim_F J = \dim_F(R/J) = 1$, so any embedding into matrices must involve some identification of components, which we call "gluing." Let us define this notion more precisely.

Definition 25C.15. Two isomorphic simple subalgebras S_1 and S_2 of R are **glued** if there are projections $\pi_i: R \to S_i$ for which $\ker \pi_1 = \ker \pi_2$.

In this case, $\varphi = \pi_2 \pi_1^{-1}$ is a well-defined isomorphism $\varphi: S_1 \to S_2$. Here is our major instance of gluing:

Remark 25C.16. Suppose a simple algebra S is a subdirect product of simple algebras $S_1, \dots, S_t$; cf. Digression 13.30. Then the natural onto maps $\pi_k: S \to S_k$ are also 1:1, since $\ker \pi_k \lhd S$ and is thus 0. Thus, the maps $\ker \pi_k \ker \pi_j^{-1}: S_j \to S_k$ are instances of gluing, and we can identify any element $s \in S$ with $(\pi_1(s), \dots, \pi_t(s)) \in S_1 \times \cdots \times S_t$.

For any algebra R in block upper triangular form, the semisimple algebra $\overline{R} = R/J$ has t simple components $S_u \cong M_{n_u}(F)$, $1 \le u \le t$, for some t. Each S_u is a subdirect product of glued $n_u \times n_u$ diagonal blocks $S_{u,1}, \dots, S_{u,m_u}$ for some m_u, and we view the elements of S_u as m_u-tuples (along the diagonal) as in Remark 25C.16. For example, for R as in (25C.3), $u = 1$ and $m_1 = 2$.

Gluing is always needed to put an algebra R in block upper triangular form when R is local with $J \neq 0$, since then $R/J \cong F$.

Note that the glued blocks do not have to occur consecutively; as illustrated in the example:

$$\begin{pmatrix} S_{1,1} & * & * & * & * \\ 0 & S_{2,1} & * & * & * \\ 0 & 0 & S_{1,2} & * & * \\ 0 & 0 & 0 & S_{1,3} & * \\ 0 & 0 & 0 & 0 & S_{2,2} \end{pmatrix}, \tag{25C.4}$$

where $*$ denotes some arbitrary block (which will be in the radical, because of its position). If we rearranged the $S_{u,v}$ to be next to each other, this would force some of the $*$ below the diagonal and thus ruin the block upper triangular form.

THEOREM 25C.17. *Any F-subalgebra R of $M_n(F)$ can be put into block upper triangular form (with respect to a suitable change of base of $F^{(n)}$).*

Proof. We go over the proof of Remark 15.23′ with more attention to detail. Namely, we view $M_n(F)$ as the endomorphism ring of $V = F^{(n)}$.

The assertion is standard for $J = 0$, because the block upper triangular form becomes block diagonal form. Take a simple component R_1 of R, and pick $v \in V$. Let $V_0 = Rv$, and let V_1 be an R-module complement of V_0 in V. Then R acts as a matrix algebra on V_0, and by induction the action of R on V_1 can be put into block diagonal form, so putting together the two actions yields the desired form for R on V. (It may happen that some simple component S_u of R might appear in the actions of R both on V_0 and on V_1; for example, this would occur in (25C.2). But in this case, since S_u is simple, Corollary 15A.4 yields a natural isomorphism of S_u with some matrix block $S_{u,k}$ of $\operatorname{End} V_k$ for $k = 0, 1$; combining these two isomorphisms "glues" $S_{u,0}$ to $S_{u,1}$, as desired.)

So assume $J \neq 0$, and take t such that $J^t = 0$ but $J^{t-1} \neq 0$. Note that $t \le n$ by Theorem 15.23. Let $V_0 = \{v \in V : Jv = 0\}$. Then $JV_0 = 0$, so R/J acts on V_0; also $RV_0 \subseteq V_0$ since $J(RV_0) \subseteq JV_0 = 0$, so R acts on V/V_0.

Lift V/V_0 to a complementary space V_1 of V_0 in V. Defining the (idempotent) projections $e_u : V \to V_u$, viewed as idempotent matrices in $M_n(F)$, we embed R into

$$W = e_1 M_n(F) e_1 \oplus e_1 M_n(F) e_0 \oplus e_0 M_n(F) e_0.$$

(This can be seen directly, or via Remark 13.6, noting that one of the Peirce components is 0.) Clearly $e_u J e_u \lhd e_u M_n(F) e_u$, for $u = 0, 1$, so we conclude the proof by induction on t. □

Theorem 25C.17 is the key first step in studying f.d. algebras in terms of their polynomial identities. This theory can be generalized to representable PI-rings that are not necessarily F-algebras, but then gluing becomes more complicated.

Chapter 26

Hopf Algebras

In the last half-century, a remarkably ubiquitous algebraic structure called a **Hopf algebra** has been obtained by dualizing Remark 18.19; the ensuing theory unifies some of the structures (group algebras, enveloping algebras of Lie algebras, algebraic groups, Galois extensions) that have dominated the exposition of this book. Hopf algebras also have major applications in other areas, including quantum theory. In the rest of this short chapter, we glance at some of their facets. We follow Sweedler [**Swe**], Montgomery [**Mo1**], and Cohen-Gelaki-Westreich [**CoheGW**] for the general theory. A lovely treatment from the perspective of algebraic groups is given in Abe [**Ab**]. The connection to deformation theory is explained in Shnider-Sternberg [**ShnS**]. In Hopf theory, the idea underlying a proof often may be clear, but the computations can be formidable. For this reason, many expositions of Hopf algebras leave computations to the reader; the most explicit details are usually found in the book by Dascalescu, Nastasescu, and Raianu [**DasNR**].

Coalgebras and bialgebras

Recall that any associative algebra $A = (A, \mu, \iota)$ over a field F has a multiplication map $\mu\colon A \otimes A \to A$ and a left and right unit element 1 yielding the map $\iota\colon F \to A$ given by $\alpha \mapsto \alpha \cdot 1$, satisfying the two respective commutative diagrams (18.4) and (18.5) of Remark 18.19. Reversing the arrows would yield two new maps on a vector space C, called **comultiplication** $\Delta\colon C \to C \otimes C$ and the **counit** $\epsilon\colon C \to F$, for which the following diagrams are to be commutative:

Coassociativity:

(26.1)
$$\begin{array}{ccc} C \otimes_F C \otimes_F C & \xleftarrow{\Delta \otimes 1_C} & C \otimes_F C \\ \uparrow{\scriptstyle 1_C \otimes \Delta} & & \uparrow{\scriptstyle \Delta} \\ C \otimes_F C & \xleftarrow[\Delta]{} & C \end{array}$$

Left and right counit:

(26.2)
$$\begin{array}{ccc} F \otimes_F C & \xleftarrow{\epsilon \otimes 1_C} & C \otimes_F C \\ & \overset{\cong}{\nwarrow} & \uparrow{\scriptstyle \Delta} \\ & & C \end{array} \qquad \begin{array}{ccc} C \otimes_F F & \xleftarrow{1_C \otimes \epsilon} & C \otimes_F C \\ & \overset{\cong}{\nwarrow} & \uparrow{\scriptstyle \Delta} \\ & & C \end{array}$$

In other words, if we write $\Delta(a) = \sum a_{1i} \otimes a_{2i}$ for $a \in C$, then comultiplication Δ and the counit ϵ should satisfy the following equations:

(CA1)
$$\sum \Delta(a_{1i}) \otimes a_{2i} = \sum a_{1i} \otimes \Delta(a_{2i});$$

(CA2)
$$\sum \epsilon(a_{1i}) a_{2i} = a = \sum a_{1i} \epsilon(a_{2i}).$$

For the rest of the chapter, we see what happens when these new operations are incorporated into the structure. We assume that the base ring F is a field, although at times it would be convenient to have a more general commutative base ring.

Definition 26.1. A **coalgebra** (C, Δ, ϵ) is a vector space C with comultiplication $\Delta\colon C \to C \otimes_F C$ and counit $\epsilon\colon C \to F$ as defined above.

The notation is cumbersome, so, following Heyneman and Sweedler, we simply delete the subscript i, writing throughout

$$\Delta(a) = \sum a_1 \otimes a_2.$$

(Physicists also delete the $\sum$.) Thus (CA1) and (CA2) become

$$\sum \Delta(a_1) \otimes a_2 = \sum a_1 \otimes \Delta(a_2); \qquad \sum \epsilon(a_1) a_2 = a = \sum a_1 \epsilon(a_2).$$

We also write multiplication in the usual notation, i.e., ab for $\mu(a, b)$.

Let us dualize other standard definitions from algebra by means of this formalism of reversing arrows. For example, an algebra homomorphism

$f\colon A_1 \to A_2$ satisfies $\mu_{A_2}(f \otimes f) = f\mu_{A_1}$ and $f\iota_{A_1} = \iota_{A_2}$, so, dualizing, we define a **morphism** $(C_1, \Delta_1, \epsilon_1) \to (C_2, \Delta_2, \epsilon_2)$ **of coalgebras** to be a map $f\colon C_1 \to C_2$ satisfying

$$(f \otimes f)\Delta_1 = \Delta_2 f, \quad \epsilon_2 f = \epsilon_1, \tag{26.3}$$

i.e., $\sum f(a_1)\otimes f(a_2) = \sum f(a)_1 \otimes f(a)_2$ and $\epsilon_2 f(a) = \epsilon_1(a)$. Having defined the objects and the morphisms, we have the category of coalgebras, and we are ready to study their structure theory.

Remark 26.2. To check that a given vector space C is a coalgebra with respect to a suitable comultiplication Δ and counit ϵ, it is enough to verify that axioms (CA1) and (CA2) hold for the elements of some base of C. There are several types of elements for which this is immediate:

(i) **grouplike** elements g, for which $\Delta(g) = g \otimes g$, since

$$(1_C \otimes \Delta)\Delta(g) = g \otimes (g \otimes g) = (g \otimes g) \otimes g = (\Delta \otimes 1_C)\Delta(g).$$

In this case (CA2) requires $g = \epsilon(g)g$, so $\epsilon(g) = 1$.

(ii) **primitive** elements a, such that $\Delta(a) = a \otimes 1 + 1 \otimes a$ (assuming $\Delta(1) = 1 \otimes 1$), since

$$\begin{aligned}(1_C \otimes \Delta)\Delta(a) &= a \otimes 1 \otimes 1 \ + \ 1 \otimes (a \otimes 1 + 1 \otimes a) \\ &= (a \otimes 1 + 1 \otimes a) \otimes 1 + 1 \otimes 1 \otimes a = (\Delta \otimes 1_C)\Delta(a).\end{aligned}$$

In this case (CA2) requires $a = \epsilon(a)1 + \epsilon(1)a = \epsilon(a)1 + a$, so $\epsilon(a) = 0$.

(iii) Generalizing (ii), if $\Delta(a) = a \otimes g + h \otimes a$ with g and h grouplike, then a satisfies

$$\begin{aligned}(1_A \otimes \Delta)\Delta(a) &= a \otimes g \otimes g \ + \ h \otimes (a \otimes g + h \otimes a) \\ &= (a \otimes g + h \otimes a) \otimes g + h \otimes h \otimes a = (\Delta \otimes 1_A)\Delta(a).\end{aligned}$$

In this case, (CA2) requires $a = \epsilon(a)g + \epsilon(h)a = \epsilon(a)g + a$, so again $\epsilon(a) = 0$.

Given vector spaces V and W, we define the **twist map**

$$\tau\colon V \otimes W \to W \otimes V$$

by $\tau(v \otimes w) = w \otimes v$; cf. Proposition 18.16. If there is some ambiguity concerning V and W, we may write $\tau_{V,W}$ for τ. Since the multiplication map μ of an algebra A is commutative iff $\mu\tau_{A,A} = \mu$, we can dualize and say that a coalgebra C is **cocommutative** if $\tau_{C,C}\Delta = \Delta$. We can view duality between algebras and coalgebras explicitly via the following construction.

Definition 26.3. Given a coalgebra C and an algebra A over the field F, we define the **convolution product** $*$ on $\mathrm{Hom}(C, A)$, defining $f * g$ by

$$(f * g)(c) = (f \otimes g)\Delta(c) = \sum f(c_1)g(c_2), \quad \forall c \in C.$$

Convolution is associative (cf. Exercise 3), so $\mathrm{Hom}(C, A)$ is an algebra. In particular, taking $A = F$, we get the **dual** algebra $C^* = \mathrm{Hom}_F(C, F)$ for any coalgebra C.

The most familiar concepts in algebra often arise in this context.

Example 26.3′. Suppose V is a vector space over F with countable base $\{b_0, b_1, b_2, \dots\}$. Define comultiplication on V via $\Delta(b_k) = \sum_{i+j=k} b_i \otimes b_j$ for each $k \geq 0$. Then the power series algebra $F[[\lambda]]$ is identified with V^* under the isomorphism sending λ^k to the map given by $\{b_i \mapsto \delta_{ik} : i \geq 0\}$. The polynomial algebra $F[\lambda]$ is identified with the subalgebra of V^* consisting of all maps having finite codimension. (Compare with Exercise 10.)

Often we can obtain 'coalgebraic' concepts for a coalgebra by translating them from the corresponding algebraic notions in its dual (viewed as an algebra). For example, a coalgebra C is cocommutative iff C^* is commutative.

Similarly, if A is an algebra, one could try to define comultiplication in A^* by $(\Delta f)(a \otimes b) = f(ab)$, but this requires $A^* \otimes A^*$ to be $(A \otimes A)^*$, which is not true in general. (But see Example 26.10.)

Bialgebras.

The main idea here is to combine the two notions of algebra and coalgebra.

Definition 26.4. A **bialgebra** $A = (A, \mu, \iota, \Delta, \epsilon)$ is an associative algebra (A, μ, ι) that is also a coalgebra (A, Δ, ϵ), such that Δ and ϵ are algebra homomorphisms and μ, ι are coalgebra morphisms. In particular, $\epsilon(1) = 1$ and $\Delta(1) = 1 \otimes 1$.

Remark 26.5. Since $\ker \epsilon$ is an algebra ideal of A of codimension 1, no bialgebra (other than F itself) can be simple as an algebra.

We are finally ready for the main definition of this chapter.

Definition 26.6. A **Hopf algebra** is a bialgebra H together with a map $S\colon H \to H$ such that

$$\mu(S \otimes 1_H)\Delta = \iota\epsilon = \mu(1_H \otimes S)\Delta, \tag{26.4}$$

i.e.,

$$\sum S(a_1)a_2 = \iota\epsilon(a) = \sum a_1 S(a_2), \quad \forall a \in H. \tag{26.5}$$

Such an S is called an **antipode**.

In Exercise 3 the antipode is motivated by means of the convolution product. The antipode S of a Hopf algebra H is both an algebra anti-homomorphism and a coalgebra anti-morphism; cf. Exercise 5. Note that the definition does not require S to be 1:1, although this is known to hold when H is finite-dimensional, in which case $S^m = 1_H$ for some m. This is an important property that we need later, and a major question in the theory is, "When is S an involution (i.e., $S^2 = 1_H$)?" This is the case when H is cocommutative; cf. Exercise 6. However, recent research has focused on the more general situation for which S^2 is an inner automorphism.

Remark 26.7. If $g, h \in H$ are grouplike elements, then

$$\Delta(gh) = \Delta(g)\Delta(h) = (g \otimes g)(h \otimes h) = gh \otimes gh.$$

$S(g)g = gS(g) = 1$ by Equation (26.5), so the grouplike elements form a group, where $g^{-1} = S(g)$, on which Δ restricts to a group homomorphism.

If $a, b \in H$ are primitive elements, then

$$\begin{aligned}\Delta(ab - ba) &= \Delta(a)\Delta(b) - \Delta(b)\Delta(a)\\ &= (a \otimes 1 + 1 \otimes a)(b \otimes 1 + 1 \otimes b) - (b \otimes 1 + 1 \otimes b)(a \otimes 1 + 1 \otimes a)\\ &= (ab - ba) \otimes 1 \ + \ 1 \otimes (ab - ba),\end{aligned}$$

proving that the primitive elements form a Lie algebra, on which Δ restricts to a Lie homomorphism; also $a + S(a) = 0$, implying that $S(a) = -a$.

More generally, for a as in Remark 26.2(iii), $S(a) = -h^{-1}ag^{-1}$.

These considerations lead us to two major examples of Hopf algebras:

Example 26.8. (i) Any group algebra $F[G]$ is a cocommutative Hopf algebra over F, in which each element of G is grouplike (and this also determines the counit and antipode by Remarks 26.2 and 26.7). Note that $\epsilon\colon F[G] \to F$ is the augmentation map (Example 19.17), Δ restricts to a group homomorphism, and the antipode S is the involution given in Exercise 19.28.

(ii) The universal enveloping algebra $U(L)$ of a Lie algebra L is a cocommutative Hopf algebra, in which each element of L is primitive. Indeed, we define the map $\Delta\colon L \to U(L) \otimes U(L)$ by $\Delta(a) = a \otimes 1 + 1 \otimes a$. The same computation as in Remark 26.2(ii) shows that Δ restricts to a Lie homomorphism $L \to (U(L) \otimes U(L))^-$, which by universality extends to the desired algebra homomorphism $\Delta\colon U(L) \to U(L) \otimes U(L)$. Again, the zero map on L extends to a homomorphism $\epsilon\colon U(L) \to F$, called the **augmentation map**, satisfying $a \mapsto 0$, $\forall a \in L$; likewise, the antipode S, uniquely defined by $S(a) = -a$ for $a \in L$, is called the **standard involution** of $U(L)$.

Other examples come from algebraic geometry.

Example 26.9. (i) If G is an algebraic group, then its coordinate algebra $\mathcal{A}$ has a coalgebraic structure described already in Remark 19B.13.

(ii) The coordinate algebra $\mathcal{A}$ of the matrix algebra $M_n(F)$ is the polynomial algebra in n^2 indeterminates $F[\lambda_{ij} : 1 \le i, j \le n]$, where λ_{ij} sends a matrix to its i, j entry. The counit ϵ is given by $\epsilon(\lambda_{ij}) = \delta_{ij}$, thereby making A into a bialgebra whose comultiplication is the dual of matrix multiplication, i.e., $\Delta(\lambda_{ij}) = \sum_{k=1}^{n} \lambda_{ik} \otimes \lambda_{kj}$. Although this is not a Hopf algebra, it leads to realizations of (i); cf. Exercise 12.

Example 26.10. If H is a f.d. Hopf algebra, then, by Exercise 11, its dual $H^* = \mathrm{Hom}_F(H, F)$ is also a Hopf algebra, obviously of the same dimension.

An explicit instance: Suppose G is a finite group. The group algebra $F[G]$ is a f.d. Hopf algebra, whose dual $\mathrm{Hom}_F(F[G], F)$ can be identified with the set of F-valued functions on G. If G is Abelian of exponent m and F contains a primitive m-th root of 1, $\mathrm{Hom}_F(F[G], F) \cong G$. (This is easily seen for finite cyclic groups, and then follows in general from Remark 20.25.)

The first f.d. Hopf algebra not belonging to one of these examples was discovered by Sweedler (for dimension 4) and generalized by Taft and Ng [**Ng**] to dimension n^2 for arbitrary n:

Example 26.11. Let ζ be a primitive n-root of 1. We define the algebra H with base $\{g^i a^j : 0 \le i, j < n\}$ satisfying $g^n = 1$, $a^n = 0$, and $ga = \zeta ag$; thus H has dimension n^2. Δ is given by $\Delta g = g \otimes g$ and $\Delta a = a \otimes 1 + g \otimes a$. By Remark 26.2, $\epsilon(g) = 1$ and $\epsilon(a) = 0$; $S(g) = g^{-1}$ and $S(a) = -g^{-1}a$. Then $S^2(g) = g$ and $S^2(a) = S(-g^{-1}a) = -S(a)S(g^{-1}) = g^{-1}ag = \zeta^{-1}a$, so S^2 is an automorphism of order n, implying that S has order $2n$.

Many basic results concerning these examples generalize to theorems about Hopf algebras. For example, the main step in proving that every affine algebraic group G is linear (Theorem 19B.19) is to prove that every f.d. subspace of the coordinate algebra $\mathcal{A}$ is contained in a f.d. G-subspace of $\mathcal{A}$. The key observation (Remark 19B.16) can be translated to what is known as the Fundamental Theorem of coalgebras, given in Exercise 17.

Hopf modules

One pleasant aspect of the theory is that many properties of group and Lie actions have natural formulations and proofs (albeit with new terminology and a huge amount of notation, much of which is bypassed in this overview). For example, recall that the tensor product of group modules becomes a group module under the diagonal action, which arose somewhat mysteriously in Definition 20.27 and Exercise 21.56. We can explain this in the context of Hopf algebras.

First note that any module M over an algebra A is a vector space endowed with a multiplication $\mu_M\colon A \otimes_F M \to M$ satisfying $\mu_M(\mu_A \otimes 1_M) = \mu_M(1_A \otimes \mu_M)$, and the canonical isomorphism $F \otimes_F M \cong M$ is $\mu_M(\iota \otimes 1_M)$.

Remark 26.12. Suppose V and W are modules (in this sense) over a Hopf algebra H. Then $V \otimes_F W$ is also a module, under the scalar multiplication

$$h(v \otimes w) = \sum h_1 v \otimes h_2 w.$$

More formally, if $\mu_V\colon H \otimes V \to V$ and $\mu_W\colon H \otimes W \to W$ denote the scalar multiplications for V and W respectively, then

$$\mu_{V\otimes W} = (\mu_V \otimes \mu_W)(1_H \otimes \tau_{H,V} \otimes 1_W)(\Delta \otimes 1_V \otimes 1_W)$$

is the scalar multiplication $H \otimes (V \otimes W) \to V \otimes W$.

In particular, if we view H as a right module over itself, then $H \otimes_C M$ is a module whose scalar multiplication takes Δ into account.

Of course, the next thing is to dualize the above definition of module.

Definition 26.13. A **comodule** over an F-coalgebra C is a vector space M together with an F-linear **scalar comultiplication** $\rho\colon M \to C \otimes_F M$ satisfying the coassociativity and coscalar actions:

$$(\Delta \otimes 1_M)\rho = (1_C \otimes \rho)\rho; \tag{26.6}$$

$$(\epsilon \otimes 1_M)\rho \quad \text{is the canonical isomorphism} \quad M \to F \otimes_F M. \tag{26.7}$$

A (left) **Hopf module** over a Hopf algebra H is an H-module M that is also an H-comodule for which the scalar comultiplication ρ is a map of H-modules, viewing $H \otimes_F M$ as a module via Remark 26.12; explicitly, the condition is

$$\rho(ha) = \Delta(h)\rho(a). \tag{26.8}$$

We denote $\rho(a) = \sum a_{-1} \otimes a_0$, where $a_{-1} \in H$ and $a_0 \in M$. (Unfortunately, the subscript -1 has become standard for left Hopf modules, since most texts deal with right Hopf modules, defined with positive subscripts.)

Example 26.13′. Any Hopf algebra H is a left and right Hopf module, taking $\rho = \Delta$.

Remark 26.14. The Hopf modules over a Hopf algebra H comprise the objects of a category whose morphisms $f\colon M \to N$ are module maps preserving the comodule structure in the sense that

$$\rho_N \circ f = (1_M \otimes f)\rho_M,$$

so $\rho(f(a)) = \sum a_{-1} \otimes f(a_0)$.

Remark 26.15. Dually to Remark 26.12, if V and W are comodules over a Hopf algebra H with respective scalar comultiplications ρ_V and ρ_W, then $V \otimes W$ is also a comodule under the scalar comultiplication

$$\rho_{V\otimes W} = (\mu_H \otimes 1_V \otimes 1_W)(1_H \otimes \tau_{V,H} \otimes 1_W)(\rho_V \otimes \rho_W).$$

Consequently, the tensor product of Hopf modules is a Hopf module. In particular, for any Hopf module M, $H \otimes_F M$ is a Hopf module. Thus, $H \otimes_F H$ is a Hopf module and, by induction, we define $H^{\otimes n} = H \otimes \cdots \otimes H$, taken n times. The Hopf version of the adjoint isomorphism is obtained in Exercise 41.

To understand the category of Hopf modules together with the tensor product, we digress a bit to introduce more categorical structure.

Definition 26.16. A **multiplication** defined on a category $\mathcal{C}$ is a functor $\otimes\colon \mathcal{C} \times \mathcal{C} \to \mathcal{C}$. A category $\mathcal{C}$ is called **monoidal** if it has a multiplication having a unit object $1_\mathcal{C}$, together with natural isomorphisms $\ell\colon 1_\mathcal{C} \otimes A \to A$ and $r\colon A \otimes 1_\mathcal{C} \to A$ for each object A, and also having natural isomorphisms

$$\Phi = \Phi_{A_1,A_2,A_3}\colon (A_1 \otimes A_2) \otimes A_3 \to A_1 \otimes (A_2 \otimes A_3)$$

satisfying the following conditions for any objects A_i in $\mathcal{C}$:

(i) The natural map $r \otimes 1_{A_2}\colon (A_1 \otimes 1_\mathcal{C}) \otimes A_2 \to A_1 \otimes A_2$ factors as

$$(A_1 \otimes 1_\mathcal{C}) \otimes A_2 \xrightarrow{\Phi} A_1 \otimes (1_\mathcal{C} \otimes A_2) \xrightarrow{1\otimes\ell} A_1 \otimes A_2. \tag{26.9}$$

(ii) There is a commutative diagram arising from the two canonical ways of going isomorphically

$$((A_1 \otimes A_2) \otimes A_3) \otimes A_4 \to A_1 \otimes (A_2 \otimes (A_3 \otimes A_4)).$$

This diagram is called **MacLane's pentagon**, since one way is a composition of three isomorphisms whereas the other way is a composition of two isomorphisms.

Thus, R-**Mod** (for R commutative) has a multiplication given by tensor product of modules in the usual way; multiplication on morphisms is given by $(f, g) \mapsto f \otimes g$. But the situation is far more complicated when R is noncommutative, if we do not want to pass to bimodules. For any Hopf algebra H, the category H-**Mod** is monoidal, defining the tensor product as above; cf. Exercise 27. This raises the question of when $M \otimes N \cong N \otimes M$, which we address shortly.

Definition 26.17. A **sub-comodule** N of a comodule M over an F-coalgebra C is an F-subspace N of M for which $\rho(N) \subseteq C \otimes_F N$. A **Hopf submodule** of a Hopf module is a submodule that is also a sub-comodule. A right and left Hopf submodule M of H is called a **Hopf ideal** of H; in this case one can form the factor Hopf algebra H/M. (But not all Hopf homomorphic images of Hopf algebras can be written in this way!)

We often want the Hopf module M itself to be an algebra, and so we introduce yet another definition.

Definition 26.18. An H-**module algebra** over a Hopf algebra H is an algebra A that is also an H-module for which μ_A and ι_A are H-module maps; explicitly

$$h(ab) = \sum (h_1 a)(h_2 b), \qquad h1_A = \varepsilon(h)1_A.$$

Dually, an H-**comodule algebra** is an algebra A that is also an H-comodule for which μ_A and ι_A are H-comodule maps. (Likewise, one can define module coalgebras and comodule coalgebras.)

Invariants and coinvariants.

Definition 26.19. Given a module M over a Hopf algebra H, define its **invariants** $M^H = \{v \in M : hv = \epsilon(h)v,\ \forall h \in H\}$. Dually, for a comodule M, define the **coinvariants** $M^{\mathrm{co}\,H} = \{v \in M : \rho(v) = 1 \otimes v\}$.

Example 26.20. (i) Suppose H is the group algebra $F[G]$ of a group G. Then the H-modules are just the G-modules, and M^H is the fixed submodule of M under the action of G. In the case when M is an H-module algebra, which means that G acts as automorphisms on M, we can view M as a

Galois extension of M^G, and this indeed has a suitable generalization to Hopf algebras, described in [**Mo1, Chapter 8**].

(ii) Again, for $H = F[G]$, an H-comodule algebra A is precisely a G-graded algebra. Indeed, we write $\rho(a) = \sum g \otimes a_g$, noting by Equation (26.7) that $a = \sum \varepsilon(g) a_g = \sum a_g$. (This could also be viewed as a $\mathrm{Hom}_F(F[G], F)$-module algebra in the notation of Example 26.10.)

Furthermore, (26.6) shows that $\rho(a_g) = a_g$, and thus $A = \bigoplus_{g \in G} A_g$; we have $A_g A_h \subseteq A_{gh}$ since A is an $F[G]$-comodule algebra. Note that the algebra of coinvariants is A_e.

(iii) If H is the universal enveloping algebra $U(L)$ of a Lie algebra L, then the H-modules are the Lie modules over L. Now an H-module algebra is an algebra A on which L acts as derivations, and A^H is the corresponding ring of constants. Furthermore, in characteristic p we could use the restricted enveloping algebra of a f.d. restricted Lie algebra, which is finite-dimensional, cf. Exercise 21C.8; this "explains" the phenomenon of a Galois theory for derivations of inseparable field extensions, originally developed by MacLane and Jacobson.

The coinvariants play the following basic role in Hopf theory.

Theorem 26.21 (The Fundamental Theorem of Hopf Modules). *Any Hopf module M is isomorphic to $H \otimes M^{\mathrm{co}\,H}$ as Hopf modules (the latter under the "trivial action" $h'(h \otimes a) = (h'h \otimes a)$).*

To prove Theorem 26.21, one can define $\varphi\colon H \otimes M^{\mathrm{co}\,H} \to M$ by means of the balanced map $H \times M^{\mathrm{co}\,H} \to M$ given by $(h, a) \mapsto ha$.

The reverse direction is somewhat trickier, in part because of the notation. We write $\rho(a) = \sum a_{-1} \otimes a_0$, where $a_{-1} \in H$, $a_0 \in M$, and $\Delta(a_{-1}) = \sum a_{-1,1} \otimes a_{-1,2}$, and define $\psi\colon M \to H \otimes M^{\mathrm{co}\,H}$ by

$$a \mapsto \sum a_{-1,1} \otimes S(a_{-1,2}) a_0.$$

One computes that $\psi = \varphi^{-1}$.

(The verification is sketched in Exercise 24; details can be found in [**DasNR, Theorem 4.4.6**].) A construction is given in Exercises 34–37 that unifies many familiar ring-theoretic and group-theoretic constructions.

Quasi-triangular Hopf algebras and the quantum Yang-Baxter equations (QYBEs)

Given H-modules V and W, one views $V \otimes W$ and $W \otimes V$ as H-modules via Remark 26.12, thereby raising the question of whether $V \otimes W \cong W \otimes V$

as H-modules. When the Hopf algebra H is cocommutative, it is easy to check that $V\otimes W \cong W\otimes V$ under the twist map $\tau\colon v\otimes w \mapsto w\otimes v$. But physicists found that this particular isomorphism fails in their models of quantum mechanics, and they proposed a different isomorphism.

Definition 26.22. Suppose V is a vector space and $R\colon V\otimes V \mapsto V\otimes V$ is a linear transformation. Define $R^{ij}\colon V\otimes V\otimes V \to V\otimes V\otimes V$ by

$$R^{12} = R\otimes 1_V, \quad R^{23} = 1_V\otimes R, \quad R^{13} = (1_V\otimes\tau)(R\otimes 1_V)(1_V\otimes\tau).$$

In other words, R^{12} acts like R on the first two tensor components, R^{23} acts like R on the last two tensor components, and R^{13} acts like R on the outer two tensor components. The transformation R is said to satisfy the **Quantum Yang-Baxter equation** (QYBE) if R is invertible and

$$R^{12}R^{13}R^{23} = R^{23}R^{13}R^{12}. \tag{26.10}$$

Note that if R is a solution to the QYBE, then $B = \tau\circ R$ satisfies

$$B^{23}B^{12}B^{23} = \tau^{23}R^{23}\tau^{12}\tau^{23}\tau^{23}R^{12}\tau^{23}R^{23} = \tau^{13}R^{12}R^{13}R^{23}$$
$$= \tau^{13}R^{23}R^{13}R^{12} = \tau^{12}(\tau^{12}\tau^{13}R^{23}\tau^{23})(\tau^{12}R^{13}\tau^{12})\tau^{12}R^{12} = B^{12}B^{23}B^{12},$$

which is precisely the braid relation of Exercises 19A.24ff.

Obviously, the identity map on $V\otimes V$ and the twist map both satisfy the QYBE; a subtler example is given in Exercise 31. Interest in Hopf algebras took a quantum leap with the discoveries by Jimbo [**Ji**] and by Drinfel′d [**Dr1**], [**Dr2**] that Hopf algebras often provide models for solutions to the QYBE. This leads us to our next definition.

Definition 26.23. A Hopf algebra H is **almost cocommutative** if the antipode S is invertible and if there is an invertible element $R\in H\otimes H$ such that

$$\tau(\Delta(a)) = R\Delta(a)R^{-1}$$

for all $a\in H$. The designation R for this element is standard, and we often write (H,R) instead of H to emphasize the importance of the element R.

Left multiplication by R gives us a map $H\otimes H\to H\otimes H$, so we can define R^{ij} as in Definition 26.22; explicitly, writing $R=\sum a_i\otimes b_i$, we define

$$R^{12} = \sum a_i\otimes b_i\otimes 1; \qquad R^{23} = \sum 1\otimes a_i\otimes b_i; \qquad R^{13} = \sum a_i\otimes 1\otimes b_i.$$

An almost cocommutative Hopf algebra (H,R) is called **quasi-triangular** if

$$(\Delta\otimes 1_H)R = R^{13}R^{23} \qquad \text{and} \qquad (1_H\otimes\Delta)R = R^{13}R^{12}.$$

H is called **triangular** if $\tau(R) = R^{-1}$.

Example 26.24. Any cocommutative Hopf algebra is triangular, taking $R = 1\otimes 1$. On the other hand, there may be other possibilities for R, even for the same Hopf algebras.

Remark 26.25. A computation using Definition 26.23 shows that any quasitriangular Hopf algebra provides a solution to the QYBE.

Furthermore, when the Hopf algebra H is quasitriangular, there is a natural isomorphism $V\otimes W\to W\otimes V$ for H-modules V and W; cf. Exercise 30.

Drinfel′d [**Dr1**] showed that any Hopf algebra with an invertible antipode can be embedded naturally in a quasitriangular Hopf algebra whose underlying vector space is $(H^*)^{\mathrm{cop}}\otimes H$; cf. [**CoheGW, Definition 4.3.8**].

Quantum groups provide excellent examples for this theory. An instance is given in Exercise 39, which can be generalized to quantum enveloping algebras for any semisimple Lie algebra, by means of quantizing the Cartan matrix of Theorem 21.108; this is done elegantly in Jantzen [**Jan, Chapter 4**]. It turns out that quantum enveloping algebras also provide solutions to the QBYE, and this remarkable fact fits neatly into the structure theory of Hopf algebras! [**Jan, Chapter 3**] describes R explicitly in these terms. Conversely, Brown and Goodearl [**BrokaG**] show how the goal of obtaining Hopf operations enables one to define the appropriate quantum structure.

Hopf cohomology.

Hopf modules "explain" some of the fine points in cohomology theory.

Remark 26.26. Suppose H is any bialgebra. F becomes an H-module via the augmentation map ϵ, and we define the **Sweedler complex A** by taking $A_n = H^{\otimes n}$, with boundary operator given by

$$d(h_1\otimes\cdots\otimes h_n) = \sum_{i=1}^{n} h_1\otimes\cdots\otimes h_{i-1}\otimes h_{i+1}\otimes\cdots\otimes h_n.$$

This corresponds to Example 25.65 when H is the group algebra $F[G]$. More generally, the Sweedler resolution is free and is f.g. whenever H is a f.d. Hopf algebra. This follows at once from the following result:

PROPOSITION 26.27. *Suppose H is a Hopf algebra with invertible antipode S, and K is a Hopf subalgebra of H. Given a K-Hopf module M, let $U = H\otimes_K M$ viewed as a K-Hopf module, and $V = H\otimes_K M$, viewed as a K-module under the "trivial action" of Remark 18.9′′: $k(h\otimes a) = h\otimes ka$. Then $U\cong V$ as K-modules.*

Proof. Define $\varphi: V \to U$ by

$$h \otimes a \mapsto \rho(a)(h \otimes 1) = \sum a_{-1}h \otimes a_0$$

and $\psi: U \to V$ by $h \otimes a \mapsto \sum S^{-1}(a_{-1})h \otimes a_0$. A (relatively) short verification (see [**DasNR, p. 294**] for the details) shows that φ and ψ are inverses. □

Thus, the Hopf theory gives a general framework in which to unite the various homological theories of groups and Lie algebras.

To get Hopf cohomology, we perform the same procedure as in group cohomology and Lie cohomology by the analogous device of going to the dual Hopf algebra H^* of Example 26.10. Recalling the definition of comultiplication in A^* as $\Delta f(a,b) = f(ab)$, let us rewrite the formula (25.15) more suggestively in Hopf terminology. Namely, we define Δ_i as applying the comultiplication Δ in the i component; also the natural projection $A^{(n+1)} \to A^{(n)}$ erasing the first component has a dual operation, the "coprojection" operation ν given by $\nu f(h_1, \dots h_{n+1}) = h_1 f(h_2, \dots, h_n)$. Now, by Equation (25.15), the following holds on all grouplike elements:

$$\delta^n f = \nu f + \sum_{i=1}^{n} (-1)^i \Delta_i f + (-1)^{n+1} f \otimes \mathbf{1},$$

where $\mathbf{1}$ takes on the constant value 1.

Finite-dimensional Hopf algebras

Having seen how Hopf algebras unify so many different theories, the algebraist is motivated to try to classify Hopf algebras, starting with the finite-dimensional ones (which include group algebras of finite groups and their duals, restricted enveloping algebras of f.d. Lie algebras, and quantum algebras in the "degenerate case" where q is a root of 1, cf. Definition 16A.1). Here is a good starting point.

Corollary 26.28 (Nichols-Zoeller [**NiZ**]). *If K is a Hopf subalgebra of a f.d. Hopf algebra H, then H is free as a K-module, and* $\dim K \mid \dim H$.

Proof. In the setup of Proposition 26.27, take $M = K$. Then $H \cong U$, which is a free K-module by Theorem 26.21, as desired. □

We call a Hopf algebra **semisimple** if it is semisimple as a ring. Appealing to the Wedderburn structure theory, it is natural to examine semisimple f.d. Hopf algebras, starting with a Hopf version of Maschke's Theorem. The situation is clarified somewhat by means of the following notion.

Definition 26.29. A **left integral** of a Hopf algebra H is an element $t \in H$ such that $ht = \epsilon(h)t$, $\forall h \in H$. $\int_H^l$ denotes the space of left integrals of H. A **right integral** of H is defined analogously, and $\int_H^r$ denotes the space of right integrals of H.

The left integral enables us to carry out the averaging procedure of Lemma 19.25 that was used in the proof of Maschke's Theorem, thereby leading to the following generalization for f.d. Hopf algebras:

Theorem 26.30. *A f.d. Hopf algebra H over a field F is semisimple iff* $\varepsilon(\int_H^l) \neq 0$.

Proof. ($\Rightarrow$) $\ker \varepsilon \lhd H$, so, by Theorem 14.13, has a nonzero complement L as a left ideal in H. We claim that any $0 \neq t \in L$ is a left integral. Indeed, for any $h \in H$, $h - \varepsilon(h)1 \in \ker \varepsilon$, so $(h - \varepsilon(h)1)t \in L \cap \ker \varepsilon = 0$, implying $ht = \varepsilon(h)t$. But $\varepsilon(t) \neq 0$ since $t \notin \ker \varepsilon$.

($\Leftarrow$) We show that H is complemented as a left H-module, by the averaging process. Pick $t \in \int_H^l$ such that $\varepsilon(t) = 1$, and write $\Delta(t) = \sum t_1 \otimes t_2$.

Suppose $L < H$, and let $\pi: H \to L$ be a projection as vector spaces over F. Define $\tilde{\pi}: H \to L$ by

$$\tilde{\pi}(a) = \sum t_1 \pi(S(t_2)a) = \mu(1 \otimes \pi)(1 \otimes S)\Delta(t)(1 \otimes a), \quad \forall a \in H. \tag{26.11}$$

For any $h \in H$, writing $h = \sum \varepsilon(h_1)h_2$ and pushing through Equation (26.11), one generalizes the computation of Lemma 19.25 and checks that $\tilde{\pi}(ha) = h\tilde{\pi}(a)$; we leave the 6-line verification to the reader. Hence $\tilde{\pi}$ is a projection of H-modules, implying that H is complemented, as desired. □

Remark 26.31. Larson and Sweedler showed that any f.d. Hopf algebra H over a field F has a left integral and a right integral, and moreover $\dim_F \int_H^l = 1 = \dim_F \int_H^r$, and $S(\int_H^l) = \int_H^r$. Details can be found in [**Swe**], [**Ab**], or [**Mo1**]. It follows that H is a Frobenius algebra, as defined in Exercise 19.24.

Of course, along with semisimplicity comes its dual:

Definition 26.32. A coalgebra A is **simple** if A has no proper subcoalgebras. *A* **cosemisimple** coalgebra is a finite direct sum of simple subcoalgebras.

In the 1980s, Larson and Radford proved two basic results in characteristic 0:

1. Any f.d. cosemisimple Hopf algebra H is semisimple.

2. A f.d. Hopf algebra H is cosemisimple (and thus semisimple) iff $S^2 = 1_H$.

Perhaps the last word in this line comes in the following theorem of Etingof and Gelaki [**EtG2**]:

A f.d. Hopf algebra H is semisimple and cosemisimple iff $S^2 = 1$ and $\mathrm{char}(F)$ does not divide $\dim_F H$.

Kaplansky's conjectures.

In order to stimulate the development of a structure theory for f.d. Hopf algebras along the classical lines of group theory, Kaplansky [**Kapl2**] laid forth a series of conjectures which set the agenda for research for many years. Here is a sampling of his more influential conjectures, where H denotes a f.d. Hopf algebra over a field. (After stating each conjecture, we indicate its current state in a parenthetical comment.)

1. (Generalizing Lagrange's Theorem) H is free as a module over any Hopf subalgebra K; hence, $\dim_F K$ divides $\dim_F H$. (This was proved by Nichols-Zoeller; cf. Corollary 26.28.)

2. If H is semisimple, then the dimension of any simple H-module divides $\dim H$. In this case, we say that H has **Frobenius type**, in honor of Frobenius' Theorem 20.18 for group algebras. (Etingof and Gelaki [**EtG1**] proved that all f.d. quasitriangular Hopf algebras over an algebraically closed field have Frobenius type.)

3. For any n, there are only finitely many isomorphism classes of n-dimensional Hopf algebras. (Although counterexamples are now known, Stefan [**St**] verified this conjecture for semisimple Hopf algebras.)

4. If $\dim_F H$ is prime, then H is a group algebra over F; in particular, H is self-dual. (This was proved by G.I. Kac [**Kacg**] in a special case and by Y. Zhu [**Zhu**] in general.)

The classification for special dimensions.

Suppose H is a Hopf algebra of dimension n over an algebraically closed field F of characteristic 0. Let $G(H)$ denote the set of grouplike elements of H; then $F[G(H)]$ is a group algebra contained in H. H is said to be **trivial** if it is isomorphic to a group algebra or its dual, i.e., if $H = F[G(H)]$ or $H^* = F[G(H^*)]$. When must H be trivial? In the following summary, p and q denote primes. More details and explicit constructions are given in [**Mo2**].

As mentioned above, Zhu [**Zhu**] proved that H is trivial if $n = p$.

On the other hand, when $n = p^2$, Example 26.11 is nontrivial, but is not semisimple. This leads one to subdivide the theory into the semisimple and non-semisimple cases.

CASE I. H is semisimple. Masuoka proved that H is trivial for $n = p^2$. Etingof and Gelaki extended this to dimension pq for primes $p \neq q$ by verifying that such H must have Frobenius type; earlier, Gelaki-Westreich and Sommerhauser had shown independently that any pq-dimensional Hopf algebra of Frobenius type is trivial.

The situation for $n = p^3$ has also been classified by Masuoka, although new constructions come into play. They are Abelian extensions, defined in Exercise 19. There are seven (nonisomorphic) trivial semisimple Hopf algebras of dimension p^3, namely three (self-dual) group algebra of Abelian groups, and two group algebras of nonabelian groups and their duals. For $n = 8$, there is precisely one nontrivial example, called H_8, given in Exercise 33, which thus is self-dual. Moreover, H_8 is non-cocommutative. For $n = p^3$ with p odd, there are $p+1$ nontrivial examples, all of which are self-dual and non-cocommutative.

For $n = pq^2$, Gelaki discovered a new Hopf algebra, and S. Natale proved that all nontrivial Hopf algebras of Frobenius type have this form.

Y. Kashima has solved the case $n = 16$.

CASE II. H is not semisimple. Then there are infinitely many nonisomorphic Hopf algebras of dimension p^4.

It is not known whether Hopf algebras of dimension pq need be semisimple.

Exercises – Part VI

Chapter 23

1. Using Exercise 17.6, conclude that PI-theory can be carried out in two noncommuting indeterminates at the cost of losing multilinearity.

Representable affine algebras over a field

2. Prove that the Jacobson radical $\mathrm{Jac}(C)$ of any commutative affine algebra C is nilpotent, and thus is the set of nilpotent elements of C. (Hint: Use Proposition 6.37 of Volume 1.)
3. (Anan'in) Suppose R is an irreducible algebra (cf. Exercise 16.9) that is f.g. as a module over a central affine subalgebra C. Show that R is representable. (Hint, due to L. Small: Let N be the radical of C, which is nilpotent; say $N^n = 0$. C/N is an integral domain; indeed, if $ab \in N$ for $a, b \in C \backslash N$, then clearly $\mathrm{Ann}_R\, a \neq 0$ or $\mathrm{Ann}_R\, b \neq 0$, so $a \in N$ or $b \in N$ by Exercise 16.10.

   Take a transcendence base $\{c_1 + N, \ldots, c_d + N\}$ for the affine domain C/N (cf. Chapter 6 of Volume 1); then $C_0 = F[c_1, \ldots, c_d]$ is isomorphic to a polynomial algebra, so has some field of fractions L. By Exercise 16.10, $\mathrm{Ann}_R\, c = 0$ for every $c \in C_0$. Hence, $R \subseteq R \otimes_{C_0} L$, which is f.d. over L and thus representable.)
4. (Anan'in) For any algebra R that is f.g. as a module over a commutative affine subalgebra C, prove that R is representable. More generally, prove that if R is affine, satisfies ACC on two-sided ideals, and is f.g. over its center, then R is representable. (Hint: By Exercise 16.9, R is a finite subdirect product of irreducible F-algebras $R_1, \ldots, R_t$, each of which is embeddible in $M_{n_i}(L_i)$ by Exercise 3. Since each L_i contains a common field, they are all contained in a common field K. Embed matrices along the diagonal.)

5. Show that any affine algebra that is embeddible into matrices $M_n(K)$ over a commutative algebra K is embeddible into $M_n(H)$ for some commutative affine subalgebra H of K, and thus is representable. (Hint: The generators of R are determined by a finite number of structure constants in K, which generate the affine subalgebra H.)

   In these exercises, some of the basic structural results of commutative algebra are lifted directly to representable affine algebras.
6. Prove that the Jacobson radical J of a representable affine algebra R is nilpotent. (Hint: By Exercise 5, one has $R \subseteq M_n(H)$ for an affine domain H; then $\mathrm{Jac}(H) = 0$ by Exercise 2. For any maximal ideal P of H, H/P is an affine field and thus is finite-dimensional; hence, the image of R in $M_n(H/P)$ is f.d., so the image $\overline{J}$ of J is nilpotent, implying that $\overline{J}^n = 0$. Conclude that $J^n \subseteq \bigcap M_n(P) = M_n(\cap P) = 0$, where P runs over the maximal ideals of H.)

Nonrepresentable affine algebras

7. Suppose $T = F\{a_1, \ldots, a_\ell\}$ is a commutative, affine F-algebra and L is a f.g. ideal of T. Show that

$$R = \begin{pmatrix} F + L & T \\ L & T \end{pmatrix}$$

   is affine. (Hint: $R = F\{e_{11}, e_{12}, a_i e_{22}, b_j e_{21} : 1 \leq i \leq \ell,\ 1 \leq j \leq m\}$, where $L = \sum_{j=1}^{m} T b_j$.)
8. In Exercise 7, let $F = \mathbb{Q}$ and $T = \mathbb{Q}[\lambda, \mu]$ for commuting indeterminates λ, μ, and $L = T\lambda$. For any subset I of $\mathbb{N}$, let

$$J_I = T\lambda^2 + \sum_{i \in I} T\lambda\mu^i,$$

   and $A_I = \left(\begin{smallmatrix} J_I & J_I \\ L^2 & L^2 \end{smallmatrix}\right) \lhd R$. There are only countably many isomorphisms among the various R/A_I, so conclude that uncountably many of the R/A_I are mutually nonisomorphic.

Multilinearization

9. Specializing $x_1 \mapsto 0$, show that any identity f can be written as the sum of two identities $g + h$, where x_1 appears in all monomials of g, whereas x_1 does not appear in any monomial of h. This procedure should be performed before each step of the multilinearization procedure to assure that it works properly.
10. The polynomial $\sum_{\pi \in S_d} x_{\pi 1} \cdots x_{\pi d}$ is the multilinearization of x^d. Is the standard polynomial s_d the multilinearization of a suitable nonmultilinear polynomial?

11. By Lagrange's Theorem, any field F of q elements satisfies the identity $f(x_1) = x_1^q - x_1$. What is the multilinearization of f?
12. Any finite ring satisfies an identity of the form $f(x_1) = x_1^n - x_1^m$. What is the multilinearization of f?
13. The linearization $x_1x_2 + x_2x_1 = 0$ of the Boolean identity $x^2 - x$ does not imply the Boolean identity. Conclude that the polynomial algebra over a Boolean algebra is not Boolean.
14. Check that multilinearizing $f(x_1, x_2) = x_1^2x_2 + x_1x_2^2$ yields two different results, depending on which indeterminate one starts with.
15. (Polarization) Suppose F is an infinite field. Define the **polarization** of $f(x_1, \dots, x_m) \in F\{X\}$ with respect to x_i to be its formal derivative at x_i in the direction of some new indeterminate x; i.e., take

$$\frac{f(\dots, x_i + \alpha x, \dots) - f(\dots, x_i, \dots)}{\alpha}$$

and then specialize $\alpha \mapsto 0$. Show how polarization can substitute for the linearization procedure.
16. Prove that every identity of an algebra over a field of characteristic 0 is a consequence of its multilinearizations. (Hint: Specialize back $x' \mapsto x_i$ after the linearization procedure.)
17. A polynomial $f(x_1, \dots, x_m)$ is called **completely homogeneous** if $\deg_i f = \deg_i h$ for every monomial h of f and every indeterminate x_i. If R is an algebra over an infinite field F, prove that every identity of R is a sum of completely homogeneous identities. (Hint: Apply a Vandermonde argument to $f(\alpha_1 x_1, \dots, \alpha_m x_m)$ for $\alpha_i \in F$.)

Alternating polynomials

18. Show that any t-alternating polynomial f is the alternator of f_1, where f_1 is the sum of those monomials of f in which $x_1, \dots, x_t$ occur in ascending order.

Rosset's proof of the Amitsur-Levitzki Theorem

19. Show that $\operatorname{tr}[x_1, x_2]$ is a trace identity for $M_n(C)$ for all n (and all commutative rings C).
20. (Kostant) Show that $\operatorname{tr}(s_{2k}(x_1, \dots, x_{2k}))$ is a trace identity of $M_n(C)$ for all n and k. (Hint: $\operatorname{tr}(a_1a_2 \cdots a_{2k}) = \operatorname{tr}(a_2 \cdots a_{2k}a_1)$.)
21. Let $G = E(V)$ be the Grassmann algebra. Grading $G = G_0 \oplus G_1$, show that one can also grade $M_n(G) = M_n(G_0) \oplus M_n(G_1)$, where $M_n(G_1) = \sum M_n(G_0)e_i$. If $a_1, \dots, a_{2n} \in M_n(G_0)$ show that

$$\left(\sum_{i=1}^{2n} a_i e_i\right)^k = \sum s_k(a_{i_1}, \dots, a_{i_k}) e_{i_1} \cdots e_{i_k}.$$

22. (Amitsur-Levitzki Theorem.) Prove that the standard polynomial s_{2n} is an identity of $M_n(C)$ for any commutative ring C. (Hint according to Rosset: One may assume that $C = \mathbb{Q}$. By Exercise 21, it suffices to show that $b = (\sum_{i=1}^{2n} a_i e_i)^2$ satisfies $b^n = 0$. But $b \in M_n(G_0)$, and G_0 is commutative of characteristic 0, so using Newton's formulas one needs to show that the traces of the powers of b are 0; this follows from Exercise 20.
23. Show that the algebra of upper triangular matrices does not have any central polynomials. (Hint: Evaluate any multilinear nonidentity on matrix units to get αe_{ij} for $j \geq i$.)
24. Prove that any PI-algebra has IBN.

Hilbert series of PI-algebras

25. For any algebra $R = F\{1, r_1, \dots, r_\ell\}$ and any element $c \in \operatorname{Cent}(R)$, verify that R and $R[c^{-1}] = F\{1, c, c^{-1}, r_1c^{-1}, \dots, r_\ell c^{-1}\}$ have the same Hilbert series.
26. (Bell) Prove that any prime affine PI-algebra has a rational Hilbert series. (Hint: Localize at an evaluation of an alternating central polynomial to get a free module over a commutative affine algebra.)

PI-algebras without 1

27. For any C-algebra R_0 without 1, adjoin 1 formally to R_0 by defining $R = C \oplus R_0$ with multiplication

$$(c_1, r_1)(c_2, r_2) = (c_1c_2, c_1r_2 + c_2r_1 + r_1r_2).$$

Let $\bar{R} = R/\operatorname{Ann}_R R_0$. Show that R is PI-equivalent to R_0. Furthermore, if R_0 is semiprime, then R is semiprime, implying that R_0 has a central polynomial, and the structure theory of the text passes over to R_0.
28. Show that any prime algebra R without 1 satisfying a nontrivial identity f (in the sense that f has a coefficient c such that $c1 \neq 0$) is PI. (Hint: Prove it first for R primitive, and then pass to $R[\lambda]$.)

T-ideals and their structure theory

29. If a set $\mathcal{V}$ of algebras is closed under subalgebras, homomorphic images, and direct products, show that $\mathcal{V}$ is the variety $\mathcal{V}(\mathcal{I}(\mathcal{V}))$. (Hint: For each $A \in \mathcal{V}$ and each subset $\{a_i : i \in I\}$ of A, define a homomorphism $\varphi_{\{(a_i)\}}: F\{X\}/\mathcal{I}(\mathcal{V}) \to A$ by $\bar{x}_i \mapsto a_i$. The intersection of the kernels of the $\varphi_{\{(a_i)\}}$ are 0, so $F\{X\}/\mathcal{I}(\mathcal{V})$ can be embedded into a direct product of algebras from $\mathcal{V}$, and thus belongs to $\mathcal{V}$.)
30. (Amitsur) Prove that any algebra R satisfying a PI f of degree d satisfies the identity s_d^k for suitable k, where k depends only on f. (Hint: Replace R by its relatively free algebra, and let $N = N(R)$.

R/N is a semiprime PI-algebra and thus has some PI-class $n \leq [d/2]$, implying that s_{2n} is a PI of R/N, i.e., $s_{2n}(\bar{x}_1, \ldots, \bar{x}_{2n}) \in N$. But this means some power $s_{2n}(\bar{x}_1, \ldots, \bar{x}_{2n})^k = 0$.)

31. Verify the following identities in any (associative) algebra: $[x, yz] = [x, y]z + y[x, z]$, $[xy, z] = x[y, z] + [x, z]y$, $[x, yz] = [xy, z] + [zx, y]$.
32. Any identity of $M_n(R)$ can be translated to a set of n^2 identities of R, according to each matrix entry. Conclude that if algebras R_1 and R_2 are PI-equivalent, then so are $M_n(R_1)$ and $M_n(R_2)$.
33. Define **higher (Lie) commutators** inductively on length by saying $[f, g]$ is a higher commutator if f and g are letters or higher commutators of smaller length. A polynomial is called **Spechtian** if it is a sum of products of higher commutators. Show that every multilinear identity is a sum of Spechtian identities. (Hint: Replacing $x_m x_i$ by $x_i x_m + [x_m, x_i]$, show that any identity $f(x_1, \ldots, x_m)$ can be written as $f = g(x_1, \ldots, x_{m-1})x_m + h$, where x_m does not appear in g, and h is a sum of terms in each of which x_m appears in a Lie commutator. Specializing $x_m \mapsto 1$ shows that g, and thus h, are identities. Continue this argument in turn with each appearance of x_m, repeat the procedure with x_{m-1}, and so forth.)

Identities of Grassmann algebras

34. (Latyshev's Lemma.) Prove that the Grassmann identity $[[x_1, x_2], x_3]$ implies the identity

$$[x_1, x_2][x_3, x_4] + [x_1, x_3][x_2, x_4].$$

(Hint: Apply Exercise 31 repeatedly to $0 = [[ad, b]c]$.)

35. Prove that the Grassmann identity implies the identity $[x_1, x_2][x_1, x_3]$. (Hint: Specialize x_3 to x_1 in Exercise 34.)
36. In the relatively free algebra of the Grassmann identity, show using Exercise 35 that any Spechtian polynomial can be reduced to a product of commutators, i.e., $[x_{i_1}, x_{i_2}][x_{i_3}, x_{i_4}] \cdots$. Since these are not identities of the Grassmann algebra G of an infinite-dimensional vector space, conclude in characteristic 0 that the T-ideal $\mathrm{id}(G)$ is generated by the Grassmann identity.
37. For any completely homogeneous polynomial $f(x_1, \ldots, x_n)$, where $d_i = \deg_i f$, prove that there is a multilinear polynomial $g(x_1, \ldots, x_n)$ such that f equals $x_1^{d_1 - 1} \cdots x_n^{d_n - 1} g(x_1, \ldots, x_n)$ on the Grassmann algebra. (Hint: Use the Grassmann identity to rearrange the x_i in ascending order, leaving commutators on the right, and then Exercise 36 shows that any nonlinear part that remains is 0.)

In the next few exercises, due to Regev, write $G(p)$ for the infinite-dimensional Grassmann algebra in characteristic p.

38. Show that $\sum_{\pi \in S_p} x_{\pi 1} \cdots x_{\pi p} \in \mathrm{id}(G(p))$. (Hint: This is clear unless all substitutions are even, in which case one gets $p! x_1 \cdots x_p = 0$.)
39. Prove that $x^p \in \mathrm{id}(G(p))$. (Hint: Any evaluation reduces to Exercise 38.)
40. (Compare with Exercise 30.) Prove that $M_n(G(p))$ satisfies the identity $s_{2n}^{n^2 p + 1}$. (Hint: First translate to identities of G, using Exercise 32. In Exercise 37, one reduces to some x_i^p, so use Exercise 39.)
41. A T-ideal $\mathcal{I}$ of $F\{X\}$ is called T-**prime** if there do not exist T-ideals $A, B \supset \mathcal{I}$ such that $AB \subseteq \mathcal{I}$. Show that $\mathcal{M}_n$ is a T-prime T-ideal, as is the T-ideal of the Grassmann algebra G of an infinite-dimensional vector space. Other T-prime T-ideals include the identities of the superalgebras $M_{k,\ell}(F)$, defined to be $M_n(F)$ with the $\mathbb{Z}/2$-grade

$$\deg(e_{ij}) = \begin{cases} 0 & \text{if} \quad 1 \leq i, j \leq k \quad \text{or} \quad k+1 \leq i, j \leq k + \ell; \\ 1 & \text{if} \quad 1 \leq i \leq k < j \leq k + \ell \quad \text{or} \quad 1 \leq j \leq k < i \leq k + \ell. \end{cases}$$

These are gr-simple rings whenever F is a field.

42. Define a T-**ideal** of an arbitrary algebra R to be the set of evaluations on R of all polynomials of a given T-ideal of $F\{X\}$, and define T-**prime** as in Exercise 41. Viewing Kemer's Theorem in terms of the ACC on T-ideals, mimic the familiar theory of commutative Noetherian rings (Chapter 9 of Volume 1), and show that in any F-algebra, a suitable finite product of T-prime T-ideals is 0. Also prove that any T-ideal has only finitely many T-prime T-ideals minimal over it.

Young diagrams in PI-theory

In the following exercises, R is a given PI-algebra. I_λ is the simple component of $F[S_n]$ corresponding to the partition λ of n, $n_\lambda = f^\lambda$ is the number of standard Young diagrams for the partition λ, and h_{ij} denote the hook numbers; cf. Theorem 19.64 and the subsequent discussion.

43. If $n_\lambda > c_n(R)$, show that $I_\lambda \subseteq \mathcal{I}_n(R)$, notation as in Definition 23.55. (Hint: By a dimension count, every minimal left ideal of $F[S_n]$ contained in I_λ must intersect $\mathcal{I}_n(R)$ and thus lie in $\mathcal{I}_n(R)$. But I_λ is the sum of minimal left ideals.)
44. Notation as in Theorem 23.57, show that $n_\lambda > \left(\frac{uv}{u+v}\right)^n \left(\frac{2}{e}\right)^n$ when λ is a $u \times v$ rectangle and $n = uv$. (Hint: $\frac{1}{2}uv(u+v) = \sum h_{ij}$, and since the arithmetic mean is at least the geometric mean,

$$\frac{u+v}{2} \geq \sqrt[n]{\prod h_{ij}}.$$

Conclude by recalling $n! \geq (n/e)^n$ from calculus.)

45. When R satisfies a PI of degree d, show that, for any Young diagram T_λ of shape λ containing a $u \times v$ rectangle for $\frac{uv}{u+v} \geq \frac{1}{2}(d-1)^4 e$, the polynomials corresponding to semi-idempotents of T_λ are identities of R. (Hint: Combine Exercises 43 and 44.)

Appendix 23A

The Grassmann involution on polynomials

1. Show that $z_1 y z_2 + z_2 y z_1$ and $z_1 z_2 z_3 + z_3 z_2 z_1$ are superidentities of any supercommutative algebra.
2. Verify that $\varepsilon(\sigma\eta, I) = \varepsilon(\sigma, I)\varepsilon(\eta, \sigma^{-1}(I))$ for all $\sigma, \tau \in S_n$.

Superalgebras

3. Show that the tensor product $R \otimes T$ of two superalgebras is a superalgebra under the grading $(R \otimes T)_u = \bigoplus_{g+h=u}(R_g \otimes T_h)$.
4. Show that $V^n = 0$ in $E(V)$ whenever V is generated by n elements.
5. Define a gr-prime superalgebra as one in which the product of nonzero graded ideals must be nonzero. Verify that $M_n(E(V))$ is gr-prime if V is free of infinite rank.
6. Define supercentral polynomials via Example 23A.14. Kemer proved that every gr-prime PI-superalgebra in characteristic 0 is super-PI-equivalent to $M_n(\mathbb{Q})$, $M_n(E)$, or $M_{k,\ell}(\mathbb{Q})$. Prove that all of these have supercentral polynomials.

Appendix 23B

1. Verify that any simple alternative, nonassociative algebra satisfies the central polynomial $[x, y]^2$, by showing that it is PI-equivalent to the split octonion algebra. (Hint: When the center is infinite, extend scalars to the algebraic closure; show that one has the split octonion algebra, yielding the first assertion. For the second assertion, pass back to Example 21B.15(iv), in which every nonzero element is invertible, and note that any noncommutative subalgebra generated by two elements is associative and thus a quaternion algebra.)

The locally nilpotent radical

2. Show that any associative algebra A has a unique maximal locally nilpotent ideal N as defined in Definition 23B.9, and A/N has no nonzero locally nilpotent ideals. (Hint: One must show for any locally nilpotent ideal $(I + N)/N$ that $I + N = N$. But any finite subset of $I + N$ is nilpotent modulo N, and thus is nilpotent.)
3. Where does the argument of Exercise B2 fail for nonassociative algebras?

Lie identities

4. Show that any Lie algebra of characteristic 3 satisfying the Engel ad-identity $e_2 = X^2$ is Lie nilpotent of class ≤ 3. (Hint: Write out the ad-identity $X_1[X_2, X_3] + [X_2, X_3]X_1$ and rearrange the order to get $4X_1X_2X_3$.)
5. Let $\mathcal{F}$ denote the free group $\{x_0, x_1, \dots, \}$, and let H denote the normal subgroup generated by x_0 and the group commutators $(x_0, \mathcal{F})$. Let $\hat{H}'$ denote the subgroup of H generated by H' and $\{a^p : a \in H\}$. Noting that $V = H/\hat{H}'$ is a vector space over $\mathbb{Z}/p$, prove that $L_{\hat{\gamma}}(G)$ defined in (23B.5) is a restricted Lie algebra. Furthermore, show that any element $a \in \mathcal{F}$ induces a linear transformation ad_a of $\hat{H}$ satisfying $\mathrm{ad}_{a^n} = \mathrm{ad}_a^n$ for any p-power n.
6. For any nilpotent p-group G of exponent $n = p^k$, prove that $L_{\hat{\gamma}}(G)$ satisfies the multilinearized n-Engel identity $\tilde{e}_n$ and the weak Engel condition $e_{S,2n}$ for some generating set S. (Hint (Zelmanov): Notation as in Exercise B5, and letting (a, b) denote the group commutator in $\mathcal{F}$, modify Equation (17.2) to show that $\mathrm{End}\, V$ satisfies the ad-identities $(\mathrm{ad}_x + 1)(\mathrm{ad}_y + 1) = \mathrm{ad}(x, y) + 1$, $\mathrm{ad}_x^n = \mathrm{ad}_{x^n}$, and thus

$$\big((\mathrm{ad}_{x_1} + 1) \cdots (\mathrm{ad}_{x_n} + 1)\big)^n = \mathrm{ad}^n_{x_1 \cdots x_n} + 1.$$

   Specializing $x_1 \mapsto 1$, show that the sum of all ad-monomials containing X_1 is 0; next, specializing $x_2 \mapsto 1$, show that the sum of all ad-monomials also containing X_2 is 0, and so forth; conclude that the sum of all ad-monomials containing each of $X_1, X_2, \dots, X_n$ is 0. Interpret this in $L_{\hat{\gamma}}(G)$ to obtain the identity $\tilde{e}_n$; then verify $e_{S,2n}$ by substituting $X_1^n(x_0)$ in place of x_0.)

n-thick sandwiches

Suppose R is an enveloping (associative) algebra of a Lie algebra L. An element $a \in L$ is called an n-**thick sandwich** of the pair (L, R) if $a^2 = aa_1 \cdots a_m a = 0$, $\forall a_i \in L$, for all $m \leq n$. Given L and R, write $\mathbf{S}_n = \mathbf{S}_n(L, R)$ for the n-thick sandwiches of (L, R).

7. Suppose L is generated by n elements $u_1, \dots, u_n \in \mathbf{S}_1$; write $R(k) = L + L^2 + \cdots + L^k$. Show that $R = R(2n)$. (Hint: Let $L_i = \mathbb{Z}u_i + [L, u_i]$; then $L = \sum_{i=1}^n L_i$, and $L_i^3 = 0$ in R. It suffices to show that any product $a_1 \cdots a_{2n+1} \in R(2n)$, where $a_i \in L_{j_i}$ for suitable j_i. Note that three of the j_i are equal, but, mod $R(2n)$, one can rearrange the order of the a_i and make the three occurrences consecutive.)
8. Verify the following inclusions for any $a \in L$: $R(k)a \subseteq aR(k) + R(k)$, $R(k)a \subseteq R(1)aR(k\text{–}1) + aR(k) + R(k\text{–}1)$, and $R(k)a \subseteq R(2)aR(k\text{–}2) + R(1)aR(k\text{–}1) + aR(k) + R(k\text{–}2)$. Note the analogous version for $aR(k)$.

9. Prove that $[\mathbf{S}_k, \mathbf{S}_1] \subseteq \mathbf{S}_k$, and $[\mathbf{S}_k, \mathbf{S}_2] \subseteq \mathbf{S}_{k+1}$. (Hint: Apply Exercise B8 repeatedly.)
10. Prove that $a_1 \cdots a_k R(k) a_1 \cdots a_k = 0$ for any commuting elements $a_1, \ldots, a_k$ of $\mathbf{S}_1$. (Hint: By induction and Exercise B8.)
11. When L is generated by elements in $\mathbf{S}_1$, show that $\mathbf{S}_2$ is contained in the locally nilpotent radical of R; cf. Exercise B2. (Hint: One may assume that R has no locally nilpotent ideals, and needs to show that $\mathbf{S}_2 = 0$. By Exercise B7, each element of $\mathbf{S}_{2n}$ generates a square-zero ideal of R, so is zero. So by reverse induction it suffices to assume that $\mathbf{S}_{i+1} = 0$ for $i \geq 2$ and prove that $\mathbf{S}_i = 0$. But Exercise B9 shows that $[\mathbf{S}_i, \mathbf{S}_i] = 0$, so Exercise B10 implies $(R\mathbf{S}_i R)^2 = 0$.)
12. (Kostrikin) Prove that if $a_1, \ldots, a_4 \in \mathbf{S}_1$ such that, for all a in L,

$$c = \mathrm{ad}_{a_{\pi 1}} \mathrm{ad}_{a_{\pi 2}} \mathrm{ad}_{a_{\pi 3}} \mathrm{ad}_{a_{\pi 4}}(a) = \mathrm{ad}_{a_1} \mathrm{ad}_{a_2} \mathrm{ad}_{a_3} \mathrm{ad}_{a_4}(a), \quad \forall \pi \in S_4,$$

then $c \in \mathbf{S}_2$. (Hint: First note that $ca_i = a_i c = 0$ for $1 \leq i \leq 4$, implying that $caa_i = -a_i ac$ for all $a \in L$. Now apply Exercise B8 repeatedly to show that $cR(2)a_{i_1}a_{i_2}a_{i_3}$, $cR(2)a_{i_1}a_{i_2}aa_{i_3}a_{i_4}$, $cR(2)a_{i_1}aa_{i_2}a_{i_3}a_{i_4}$, and $cR(2)aa_{i_1}a_{i_2}a_{i_3}a_r$ are all zero.)
13. Suppose R is generated by elements $u_1, \ldots, u_n \in \mathbf{S}_1$ but is not nilpotent. Show that, for any m, there are Lie products b_j in the $u_1, \ldots, u_n$, $0 \leq j \leq m$, such that for each $k \in \mathbb{N}$,

$$b_0 b_{i_1} b_{i_2} b_0 b_{i_3} b_{i_4} b_0 \cdots b_0 b_{i_{2k-1}} b_{i_{2k}} b_0 \neq 0 \tag{E23B.1}$$

for suitable $1 \leq i_1, \ldots, i_{2k} \leq m$. (Hint: One may assume that R has no locally nilpotent ideals. Applying induction on n, one is done unless $u_1, \ldots, u_{n-1}$ generate a nilpotent subalgebra. Take some Lie product c_0 in $u_1, \ldots, u_{n-1}$ such that $[c_0, u_i] = 0$ for $1 \leq i \leq n-1$. Then $[c_0, u_n] \neq 0$. Take a nonzero Lie product $[c_0, u_n, u_{i_1}, \ldots, u_{i_m}]$ with m maximal, and then take Lie products $r_1, \ldots, r_t$ of $u_1, \ldots, u_{n-1}$ of total degree m with t maximal, such that

$$c_1 = [c_0, u_n, r_1, \ldots, r_t] \neq 0.$$

Then $[c_0, u_n, r_{\pi 1}, \ldots, r_{\pi t}] = c_1$ for any $\pi \in S_t$. If $t \geq 3$, then $c_1 \in \mathbf{S}_2$, by Exercise B12. But $\mathbf{S}_2 = 0$, by Exercise B11, a contradiction. Also $t \neq 1$. Thus, $t = 2$. Continue, opening up the expression in R, and take $b_0 = u_n$.)
14. Prove that R is nilpotent whenever R is generated by a finite subset of $\mathbf{S}_1$. (Hint: Otherwise take b_i satisfying (E23B.1). Let R_1 be the subring generated by all $[b_i, b_j, b_0]$, $1 \leq i, j \leq m$. Let

$$I_1 = \mathrm{Ann}_{R_1} b_0 = \{r \in R : rb_0 0 = 0\} \lhd R_1.$$

Then one can apply Exercise 13 to $\overline{R_1} = R_1/I_1$ to get elements $\overline{b'_i} \in \overline{R_1}$ satisfying $b'_0 b'_1 b'_2 b'_0 b'_3 b'_4 b'_0 \cdots b'_0 b'_{2k-1} b'_{2k} b'_0 b_0 \neq 0$. Let R_2 be the subring generated by the $[b'_i, b'_j, b'_0]$, and continue the procedure. By Exercises B7 and B10, some product $\cdots b''_0 b'_0 b_0$ generates a nilpotent ideal of R, contrary to hypothesis.)
15. Finish the proof of Theorem 23B.16. (Hint: By Exercise B14, the algebra generated by the adjoints of the sandwiches is nilpotent.)

Zelmanov's Theorem

16. (Zelmanov's version of Shirshov's Lemma.) Say an associative word in X is **special** if it is the leading word appearing in some Lie word. The word w is **Zelmanov d-decomposable** if it can be written as a product of subwords $w = w' w_1 w'_1 w_2 w'_2 \cdots w_d w'_d w''$ with each w_i special and $w_1 \succ w_2 \cdots \succ w_d$. Mimicking Theorem 23A.3, show for any ℓ, k, d that there is $\beta = \beta(\ell, k, d)$ such that any Zelmanov d-indecomposable word w of length $\geq \beta$ in ℓ letters must contain a nonempty subword of the form u^k, with u special.
17. Suppose the restricted Lie algebra L is f.g. over $\mathbb{Z}/p$ and satisfies the linearized Engel identity $\tilde{e}_m$ as well as the weak Engel identity $e_{S,n}$ for some m, n. Furthermore, assume, in a suitable associative enveloping algebra R of L for some $q \leq m$, that any elements $a_1, \ldots, a_q \in L$ satisfy $\sum_{\pi \in S_q} a_{\pi 1} \cdots a_{\pi q} = 0$. Prove that R_0, the associative subalgebra without 1 generated by L, is nilpotent. (Hint: Tensor R_0 by the algebra of dual numbers (cf. Exercise 4.1 of Volume 1), written as $(\mathbb{Z}/p)[\varepsilon_1, \varepsilon_2, \ldots]$ with each $\varepsilon_i^2 = 0$. By Exercise B16, any long enough nonzero word w in the generators is Zelmanov d-decomposable for d as in Corollary 23B.7. Write $w = w' w_1 w'_1 w_2 w'_2 \cdots w_d w'_d w''$; by considering $\hat{w} = \sum w_i \otimes \varepsilon_i$ and examining leading terms, show that $R_0^d = 0$.)

Chapter 24

1. Verify the following short, direct proof of Frobenius' Theorem that the only $\mathbb{R}$-cda other than R is $\mathbb{H}$. First note that $\mathbb{R}[a] \cong \mathbb{C}$ is a maximal subfield of D for any $a \in D \backslash \mathbb{R}$.

STEP I. Any two noncommuting elements $a, b \in D$ generate a copy $Q = \mathbb{R} \oplus \mathbb{R}i \oplus \mathbb{R}j \oplus \mathbb{R}ij$ of $\mathbb{H}$, where $i \in \mathbb{R}[a]$. (Hint: Taking $i \in \mathbb{R}[a]$ with $i^2 = -1$, one has $\mathbb{R}[i] = \mathbb{R}[a]$. Pick $b \notin \mathbb{R}[i]$ with $b^2 \in \mathbb{R}$. Let $c = bi + ib$. Then $c \in C_D(b) = \mathbb{R}[b]$ and thus $bi \in -ib + \mathbb{R}[b]$. Hence, $\mathbb{R}[i] + \mathbb{R}[i]b$ is a division algebra D_0 of dimension 4. Now let $d = bi - ib$. Then $di = -id$, implying that $\mathbb{R}[d^2] \subset \mathbb{R}[i]$ and thus $d^2 \in \mathbb{R}$.

But $d \notin \mathbb{R}$ implies $d^2 < 0$, so $j = \frac{d}{|d|}$ satisfies $ji = -ij$ and $j^2 = -1$. Thus, D_0 contains $Q = \mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}ij$, so $D_0 = Q$.)

STEP II. $D = Q$ of Step I. (Hint: By Step I, for any $d \in D \setminus Q$, i and d generate some algebra $Q' = \mathbb{R}+\mathbb{R}i+\mathbb{R}j'+\mathbb{R}ij'$ with $ij' = -j'i$. Thus, $j'j^{-1} \in \mathbb{R}[i]$, implying that $d \in Q$, a contradiction.)

Cyclic algebras

2. Verify that the cyclic algebra (K, σ, β) is simple, using skew polynomial rings. (Hint: If $\langle \lambda^n - \beta \rangle \subset I \lhd K[\lambda; \sigma]$, then $I = K[\lambda; \sigma]f$, where $f \in \mathrm{Cent}(K[\lambda; \sigma])$ divides $\lambda^n - \beta$.)
3. Verify that the symbol algebra $(3, -1)_2$ over $\mathbb{Q}$ is a division algebra. (Hint: 3 is not a sum of two squares in $\mathbb{Q}$, since $3m^2$ is not a sum of two squares in $\mathbb{Z}$ for any $m > 0$.)
4. Generalizing Exercise 3, show that $(4k+3, -1)_2$ is a division algebra for any integer k.
5. Verify that $(\alpha, \beta; \zeta)_n \sim (\alpha, \beta; \zeta^{-1})_n^{\mathrm{op}}$. (Hint: If $zy = \zeta yz$, then $yz = \zeta^{-1}zy$.)
6. Prove that $(\alpha, \beta)_{mn}^{\otimes m} \sim (\alpha, \beta)_n$. (Hint: If $zy = \zeta yz$, then $zy^m = \zeta^m y^m z$.) Conclude that any symbol has an m-th root that also is a symbol, presuming F has enough roots of 1.
7. Show that the symbol $(\alpha, 1-\alpha)_n \sim 1$ for any $\alpha \in F$. More generally, $(\alpha, \beta)_n \sim (\pm\frac{\alpha}{\beta}, \alpha+\beta)_n$. (Hint: If $zy = \rho yz$, consider yz^{-1} and $y + z$.)
8. (Wedderburn's criterion.) Prove that a cyclic algebra $R = (K, \sigma, \beta)$ of degree n is a division algebra of exponent n, if β^j is not a norm from K for all $1 \leq j < n$. (Hint: $R^{\otimes j}$ is not split for these j, so $\exp(R) = n$.)

The Koethe-Noether-Jacobson Theorem

Recall the Lie terminology ad_a given by $\mathrm{ad}_a(r) = [a, r] = ar - ra$.

9. Prove that $\mathrm{ad}_a^n(b) = \sum_{i=0}^n (-1)^i \binom{n}{i} a^{n-i} b a^i$. In particular, in odd characteristic p, $\mathrm{ad}_a^p(b) = a^p b - b a^p$. (Hint: As in the proof of Leibniz' rule (Proposition 6B.5 of Volume 1).)
10. Suppose $ab - ba = 1$ in a cda D. Show that $F[ba]$ has a nontrivial automorphism over F. (Hint: $a(ba)a^{-1} = ab = ba + 1$.)
11. Reprove the Koethe-Noether-Jacobson Theorem. (Hint: Assume that every element in $D \setminus F$ is purely inseparable. Take $a \in D \setminus F$ such that $a^p \in F$. By Exercise 10, it suffices to find b such that $ab - ba = 1$. So take d with $ad \neq da$. Then $\mathrm{ad}_a^p = 0$ by Exercise 9, so there is some $m < p$ such that $c = \mathrm{ad}_a^m d \neq 0$ but $\mathrm{ad}_a c = 0$. Let $b = \mathrm{ad}_a^{m-1}(d)c^{-1}$; then $\mathrm{ad}_a(b) = 1$.)

Properties of the Brauer group

12. In Theorem 24.57, show that L_u is a maximal subfield of $D \otimes_F L$.

13. When $\mathrm{ind}(A) = pm$ for a given prime number p, show that $\mathrm{ind}(A^{\otimes p})$ divides m. (Hint: Application of Lemma 24.61.)
14. Show that $[R] \mapsto [R]^k$ yields an automorphism of $\mathrm{Br}(F)_m$, for each $k \in \mathbb{N}$ relatively prime to m.

Applications of the reduced norm and trace; cf. Appendix 4B of Volume 1

15. Show that the reduced norm $\mathrm{N}_{\mathrm{red}}(\widehat{d})$ of a generic element $\widehat{d} = \sum_{i=1}^{n^2} \lambda_i b_i$ (computed on elements of D as a function in $\lambda_1, \ldots, \lambda_{n^2}$) defines a non- degenerate (homogeneous) form over F of degree n, on n^2 variables.
16. Prove that the Brauer group of a C_1 field (cf. Exercise 4B.17ff. of Volume 1) is trivial. (Hint: It suffices to find a nonzero element whose reduced norm is 0; use Exercise 15, noting that $n^2 > n$ for all $n > 1$.)
17. Using Exercise 16 and Exercises 4B.18 and 4B.24 of Volume 1, prove Wedderburn's Theorem (Theorem 24.42) and Tsen's Theorem, that there are no noncommutative cda's over $\mathbb{C}(\lambda)$ (or more generally, over any field of transcendence degree 1 over $\mathbb{C}$).
18. For any csa R of degree n over an infinite field F, prove that the set of separable elements of R having degree n is Zariski dense. (Hint: View the generic element $\hat{r}$ in $M_{n^2}(F(\Lambda))$. The fact that $\deg \hat{r} = n$ means that writing $\hat{r}^i = \sum_{j=1}^{n^2} h_{ij} b_j$, for $h_{ij} \in F(\Lambda)$, $0 \leq i < n$, the $n \times n^2$ matrix (h_{ij}) has a nonsingular $n \times n$ minor. Specializing the h_{ij} to elements of F shows that there is a nonzero Zariski open subset of R for which the corresponding minor is nonzero, and thus $\deg r = n$ for these specializations r. The same argument holds for separability.)
19. Show that any cda D has an element $d \neq 0$ for which $\mathrm{tr}_{\mathrm{red}}(d) = \mathrm{tr}_{\mathrm{red}}(d^{-1}) = 0$. (Hint (Haile): Take a_0 separable of degree n and $b \in D$ with $a_0 b \neq b a_0$. Define $f\colon F[a_0] \to F$ by $a \mapsto \mathrm{tr}_{\mathrm{red}}\, a[a_0, b]^{-1}$. Take any $0 \neq a \in \ker f$ and put $d = a[a_0, b]^{-1} = [a_0, ba^{-1}]^{-1}$.)
20. (Structure of division algebras of degree dividing 12.) Let $n = \deg D$; for convenience, assume that F contains a primitive n-th root of 1. Prove that D is cyclic for $n = 2$ or 3. For $n = 4$, prove that D has a maximal subfield K with Galois group $C_2 \times C_2$. Applying Theorem 24.66, conclude that every division algebra of degree dividing 12 is a crossed product. (Hint: For $n = 3$, this follows at once from Exercise 19 and Newton's identities. For $n = 4$, Exercise 19 gives an element $d \in D$ such that d^2 is quadratic, and thus D has an element $a \notin F$ with $a^2 \in F$. Take $b \in D$ with $ba = -ab$. If $b^2 \in F$, then a and b generate a quaternion subalgebra, so use the Double Centralizer Theorem. If $b^2 \notin F$, then $K = F[a, b^2]$.)

21. Prove that every csa is generated as an algebra by two elements, and in fact there are a and b such that $\{a^i b^j : 0 \le i, j \le n-1\}$ are a base. (Hint: As in Proposition 24.53.)

Crossed products

22. Suppose a csa R of degree n has a maximal subfield K that is Galois over F, having Galois group G. Show that R is a crossed product. (Hint: The Skolem-Noether Theorem implies that for each $\sigma \in G$ there is some $z_\sigma \neq 0$ in D such that $z_\sigma a z_\sigma^{-1} = \sigma(a)$ for all a in K, i.e., $z_\sigma a = \sigma(a) z_\sigma$. It follows that $\{z_\sigma : \sigma \in G\}$ is a base of R over K. Show that $z_\sigma z_\tau (z_\sigma \tau)^{-1}$ commutes with K, so $z_\sigma z_\tau (z_\sigma \tau)^{-1} = c_{\sigma,\tau}$ for some $0 \neq c_{\sigma,\tau} \in K$, i.e., $z_\sigma z_\tau = c_{\sigma,\tau} z_{\sigma\tau}$. Now follow Example 24.45.)
23. The factor set $\{c_{\sigma\tau} : \sigma, \tau \in G\}$ of Example 24.45 is said to satisfy the **coboundary condition** if there are $b_\sigma \in K^\times$ such that $c_{\sigma\tau} = b_\tau \tau(b_\sigma)$ for all $\sigma, \tau \in G$; show in this case that $(K, G, f) \sim 1$. (Hint: Replacing z_σ by $b_\sigma^{-1} z_\sigma$, one may assume that $c_{\sigma,\tau} = 1$, $\forall \sigma, \tau$.)
24. (Saltman) Show that $R = \bigoplus_{\tau \in G} K z_\tau$ is a crossed product iff $\bigcup K^\times z_\tau$ is a group.
25. Prove that $(K, G, (c_{\sigma,\tau})) \otimes (K, G, (d_{\sigma,\tau})) \sim (K, G, (c_{\sigma,\tau} d_{\sigma,\tau}))$ in $\mathrm{Br}(F)$. (Hint: Use the argument of Proposition 24.48′.)

Characteristic p

Assume throughout that $\mathrm{char}\, F = p > 0$.

26. Suppose $\deg R = p$, and R has a maximal subfield E that is cyclic Galois over F. By Exercise 4.70 of Volume 1, one can write $E = F[a]$ such that $a^p - a = \alpha \in F$ and $\sigma(a) = a + 1$; take $z \in R^\times$ with $zaz^{-1} = a + 1$, and let $\beta = z^p \in F$. Then R is denoted as the **Artin-Tate** symbol $[\alpha, \beta)$. Conversely, given any α, β in F, construct the algebra having Artin-Tate symbol $[\alpha, \beta)$, as follows: Take a field extension $E = F[a]$ with $a^p - a = \alpha$, and define $R = \bigoplus_{i=0}^{p-1} K z^i$, where $z^p = \beta$ and $za = (a + 1)z$. This is a simple algebra of degree p.
27. Prove that $[\alpha_1 + \alpha_2, \beta) \sim [\alpha_1, \beta) \otimes [\alpha_2, \beta)$. $[\alpha, \beta_1 \beta_2) \cong [\alpha, \beta_1) \otimes [\alpha, \beta_2)$.
28. Show that $[\alpha^p - \alpha, \beta) \sim 1$; hence $[\alpha^p, \beta) \sim [\alpha, \beta)$.
29. Show that $[\alpha, \beta)^{\otimes p} \sim 1$.
30. Letting σ be the Frobenius map $a \mapsto a^p$, show that $\sigma(R) \sim R^{\otimes p}$ for any p-algebra R. (Hint: Assume that R is a crossed product, and then match the structure constants.)
31. Fixing an algebraic closure $\bar{F}$ of F, define $F^{1/p} = \{a \in \bar{F} : a^p \in F\}$, and inductively define $F^{1/p^t} = (F^{1/p^{t-1}})^{1/p}$. Show that for any p-algebra R there is some t such that F^{1/p^t} is a splitting field for R and $\exp(R)$ is p^t for the minimal such t. Thus, any p-algebra is split by a purely inseparable, finite-dimensional field extension of F. (Hint: Exercise 30.)

Constructions of division algebras

32. If $\mathrm{UD}(n, F)$ is a crossed product with respect to the group G, show that the elements defining the crossed product conditions can be specialized to any csa of degree n over F, and thus every csa of degree n over F is a crossed product with respect to G. (Hint: Translate the assertion to equations involving specific elements of $\mathrm{UD}(n, F)$, and then appeal to Remark 24.55.)
33. (The "generic" tensor product of two quaternion algebras.) Taking indeterminates $\lambda_1, \lambda_2, \mu_1, \mu_2$ over a field F_0 of characteristic $\neq 2$, let $F = F_0(\lambda_1, \lambda_2, \mu_1, \mu_2)$, and let $D = (\lambda_1, \mu_1)_2 \otimes_F (\lambda_2, \mu_2)_2$. Take y_i, z_i with $y_i^2 = \lambda_i$, $z_i^2 = \mu_i$, and $y_i z_i = -z_i y_i$. Show that the F_0-subalgebra W of D generated by the y_i and z_i is a skew polynomial ring, without zero divisors, and thus D is a division algebra. Also, the maximal subfield $F(y_1, y_2)$ is Galois over F with Galois group $C_2 \times C_2$. On the other hand, D is not cyclic. (Hint: If D were cyclic, there would be a cyclic maximal subfield K of D and $z \in W$ such that $zaz^{-1} = \sigma(a)$ for all $a \in K$. Thus, $z^2 a = \sigma^2(a) z^2$, so one can choose a such that $z^2 a = -az^2$ and assume that $a \in W \cap K$. Taking the "leading monomials" under the lexicographic order, one could assume that a, z are monomials. But then $z^2 \in F$, a contradiction.)
34. Notation as in Example 24.5, for $n = p^2$, show that the generic symbol $(\mu^n, \lambda^n)_n$ is also a crossed product with respect to $G = C_p \times C_p$, since $F_1(\mu^p, \lambda^p)$ is a maximal subfield with Galois group G.
35. For any prime powers m, n of p, where $m | n$, show that a tensor product of suitable generic symbols is a division algebra of degree n and exponent m.
36. Using the idea of Exercise 33, show that if a tensor product of generic symbols (using distinct commuting indeterminates λ_i) is a crossed product with respect to G, then the elements used to express this via Exercise 35 may be taken to be monomials in the λ_i. Verify the assertions of Example 24.56 by means of this observation.
37. For $\mathrm{char}(F) \neq 2$, prove that the center of $\mathrm{UD}(2, F)$ (i.e., taking $Y = \{y_1, y_2\}$) is purely transcendental over F, generated by the elements $a_1 = \mathrm{tr}(y_1)$, $a_2 = \mathrm{tr}(y_2)$, $a_3 = \mathrm{tr}(y_1 y_2)$, $a_4 = \det(y_1)$, and $a_5 = \det(y_2)$. (Hint: Let $L = F(a_1, \ldots, a_5)$, which has transcendence degree 5. Then $\det(y_1 y_2) \in L$ by Newton's formulas. Let V be the L-subspace of $\mathrm{UD}(2, F)$ spanned by $1, y_1, y_2, y_1 y_2$. Then $y_1^2, y_2^2, (y_1 + y_2)^2 \in V$, so

$$y_1 y_2 + y_2 y_1 = (y_1 + y_2)^2 - y_1^2 - y_2^2 \in V,$$

and it follows that V is an L-subalgebra of $\mathrm{UD}(2,F)$ containing y_1 and y_2, and so is all of $\mathrm{UD}(2,F)$.)

38. Show that division algebras of degree n exist in any characteristic, for any n. (Hint: Using Corollary 24.65, assume that n is a prime power, and separate the cases $p \nmid n$ and $p \mid n$.)

Wedderburn's Factorization Theorem

39. (Wedderburn's original computation.) When $f(\lambda) = h(\lambda)g(\lambda)$, with $g(d) \neq 0$, show that $g(d)dg(d)^{-1}$ is a root of h.
40. When d_1 is a root of $f = (\lambda - d_3)(\lambda - d_2)(\lambda - d_1) \in F[\lambda]$, show that $d_3 = [d_2, d_1]d_1[d_2, d_1]^{-1}$. (Hint: d_2 is not a root of $g = (\lambda - d_2)(\lambda - d_1)$ since $g(d_2) = [d_1, d_2]$; apply Exercise 39 with $h = \lambda - d_3$.)
41. Using Wedderburn's Factorization Theorem, reprove that any division algebra D of degree 3 is cyclic. (Hint: Take $d \in D$, and factor its minimal polynomial $f_d \in F[\lambda]$ as $(\lambda - d_3)(\lambda - d_2)(\lambda - d_1)$, where $d_1 = d$. Let $b = [d_1, d_2]$. Thus, $d_3 = bd_2b^{-1}$. But $d_3 + d_2 + d_1 = \mathrm{tr}_{\mathrm{red}}\, d$; hence, $[d_3, d_2] = -[d_1, d_2] = -b$.

    Also $f = (\lambda - d_1)(\lambda - d_3)(\lambda - d_2)$, implying that $d_1 = bd_3b^{-1}$. Hence, $b(d_3d_2^{-1})b^{-1} = -(d_3 + d_2)d_3^{-1} = -1 - d_2d_3^{-1}$; so $K = F[d_3d_2^{-1}]$ has a nontrivial automorphism σ over F given by conjugation by b and thus is cyclic. Conclude that $D \cong (K, \sigma, b^3)$.)
42. When $\deg D = 3$, show that any element d of reduced norm 1 is a multiplicative commutator. (Hint: Notation as in Exercise 41, $b^3 = (db)^3$, so by the Skolem-Noether Theorem there is $a \in D$ such that $db = aba^{-1}$.)
43. Show that any element $d \in D$ of reduced trace 0 is an additive commutator when $\deg D = 3$ and $\mathrm{char}(F) \neq 3$. (Hint: Same philosophy as Exercise 42, but now using $d_1 + d_2 + d_3 = 0$.)
44. Prove that there are no noncommutative ordered cda's. (Hint: In view of Wedderburn's Factorization Theorem, an element of reduced trace 0 is neither positive nor negative, since a finite sum of conjugates is 0.)
45. Suppose that D is a cda over F, and $p \in D[\lambda]$ is irreducible in $D[\lambda]$ of degree m. Write $\frac{1}{p} = \frac{q}{c}$ for $q \in D[\lambda]$ and $c \in F[\lambda]$. The polynomial c of minimal degree is called the **minimal polynomial for** p. Show that c is irreducible in $F[\lambda]$ and divides the reduced norm of p (as a polynomial in $F[\lambda]$).
46. Notation as in Exercise 45, show that $\bar{D} = D[\lambda]/cD[\lambda]$ is a simple algebra with center $F[\lambda]/F[\lambda]c$ and $M = D[\lambda]/p$ is a simple $\bar{D}$-module; also, $m \mid \deg c$. (Hint: Take a Jordan-Holder decomposition of $\bar{D}$.)

Corestriction and the Merkurjev-Suslin Theorem

47. Let L/F be a separable field extension with normal closure E. Take a transversal $\{\sigma_1, \ldots, \sigma_t\}$ of $H = \mathrm{Gal}(E/L)$ in $G = \mathrm{Gal}(E/F)$. Given a

csa R over L, extend each $\sigma \in G$ to an automorphism of $\hat{R}$ as follows: Write $\sigma_i\sigma = \tau_i\sigma_{\pi i}$ for $\tau_i \in H$, and define

$$\sigma\Big(\bigotimes_i (r_i \otimes \gamma_i)\Big) = \bigotimes_{i=1}^{t} (r_{\pi i} \otimes \tau_i(\gamma_{\pi i}))\,.$$

Then define $\mathrm{cor}_{L/F}(R)$ to be the fixed subalgebra $\widehat{R}^G$. Show that $\mathrm{cor}_{L/F}\colon \mathrm{Br}(L) \to \mathrm{Br}(F)$ is a well-defined homomorphism. (Hint: First show that a different transversal of $\mathrm{Gal}(E/L)$ in $\mathrm{Gal}(E/F)$ would provide the same result. Next show that if $\tilde{E}$ is another Galois extension of L containing F, then by taking the compositum of E and $\tilde{E}$, one could assume that $E \subseteq \tilde{E}$, and then $\mathrm{Gal}(\tilde{E}/E) \triangleleft \mathrm{Gal}(\tilde{E}/F)$, so again one gets the same result.)

48. If $R = (a, b; L)_n$, then $\sigma(R) \cong (\sigma(a), \sigma(b))$. For $a \in F$, conclude the **Projection Formula** $\mathrm{cor}_{L/F}(R) \sim (a, N_{L/F}(b))$.
49. Assume that F contains a primitive p-th root of 1. Prove that any cda D of degree p is similar to the corestriction of a symbol algebra having center L, for a suitable separable extension L of F whose dimension divides $(p-1)!$. (Hint: Take a maximal separable subfield K, whose Galois closure is E. Thus, $[E:F] = pm$ where $m|(p-1)!$. E has a subfield L of dimension m over F, with $D \otimes_F L$ cyclic of degree p, and thus is a symbol. Corestrict from L to F, applying Theorem 24.74.)
50. Assume that F has primitive p^k-roots of 1 for all k. Show that once we know that the Merkurjev-Suslin Theorem holds for all csa's of exponent p, then it holds for all algebras of index p^k, for all k. (Hint: Induction on $p^m = \exp(R)$; $\exp(R^{\otimes p}) = p^{m-1}$, so, by induction, $R^{\otimes p}$ is similar to a tensor product of symbols $(\alpha_1, \beta_1)_{m_1} \otimes \cdots \otimes (\alpha_t, \beta_t)_{m_t}$, where each $\alpha_i, \beta_i \in F$ and m_i is a p-power. Then the central simple algebra $R \otimes (\alpha_1, \beta_1)^{\mathrm{op}}_{pm_1} \otimes \cdots \otimes (\alpha_t, \beta_t)^{\mathrm{op}}_{pm_t}$ has exponent p, so by hypothesis is similar to a tensor product of symbols of degree p.)
51. Show that $\mathrm{Br}(F)$ is divisible whenever F has m-roots of 1 for all m. (Hint: Apply Merkurjev-Suslin to Exercise 6.)

Division algebras with valuation

52. Show that the value group $\mathcal{G}_D$ of D is Abelian whenever D is algebraic over its center F and has a valuation v. (Hint: First note that $\mathcal{G}_F$ is in the center of $\mathcal{G}_D$. For any $a, b \in D$, taking $\sum \alpha_i a^i = 0$ shows that $v(\alpha_i a^i) = v(\alpha_j a^j)$ for suitable $i < j$, implying that $(j-i)v(a) \in \mathcal{G}_F$. But then $(j-i)v(bab^{-1}) = v(ba^{j-i}b^{-1}) = (j-i)v(a)$, implying that $v(bab^{-1}) = v(a)$.)
53. Suppose the field F is complete with respect to a discrete valuation v, and D is a finite-dimensional division algebra over F. Extending v

to D, show that the value group $\mathcal{G}_D$ is cyclic and hence is generated by $v(\pi)$ for some $\pi \in D$.

54. Prove the inequality $e(D/F)f(D/F) \leq [D:F] = n^2$, equality holding when the valuation is discrete and the field F is complete with respect to it. (Hint: Follow the standard commutative proof; cf. Propositions 12.54 and 12.58 of Volume 1. Namely, pick $a_1, \ldots, a_e \in \mathcal{O}_D$ such that their values are in distinct coset representatives of Γ_D with respect to Γ_F, and $b_1, \ldots, b_f \in \mathcal{O}_D$ such that their images in $\bar{D}$ are linearly independent over F. First show that the $a_i b_j$ are linearly independent over F, by considering $0 = \sum \alpha_{ij} a_i b_j = \sum_i \left(\sum_j \alpha_{ij} b_j \right) a_i$; each summand is 0, so each coefficient is 0, yielding the first assertion. For the second assertion, take a generator π of Γ_D and show that $\{\pi^i b_j : 0 \leq i < e, 1 \leq j \leq f\}$ spans E over F, by inductively building a series converging to any element of E.)
55. Suppose v is a discrete valuation of a finite-dimensional division algebra D over F. Show that the generator π of $\mathcal{G}_D$ can be taken to be any element of $P_D \setminus P_D^2$. (Indeed, $P_D = V_D \pi$, so any element $a \in P_D$ can be written as $b\pi$, where $b \in V_D$; $v(a) = v(\pi)$ iff $b \notin P_D$.)
56. In Remark 24.84(ii) and thus in Theorem 24.85, show that ζ can be taken to be a root of 1. Also, K contains an isomorphic copy of every unramified field extension of F inside D. (Hint: $\bar{\zeta}$ is a root of 1, so apply Hensel's Lemma.)
57. Notation as in Theorem 24.85, take $\pi \in P \setminus P^2$. Show that the automorphism of $\bar{D}$ induced by conjugation by π is given by $\bar{\zeta} \mapsto \bar{\zeta}^q$ for some q. Hence, one can replace π by $\pi' = \sum_{j=1}^{m-1} \zeta^{-jq} \pi \zeta^j$ and conclude that $\pi \zeta \pi^{-1} = \zeta^q$.
58. (An explicit description of the cyclic algebra (K, σ, α) obtained in Theorem 24.85) For π as in Exercise 57, show that any element d of D can be written as an infinite sum $\sum_{i \geq m} \alpha_{ij} \zeta_i \pi^j$, where ζ_i is some power of ζ. Conclude from this that $D = \sum_{j=0}^{n-1} K \pi^j$. Thus, $D = (K, \sigma, \pi^n)$. (Hint: As in Theorem 12.65 of Volume 1.)

Appendix 24A

Plücker coordinates

1. Given a subspace W of V and $v \in V$, show that $v \in W$ iff $\omega_W v = 0$ in the Grassmann algebra $E(V)$.
2. Fix a base $\{e_j : 1 \leq j \leq n^2\}$ of V. Given any $\omega \in E^d(V)$, write

$$\omega = \sum_{j_1 < \cdots < j_d} \alpha_{\mathbf{j}} e_{j_1} \cdots e_{j_d},$$

where $\mathbf{j} = (j_1, \ldots, j_d)$, and call the $\alpha_{\mathbf{j}}$ the **Plücker coordinates** of ω. So far $\alpha_{\mathbf{j}}$ is defined only for $j_1 < \cdots < j_d$. For arbitrary $j_1, \ldots, j_d$ we define $\alpha_{\mathbf{j}} = 0$ if two of the j_u are the same; if $j_1, \ldots, j_d$ are distinct, pick the permutation π such that $\pi(j_1) < \pi(j_2) < \cdots < \pi(j_d)$ and put $\alpha_{\mathbf{j}} = \operatorname{sgn}(\pi) \alpha_{\pi(j_1), \ldots, \pi(j_d)}$. For any base $\{b_1, \ldots, b_d\}$ of W where $b_i = \sum_{i=1}^{n^2} \alpha_{ij} e_j$, show that the Plücker coordinates of ω_W are

$$\alpha_{j_1, \ldots, j_d} = \begin{vmatrix} \alpha_{1j_1} & \cdots & \alpha_{1j_d} \\ \vdots & \ddots & \vdots \\ \alpha_{dj_1} & \cdots & \alpha_{dj_d} \end{vmatrix}.$$

3. Given $i_1, \ldots, i_{d-1}$, show that the vector $v = \sum_j \alpha_{i_1, \ldots, i_{d-1}, j} e_j$ lies in W, by direct computation.
4. Show that a vector $v = \sum \beta_j e_j \in V$ lies in a subspace W iff

$$\sum_{u=1}^{d} (-1)^u \beta_{j_u} \alpha_{j_1, \ldots, j_{u-1}, j_{u+1}, \ldots, j_{d+1}} = 0$$

for each $j_1 < \cdots < j_d$, where the α denote the Plücker coordinates of ω_W.

5. Plug Exercise A3 into Exercise A4 to obtain the **Plücker equations**

$$\sum_{u=1}^{d} (-1)^u \alpha_{i_1, \ldots, i_{d-1}, j_u} \alpha_{j_1, \ldots, j_{u-1}, j_{u+1}, \ldots, j_{d+1}} = 0$$

for all $1 \leq i_1 < \cdots < i_{d-1} \leq n^2$ and $1 \leq j_0 < \cdots < j_d \leq n^2$.

6. Picking $i_1, \ldots, i_d$ suitably and defining

$$v_k = \sum_j \alpha_{i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_{d-1}, j} e_j,$$

cf. Exercise A3, show that $v_1, \ldots, v_d$ is a base of W.

7. Using Exercise A6 to reverse the argument of Exercise A5, show that the homogeneous polynomials

$$f_{\mathbf{i},\mathbf{j}} = \sum_{u=1}^{d} (-1)^u \lambda_{i_1, \ldots, i_{d-1}, j_u} \lambda_{j_1, \ldots, j_{u-1}, j_{u+1}, \ldots, j_{d+1}}$$

(of degree $2d$) define a projective subvariety of $\mathbb{P}(E^d(V))$.

8. (A geometric criterion of when the n-dimensional subspace W of a csa R is a left ideal.) Take a base B of R over F consisting of invertible elements. Define the action of B on $\mathbb{P}(E^n(V))$ by

$$b(w_1 \cdots w_n) = (bw_1) \cdots (bw_n).$$

Prove that W is a left ideal iff $b\omega_W \in F\omega_W$, $\forall b \in B$. Describe this in terms of equations in the Plücker coordinates.

Chapter 25

Modules that are not projective

1. Verify the following example of a non-projective module over an Artinian ring: $R = \mathbb{Z}/4$ and $M = \mathbb{Z}/2$.
2. Prove that $\mathbb{Q}$ is not projective as a $\mathbb{Z}$-module.
3. Suppose R is any ring, and $s \in R$ is noninvertible. If the module R/Rs is projective, conclude that $Rs = Re$ for an idempotent e. (Hint: The exact sequence $0 \to Rs \to R \to R/Rs \to 0$ splits, yielding a projection $\pi\colon R \to Rs$.)
4. Suppose R is any ring, and $s \in R$ is a regular, noninvertible element. Show that $Rs \cong R$ is a free R-module, but R/Rs cannot be projective. (Hint: Exercise 3.)

Hereditary rings

A ring R is called **hereditary** if each left ideal is a projective as an R-module.

5. Verify that any PLID is hereditary and that the ring of upper triangular matrices over a division ring is hereditary.
6. Prove that over a hereditary ring R, every submodule of a projective module is projective. More precisely, every submodule of a free module F is isomorphic to a direct sum of left ideals. (Hint: Take a base $\{b_i : i \in I\}$ of F, let π_i denote the projection $F \to R$ given by $\sum_j r_j b_j \mapsto r_i$, and well-order the set I. For any $M \leq F$, write $M_i = M \cap (\sum_{j<i} Rb_j)$, and let f_i denote the restriction of π_i to M_i. Let $L_i = f_{i+1}(M_{i+1}) \leq R$. Then $\ker f_{i+1} = M_i$, so $M_{i+1} \cong M_i \oplus L_i$. Conclude that $M \cong \bigoplus L_i$ by transfinite induction.)

Nonfree projective modules

7. Prove that a fractional ideal P of an integral domain C is invertible (as a fractional ideal) iff P is projective as a module. (Hint: Write $P = \frac{1}{s}A$ for $A \lhd C$, and take $0 \neq a \in A$. Any module homomorphism $f_i\colon P \to C$ is given by multiplication by $f_i(\frac{a}{s})\frac{s}{a}$; now reverse the argument of Example 25.17(i).)
8. Show that a projective module P is principal iff P is a direct summand of R.
9. (Bourbaki) Suppose $q \in F[x]$ is separable quadratic with constant term 0, and consider the elliptic curve $f = y^2 - xq(x)$ and its coordinate algebra $R = F[x,y]/\langle f\rangle$. Show that R is an integral domain whose invertible elements are precisely $F^\times$. (Hint: Let $\bar{x}, \bar{y}$ denote the images of x, y in R. R has an automorphism σ given by $\bar{x} \mapsto \bar{x}$ and $\bar{y} \mapsto -\bar{y}$. Any element $a \neq 0$ can be written in the form $h_1(\bar{x}) + h_2(\bar{x})\bar{y}$; then $a\sigma(a) = h_1(\bar{x})^2 - h_2(\bar{x})^2\bar{y}^2 = h_1(\bar{x})^2 - h_2(\bar{x})^2\bar{x}q(\bar{x})$, which is nonconstant in $F[\bar{x}] \cong F[x]$, and is nonzero iff $a \neq 0$.)

10. Notation and assumptions as in Exercise 9, verify that the ideal $P = \langle \bar{x}, \bar{y}\rangle$ of R is invertible and thus projective, but not principal. (Hint: $\bar{x}^2 \in P^2$ and $\bar{x}q(\bar{x}) \in P^2$ implying that $\bar{x} \in P^2$. If P were principal, say $P = Ra$, then $a^2 = u\bar{x}$ for $u \in R$ invertible, so $u \in F$, implying that $u^2\bar{x}^2 = a^2\sigma(a)^2$, and thus $\pm u\bar{x} = a\sigma(a) = h_1^2 - \bar{x}\bar{q}h_2^2$ for suitable $h_i \in F[\bar{x}]$, which is impossible.)
11. (Kaplansky) Let C be the ring of continuous real-valued functions on the real interval $[0,1]$, and P be the ideal of functions vanishing close to 0. Show that P is a faithful and projective, but is not free.

The rank function on projective modules

In the following exercises, assume that P_i are f.g. projective modules over a commutative ring C. When in doubt, use the fact that a local isomorphism is an isomorphism; cf. Exercise 8.30 of Volume 1.

12. Show that $\operatorname{rank}(P_1 \oplus P_2) = \operatorname{rank} P_1 + \operatorname{rank} P_2$. In particular, $C^{(n)}$ has constant rank n.
13. Verify that $\operatorname{rank}(P_1 \otimes P_2) = \operatorname{rank}(\operatorname{Hom}_C(P_1, P_2)) = \operatorname{rank}(P_1)\operatorname{rank}(P_2)$.
14. Prove that $\operatorname{rank} P^* = \operatorname{rank}(P)$.
15. Show that rank is continuous with respect to the Spec topology on C and the discrete topology on $\mathbb{N}$. (Hint: One needs to show that if $\operatorname{rank}(P) = n$, then, taking suitable s such as in Exercise 55 below, $\operatorname{rank}(P_{\mathfrak{q}}) = n$ whenever $s \notin \mathfrak{q}$.)
16. Conclude from Exercise 15 that every f.g. projective module is isomorphic to a direct sum of projective modules of constant rank.
17. Prove that the following conditions are equivalent for a C-module P: (i) P is invertible; (ii) $P \otimes_C P^* \cong C$; (iii) there is a C-module Q such that $P \otimes_C Q \cong C$; (iv) P is projective of constant rank 1; (v) P is projective and $\operatorname{End}_C P \cong C$; (vi) P is f.g. and $P_{\mathfrak{m}} \cong C_{\mathfrak{m}}$ for each maximal ideal $\mathfrak{m}$ of $\operatorname{Spec} C$. (Hint: (iii) $\Rightarrow$ (i) by means of Proposition 25.24. (i) $\Rightarrow$ (ii) is clear. (ii) $\Rightarrow$ (iv) and the other implications follow from localizing and applying Exercise 8.29 of Volume 1.)
18. Show that the Picard group of a Dedekind domain is the class group.

Projective dimension

19. We call two modules M, M' **projectively equivalent** if there are projective modules P, P' such that $M \oplus P \cong M' \oplus P'$. Show that any module projectively equivalent to a projective module is projective.
20. (Schanuel's Lemma.) If $0 \to K_i \to P_i \xrightarrow{g_i} M \to 0$ are exact with P_i projective for $i = 1, 2$, show that $P_1 \oplus K_2 \cong P_2 \oplus K_1$. (Hint: Lift g_1 to a map $h\colon P_1 \to P_2$ satisfying $g_1 = g_2 h$ and define $f\colon P_1 \oplus K_2 \to P_2$ by $f(x,y) = h(x) - y$. Then f is onto, and $\ker f \cong K_1$.)

21. (Generalized Schanuel's Lemma.) Suppose

$$0 \to K \to P_n \xrightarrow{f_n} P_{n-1} \xrightarrow{f_{n-1}} \dots \xrightarrow{f_1} P_0 \xrightarrow{\epsilon} M \to 0,$$
$$0 \to K' \to P'_n \xrightarrow{f'_n} P'_{n-1} \xrightarrow{f'_{n-1}} \dots \xrightarrow{f'_1} P'_0 \xrightarrow{\epsilon'} M \to 0,$$

with each P_i, P'_i projective. Prove that

$$K \oplus P'_n \oplus P_{n-1} \oplus P'_{n-2} \oplus \cdots \cong K' \oplus P_n \oplus P'_{n-1} \oplus P_{n-2} \oplus \dots.$$

Thus, all n-syzygies of M are projectively equivalent. (Hint: Define M' to be $P'_0 \oplus \ker\epsilon \cong P_0 \oplus \ker\epsilon'$. Then

$$0 \to K \to P_n \xrightarrow{f_n} P_{n-1} \xrightarrow{f_{n-1}} \cdots \to P_1 \oplus P'_0 \xrightarrow{f_1 \oplus 1} M' \to 0,$$
$$0 \to K' \to P'_n \xrightarrow{f'_n} P'_{n-1} \xrightarrow{f'_{n-1}} \cdots \to P'_1 \oplus P_0 \xrightarrow{f'_1 \oplus 1} M' \to 0,$$

so conclude by induction on n.)

22. With notation as in Theorem 25.44, verify the inequality $\operatorname{pd} M \leq \max\{\operatorname{pd} M', \operatorname{pd} M''\}$, and show that $\operatorname{pd} M = \operatorname{pd} M'$ if $\operatorname{pd} M' > \operatorname{pd} M''$.
23. With notation as in Theorem 25.44, when K'_n, K_n are the n-th syzygies of projective resolutions $\mathbb{P}'$, $\mathbb{P}$ of M', M respectively, show that there is an exact sequence

$$0 \to K'_n \to K_n \to P''_n \to P''_{n-1} \to \cdots \to P''_0 \to M'' \to 0.$$

Conclude that $\operatorname{pd} M'' \leq \max\{\operatorname{pd} M, \operatorname{pd} M'\} + 1$. If $\operatorname{pd} M' < \operatorname{pd} M$, then $\operatorname{pd} M'' = \operatorname{pd} M$. If $\operatorname{pd} M' > \operatorname{pd} M$, then $\operatorname{pd} M'' = \operatorname{pd} M' + 1$.

24. Verify the inequality $\operatorname{pd}_{R[\lambda]} M \leq \operatorname{pd}_R M + 1$ for any $R[\lambda]$-module M. Conclude that

$$\operatorname{gl\,dim} R[\lambda] \leq \operatorname{gl\,dim} R + 1.$$

In fact, one can show equality when $\operatorname{gl\,dim} R$ is finite. (Hint for first part: Define $M[\lambda] = R[\lambda] \otimes_R M$. There is an exact sequence

$$0 \to M[\lambda] \xrightarrow{g} M[\lambda] \to M \to 0$$

of $R[\lambda]$-modules, where $g(\lambda^n \otimes a) = \sum \lambda^n \otimes \lambda a - \lambda^{n+1} \otimes a$. Apply Exercise 23.)

25. (Eilenberg's trick.) For any projective module P, show that the module $P \oplus F$ is free for some free module F. This explains why f.g. free modules are used in the definition of stably free. (Hint: Take P' such that $P \oplus P'$ is free, and consider $P \oplus P' \oplus P \oplus P' \cdots$.)

Injective modules

26. Prove that a module E is injective iff, for every monic $h: N \to M$, the natural map $h^{\#}: \operatorname{Hom}(M, E) \to \operatorname{Hom}(N, E)$ is onto.
27. Prove that the direct product of injective modules is injective.
28. (Baer's criterion.) To verify injectivity in Definition 25.28, it is enough to check Equation (25.5) for $M = R$. (Hint: Using Zorn's Lemma, extend $f: N \to E$ to $\hat{f}: \hat{N} \to E$, with $\hat{N} \subseteq M$ maximal possible. Taking $a \in M \setminus \hat{N}$, put $L = \{r \in R : ra \in \hat{N}\}$, and define $f_1: L \to E$ by $f_1(r) = f(ra)$. By hypothesis, f_1 extends to a map $R \to E$, so $\hat{f}$ extends in turn to a map $\hat{N} + Ra \to E$.)
29. Show that every direct summand of an injective module is injective. (Hint: Use Baer's criterion.)
30. Show that E is injective iff every monic $E \to M$ splits.

Divisible modules

A module M is called **divisible** if $M = sM$, for every regular element $s \in R$.

31. Verify that each homomorphic image of a divisible module is divisible. A direct sum or product of divisible modules is divisible.
32. Show that every injective module E is divisible. (Hint: Given $a \in E$, define $f: Rs \to E$ by $s \mapsto a$.)
33. Verify that every divisible module M over a PLID is injective. (Hint: Given $f: Rs \to M$, $f(s) = sa$ for some $a \in M$; use Baer's criterion with $1 \mapsto a$.)
34. Show that $\mathbb{Q}/\mathbb{Z}$ is an injective $\mathbb{Z}$-module. (Hint: $\mathbb{Q}$ is divisible.)

The Pontrjagin dual

Given an R-module M, define the **Pontrjagin dual** $M^{\#} = \operatorname{Hom}_{\mathbb{Z}}(M, \mathbb{Q}/\mathbb{Z})$, viewed naturally as a right R-module.

35. Prove that a map $f: A \to B$ is monic iff the dual map $f^{\#}: B^{\#} \to A^{\#}$ is epic. The sequence $A_1 \to A_2 \to A_3$ is exact iff the sequence $A_3^{\#} \to A_2^{\#} \to A_1^{\#}$ is exact.
36. Show that the natural map $M^{\#} \otimes_R N \to \operatorname{Hom}(N, M)^{\#}$ is an isomorphism, for every finitely presented module N.

Injective hulls

37. Suppose R is an algebra over a field C. For any flat right R-module P and any injective C-module E, show that $P^* = \operatorname{Hom}_C(P, E)$ is injective (viewed naturally as an R-module). (Hint: Given $f: N \to M$ monic, one needs to show that $f^*: \operatorname{Hom}_R(M, P^*) \to \operatorname{Hom}_R(N, P^*)$ is onto. One has $\operatorname{Hom}_C(P \otimes_R N, E) \to \operatorname{Hom}_C(P \otimes_R M, E)$ monic. Conclude with the adjoint isomorphism.)

38. Show that every module M can be embedded in an injective module. (Hint: Write $M = F/K$, where F is a free $\mathbb{Z}$-module. Then $F \otimes_{\mathbb{Z}} \mathbb{Q}$ is divisible, so M is contained in the $\mathbb{Z}$-module $N = (F \otimes_{\mathbb{Z}} \mathbb{Q})/K$, which as a $\mathbb{Z}$-module is divisible and thus injective. Hence, $\mathrm{Hom}_{\mathbb{Z}}(R, N)$ is an injective R-module, by Exercise 37.)
39. (Injective hulls.) Say E is an **injective hull** of M if M is a large submodule of an injective module E. Prove that any module M has an injective hull. (Hint: Take E_1 injective containing M and, using Zorn's Lemma, take $E < E_1$ maximal such that M is a large submodule of E.)
40. Show that the injective hull is unique up to isomorphism.
41. (Projective covers.) A submodule K of M is called **small** if $K + N \neq M$ for every proper submodule N of M. Dually to injective hull, define a **projective cover** of a module M to be an epic $\pi\colon P \to M$ with P projective such that $\ker \pi$ is small. Show that if R is a ring that has no nonzero small ideals, e.g. $R = \mathbb{Z}$, then any cyclic R-module that is not projective fails to have a projective cover. Conclude that the category R-**Mod** is not self-dual.
42. Identify the maximal ring of quotients $Q(R)$ of a nonsingular ring R (Exercise 16.23) as $\mathrm{End}_R E$, where E is the injective hull of R.

Adjoint pairs

43. Conversely to Example 25.6, for any adjoint pair (F, G) with respect to η, show that $U_D = FD$ is a universal for D, where $\nu_D = \eta_{D,FD} 1_{FD}$.
44. Given any ring R, let E be the injective hull of $\bigoplus_{L<R} R/L$. Show that E also satisfies the property that for any R-module N and any $a \in N$ there is $f\colon N \to E$ with $f(a) \neq 0$. (This is the dual property to "generator"; cf. Definition 25A.2.)
45. For any adjoint pair (F, G) of functors $F\colon W$-**Mod** $\to R$-**Mod** and $G\colon R$-**Mod** $\to W$-**Mod**, show that F is right exact and G is left exact. (Hint: To check that F is right exact, suppose $0 \to K \to M \to N \to 0$ is an exact sequence, and consider the commutative diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & \mathrm{Hom}_R(FN, E) & \longrightarrow & \mathrm{Hom}_R(FM, E) & \longrightarrow & \mathrm{Hom}_R(FK, E) \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & \mathrm{Hom}_W(N, GE) & \longrightarrow & \mathrm{Hom}_W(M, GE) & \longrightarrow & \mathrm{Hom}_W(K, GE) \end{array}$$

    where E is as in Exercise 44. The bottom row is exact, so the top row is exact, implying that $FN \to FM \to FK \to 0$ is exact.)
46. Reprove from Exercise 45 that the functor $M\otimes$ is right exact, for any R, W-bimodule M.

Chain complexes

47. Define a **bifunctor** to be a covariant functor $F \to \mathcal{C}_1 \times \mathcal{C}_2 \to \mathcal{D}$, where $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{D}$ are categories (of modules). Fixing modules M and N, define $\hat{F} = F(M, \_)$, and $\bar{F} = F(\_, N)$. For homological δ-functors $\hat{T}$ and $\bar{T}$ defined by means of $\hat{F}$ and $\bar{F}$ respectively. Show that $\hat{T}_n(N) = \bar{T}_n(M)$ for all n. (Hint: Induction on n. For $n = 1$, take P projective with an exact sequence $0 \to K \to P \to M \to 0$, and P' projective with an exact sequence $0 \to K' \to P' \to N \to 0$. Piece together various exact sequences defining $\hat{T}_1(N)$ and $\bar{T}_1(M)$, and apply a dimension shifting argument analogous to Remark 25.60.)
48. Find the analogs of Exercise 47 when F is contravariant in one position.
49. A **bicomplex**, or **double complex**, is a family $\mathbf{A} = \bigoplus_{m,n\in\mathbb{Z}} A_{m,n}$, for $A_{m,n}$ in $\mathcal{C}$, with horizontal boundary maps $d^h_{m,n}\colon A_{m,n} \to A_{m-1,n}$ and vertical boundary maps $d^v_{m,n}\colon A_{m,n} \to A_{m,n-1}$ such that

    $$d^h_{m-1,n} d^h_{m,n} = d^v_{m,n-1} d^v_{m,n} = 0 = d^h_{m,n-1} d^v_{m,n} + d^v_{m-1,n} d^h_{m,n}, \ \forall m, n.$$

    For any n, the collection $\{(-1)^m d^v_{m,n} : m \in \mathbb{Z}\}$ comprises a chain map, i.e., a morphism in $\mathbf{Ch}(\mathcal{C})$. Conclude that the category of bicomplexes can be identified with $\mathbf{Ch}(\mathbf{Ch}(\mathcal{C}))$, and resolutions of modules M and N can be combined by means of a bifunctor F to produce a double resolution of $F(M, N)$.
50. Construct the **total (chain) complex** $\mathbf{Tot}(\mathbf{A}) = \bigoplus_u \mathrm{Tot}(\mathbf{A})_u$, where $\mathrm{Tot}(\mathbf{A})_u = \bigoplus_{m+n=u} A_{m,n}$ is a complex with boundary map $d = d^h + d^v$. Use this to define homology and cohomology for bicomplexes, and derived bifunctors.
51. Reprove Exercise 47 for Tor by means of bicomplexes. (Hint: First, prove that the tensor product of chain complexes $\mathbf{M}$ and $\mathbf{N}$ yields a bicomplex $\bigoplus_u \left(\bigoplus_m M_m \otimes N_{u-m}\right)$. The tensor product of resolutions thus defines homology and cohomology, as in Exercise 49. Next, show that dimension shifting of Remark 25.60 holds for either component of the bicomplex, so the derived bifunctor of $M \otimes N$ via the bicomplex matches the functor obtained by fixing one component or the other. One can replace dimension shifting by an easy argument to pass directly to the complexes for the first or second component, by modding out sub-bicomplexes with trivial homology.)

δ-functors

52. Given a chain map $f\colon (\mathbf{A}, d) \to (\mathbf{A}', d')$, define the **mapping cone** of f to be the chain complex whose n-th component is $A_{n-1} \oplus A'_n$ with boundary map

    $$d(a_{n-1}, a'_n) = (-d(a_{n-1}), d'(a'_n) - f(a_{n-1})).$$

Show that when f is the identity map, this boundary map splits the mapping $(a_{n-1}, a_n) \mapsto (-a_n, 0)$.

53. Using Exercise 52, show that the homology functor $\mathbf{H}$ is a universal δ-functor.

Finitely presented modules

54. Show that the module K is f.g. whenever $N = M/K$ is finitely presented and M is f.g. (Hint: First assume M is projective. Write $N \cong F/K_1$ with F free and K_1 f.g. By Exercise 20, K is a summand of the f.g. module $M \oplus K_1$. In general, take the epic $f\colon M \to N$ having kernel K and an epic $\pi\colon F \to M$ with F f.g. free. Then $\ker(f\pi)$ is f.g., and $K = \pi(\ker f\pi)$.)
55. (Generic flatness.) Show that $S^{-1}M$ is a finitely presented $S^{-1}R$-module for any M finitely presented C-module and any submonoid S of $\mathrm{Cent}(R)$. If $S^{-1}M$ is free as an $S^{-1}C$-module, then there is $s \in S$ such that $M[s^{-1}]$ is free as a $C[s^{-1}]$-module. (Hint: If $M \cong F/K$, then $S^{-1}M \cong S^{-1}F/S^{-1}K$. For the second assertion, take a base $\{\frac{a_1}{z}, \dots, \frac{a_t}{z}\}$ of $S^{-1}M$; then $a_1, \dots, a_t$ span $M[z^{-1}]$ over $R[z^{-1}]$. Let $\pi\colon R[z^{-1}]^{(t)} \to M[z^{-1}]$ be onto. Then $\ker \pi$ is f.g., implying that there is $z' \in S$ with $z' \ker \pi = 0$. Take $s = zz'$.)

Flat modules

56. Show that every f.g. projective module is finitely presented, and conversely, every finitely presented flat module is projective. (Hint: Use Exercise 36.)
57. Show that $S^{-1}C$ is a flat C-module for any submonoid S of a commutative ring C.
58. Suppose E is an injective R-module and M is an R, S-bimodule that is flat as an R-module. Show that $\mathrm{Hom}_R(M, E)$ is injective as a right S-module. Use this in conjunction with Exercise 19.24 to prove that group algebras over a field are quasi-Frobenius.

The Tor functor

59. For any Abelian group A, show that $\mathrm{Tor}_0(\mathbb{Z}/m, A) = A/mA$, $\mathrm{Tor}_1(\mathbb{Z}/m, A) = \{a \in A : ma = 0\}$, and $\mathrm{Tor}_n(\mathbb{Z}/m, A) = 0$ for all $n \geq 2$.
60. For any exact sequence $0 \to K \to F \to N \to 0$ with F free, and for $M = R/A$ where $A \lhd R$, show that $\mathrm{Tor}_1(M, N) = (K \cap AF)/AK$. In particular, for $B \lhd R$, $\mathrm{Tor}_1(R/A, R/B) = (A \cap B)/AB$.

Ext and derived functors

61. Prove that $\mathrm{gl\,dim}\, R = \sup\{n : \mathrm{Ext}^n(M, N) \neq 0$ for all R-modules $M, N\}$, and also equals sup{injective dimensions of all R-modules}.

62. Prove that $\mathrm{gl\,dim}\, R = \sup\{\mathrm{pd}\, R/L : L < R\}$. (Hint: Show for any injective resolution that the n cosyzygy is injective, by means of Baer's criterion.)
63. Show that a ring R is left hereditary iff $\mathrm{gl\,dim}\, R \leq 1$.
64. Show that $\mathrm{Ext}^1(P, N) = 0$, for every exact sequence $0 \to K \xrightarrow{f} P \to M \to 0$, with P projective. Conclude that an exact sequence

$$0 \to \mathrm{Hom}(M, N) \to \mathrm{Hom}(P, N) \xrightarrow{f^\#} \mathrm{Hom}(K, N) \to \mathrm{Ext}^1(M, N) \to 0.$$

Thus, $\mathrm{Ext}^1(M, N) \cong \mathrm{Hom}(K, N)/f^\#(\mathrm{Hom}(P, N))$.

65. Define equivalence of module extensions analogously to (25.17). Show that $\mathrm{Ext}^1(M, N)$ can be identified with the equivalence classes of module extensions $0 \to N \to E \to M \to 0$. Hint: Take a projective resolution $\mathbb{P}$ of M and fill in the vertical arrows step by step, starting with 1_N, as follows:

$$\begin{array}{ccccccccccc}
\dots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{d_0} & M & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle 0} & & \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_0} & & \downarrow{\scriptstyle 1_M} & & \\
\dots & \longrightarrow & 0 & \longrightarrow & N & \longrightarrow & E & \longrightarrow & M & \longrightarrow & 0
\end{array}$$

Then f_1 yields an element in $\ker d_2^*/\operatorname{im} d_1^* = \mathrm{Ext}^1(M, N)$. It remains to find the reverse correspondence, which is done by taking the diagram:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & P_1/d_2(P_2) & \xrightarrow{\bar{d}_1} & P_0 & \xrightarrow{d_0} & M & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \tilde{f}_1} & & & & & & \\
 & & N & & & & & &
\end{array}$$

and applying the pushout (Exercise 2.19 of Volume 1).

Group cohomology

66. Show that the tensor product of two projective resolutions of $\mathbb{Z}$ viewed respectively as a G_1-module and a G_2-module is a projective resolution of $\mathbb{Z}$ as a $G_1 \times G_2$-module, by Proposition 20.25. This leads via Exercise 51 to the "cup product," which makes $H^*(G, \mathbb{Z})$ a graded-commutative ring; jumping ahead to Chapter 26, for a field F and group G, $H^*(G, F)$ is a Hopf algebra. See [**Wei, Chapter 6**] for details.
67. Write out the details of the assertions in Examples 25.63 and 25.65.
68. For the cyclic group $C_m = \langle \sigma \rangle$ of order m, define $N = \sum_{i=0}^{m-1} \sigma^i$. Show that $\mathbb{Z}$, as a trivial $\mathbb{Z}[C_m]$-module, has the free resolution

$$\dots \xrightarrow{N} \mathbb{Z}[C_m] \xrightarrow{(1-\sigma)} \mathbb{Z}[C_m] \xrightarrow{N} \mathbb{Z}[C_m] \xrightarrow{(1-\sigma)} \mathbb{Z}[C_m].$$

For any C_m-module M, let $\mathcal{I}_M = \{a \in \mathbb{Z}[C_m] : Na = 0\}$. Conclude that $H_0(C_m, M) = M/(\sigma - 1)M$, $H_{2n-1}(C_m, M) = M^{C_m}/NM$, and $H_{2n}(C_m, M) = \mathcal{I}_M/(\sigma - 1)M$ for $n > 0$. Likewise, $H^0(C_m, M) = M/(\sigma - 1)M$, $H^{2n-1}(C_m, M) = \mathcal{I}_M/(\sigma - 1)M$, and $H_{2n}(C_m, M) = M^{C_m}/NM$ for $n > 0$. (Hint: $(\sigma - 1)N = 0$ in $\mathbb{Z}[C_m]$. $\mathcal{I}_{C_m}$ is the augmentation ideal of $\mathbb{Z}[C_m]$.)

Shapiro's Lemma

69. Compute that the corestriction map is compatible with the transfer in the cohomology of $H^2(G, K^\times)$.

Lie homology and cohomology

Assume that L is a semisimple Lie algebra over a field F of characteristic 0.

70. Show that $H^n(L, M) = H_n(L, M) = 0$ for every simple L-module M and all n (Hint: Let R be the universal enveloping algebra $U(L)$, and view F as the trivial module over R via the augmentation map. The Casimir element of M annihilates $\mathrm{Tor}_n(F, M)$ and $\mathrm{Ext}^n(F, M)$ since it annihilates F, by Exercise 21C.15, but acts as a nonzero scalar on M.)
71. Prove that $H^1(L, M) = \mathrm{Deriv}(L)/\mathrm{InnDeriv}(L)$. (Hint: Letting $\mathcal{I}$ be the augmentation ideal of the enveloping algebra $U(L)$, apply the long exact sequence of cohomology to the exact sequence

$$0 \to \mathcal{I} \to U(L) \to F \to 0$$

to get the exact sequence

$$0 \to M^L \to M \to \mathrm{Hom}_L(\mathcal{I}, M) \to H^1(L, M) \to 0.$$

Apply Exercise 21C.16.) Compare with Exercise 25B.11 below.
72. Reprove Weyl's Theorem by showing that $H^1(L, M) = 0$ for any finite-dimensional Lie module M.

Appendix 25A

Morita equivalence

1. For any ring R and any idempotent e of R, verify the functor $F\colon R\text{-}\mathbf{Mod} \to eRe\text{-}\mathbf{Mod}$, where $FM = eM$ and Ff is the restriction of $f\colon M \to N$ to $eM \to eN$.
2. Verify the functor $F\colon R\text{-}\mathbf{Mod} \to M_n(R)\text{-}\mathbf{Mod}$, defined as follows: $FM = M^{(n)}$, where $M^{(n)}$ is viewed as an $M_n(R)$-module via matrix multiplication, and Ff acts componentwise for $f\colon M \to N$.

3. Prove directly that R-**Mod** and $M_n(R)$-**Mod** are equivalent categories. (Hint: The functors from Exercises A2 and A1 yield the category equivalence.)
4. Under the category equivalence of Exercise A3, show that the R-module $R^{(n)}$ corresponds to the module $M_n(R)$ over itself. Conclude that cyclicity is not a categorical module-theoretic property. On the other hand, show that this correspondence often reduces theorems about f.g. modules to cyclic modules, as seen in the proof of the Jacobson density theorem (Theorem 15A.1).
5. Show that a module is f.g. iff it is not a direct limit of proper submodules. State this in the language of Abelian categories. Thus, "f.g." can be defined categorically, as can "Noetherian."

Morita equivalence

6. Prove that any two Morita equivalent commutative rings are isomorphic. (Hint: Exercise 25.17.)
7. Suppose $R_0 = \mathcal{A}_1(\mathbb{R})$, the Weyl algebra of Example 13A.1, and let σ be the automorphism sending $\lambda \mapsto -\lambda$ and $\mu \mapsto -\mu$. Prove that the skew group ring $R_0\#\{1, \sigma\}$ of Exercise 19.35 is simple but not a matrix ring over a domain. See also Exercise 26.38.
8. Construct the **Morita ring** $\left(\begin{smallmatrix} R & M \\ M' & R' \end{smallmatrix}\right)$, where multiplication in the matrix components is taken according to how it makes sense, i.e.,

$$\begin{pmatrix} r_1 & a_1 \\ f_1 & r_1' \end{pmatrix}\begin{pmatrix} r_2 & a_2 \\ f_2 & r_2' \end{pmatrix} = \begin{pmatrix} r_1r_2 + \tau(a_1 \otimes f_2) & r_1a_2 + r_2'(a_1) \\ f_1r_2 + r_1'f_2 & \tau'(f_1 \otimes a_2) + r_1'r_2' \end{pmatrix}.$$

$(R, R', M, M', \tau, \tau')$ is a Morita context iff associativity holds in the Morita ring.
9. Verify the following assertions in a Morita context $(R, R', M, M', \tau, \tau')$, with τ, τ' onto:

M is a progenerator in R-**Mod** and **Mod**-R'.

M' is a progenerator in R'-**Mod** and **Mod**-R.

$M' \cong M^*$ as R, R'-bimodules.

$M \cong \mathrm{Hom}_{R'}(M', R)$ as R', R-bimodules.

$R' \cong (\mathrm{End}_R M)^{\mathrm{op}}$.

$R \cong \mathrm{End}_{R'} M'$.

R and R' are Morita equivalent.

The functors $\mathrm{Hom}_R(M, \_)$ and $M' \otimes_R \_$ are naturally isomorphic.

(Hint: Most parts are variations of the proof of Morita's Theorem, and some were done in Lemma 25A.17. The last part follows from

the chain of isomorphisms $\mathrm{Hom}_R(M,N) \cong \mathrm{Hom}_{R'}(M'\otimes M,\ M'\otimes N) \cong \mathrm{Hom}_{R'}(R', M'\otimes N) \cong M'\otimes N$.)

10. Hypotheses as in Exercise A9, there is a lattice isomorphism from the ideals of R to the R', R-sub-bimodules of M', given by $A \mapsto M'A$.
11. Applying Exercise A10 twice, show that there is a lattice isomorphism from the ideals of R to the ideals of R'. In particular, any ring that is Morita equivalent to a simple ring is simple.

Appendix 25B

Separable algebras

1. Suppose $C \subseteq \tilde{C} \subseteq \mathrm{Cent}(R)$. If R is separable over $\tilde{C}$ and $\tilde{C}$ is separable over C, then R is separable over C. Conversely, if R is separable over C and projective over $\tilde{C}$, then R is separable over $\tilde{C}$ and $\tilde{C}$ is separable over C. (Hint: For R separable over C, the same separability idempotent shows that R is separable over $\tilde{C}$; to see that $\tilde{C}$ is separable over C, note that $\tilde{C}$ is a direct summand of R.)
2. If K is commutative and faithfully projective over C, and if $R \otimes_C K$ is separable over K, then R is separable over C.
3. $\mathrm{Hom}_{R^e}(R,M) \cong M^R$ via $f \mapsto f(1)$. The inverse map is given by sending $a \in M^R$ to the right multiplication map $r \mapsto ra$.
4. $\mathrm{Hom}_{R^e}(R,R) \cong \mathrm{Cent}(R)$. (Hint: Take $M = R$ in Exercise B3.)

Hochschild cohomology

5. Suppose R is a separable C-algebra, and M is an R^e-module. Define $C^0(R,M) = M$ and $C^n(R,M) = \mathrm{Hom}_C(R^{(n)}, M)$ for $n > 0$, and $\delta_n\colon C^n(R,M) \to C^{n+1}(R,M)$, by taking $\delta_n f(r_1,\dots,r_n)$ to be

$$r_1 f(r_2,\dots,r_{n+1}) + \textstyle\sum_{i=1}^{n}(-1)^i f(r_1,\dots,r_ir_{i+1},\dots,r_{n+1}) + (-1)^{n+1} f(r_1,\dots,r_n)r_{n+1}.$$

Find a natural isomorphism of the n-th cohomology group with $\mathrm{Ext}^n_{R^e}(R,M)$. (Hint: Define $P_n = R^{\otimes(n+2)} \cong R^e \otimes_C R^{\otimes(n)}$, a projective R^e-module. Then $\mathrm{Hom}_{R^e}(P_n,M) \cong \mathrm{Hom}_C(P_{n-2},M) = C^n(R,M)$. If one defines $\partial_n\colon P_n \to P_{n-1}$ by

$$\partial_n(r_1\otimes\cdots\otimes r_{n+2}) = \sum(-1)^{i-1} r_1\otimes\cdots\otimes r_ir_{i+1}\otimes\cdots\otimes r_{n+2},$$

then $\partial_n^{\#}\colon \mathrm{Hom}_{R^e}(P_n,M) \to \mathrm{Hom}_{R^e}(P_{n+1},M)$ yields the desired identification.)

6. Suppose $H^2(R,\_) = 0$, and R has a nilpotent ideal N such that R/N is separable. Prove that R has a subalgebra $S \cong R/N$ that is a complement to N as a C-module. (Hint: Induction on t such that $N^t = 0$. First assume that $t = 2$. The projection $R \to R/N$ splits (over C); write $R = N' \oplus N$. Letting $\pi : R \to N'$ be the projection, define $f(r_1,r_2) = \pi(r_1r_2) - \pi(r_1)\pi(r_2)$. Then $\delta_2 f = 0$, so $f = \delta_1 g$; hence $R = S \oplus N$, where $S = \{\pi(r) + g(r) : r \in R\}$.

In general, apply the previous case to $\bar{R} = R/N^2$ and $\bar{N} = N/N^2$.)

7. (Wedderburn's Principal Theorem.) Suppose R is a finite-dimensional algebra over a perfect field F, and let $J = \mathrm{Jac}(R)$. Prove that R has a semisimple subalgebra $S \cong R/J$ and a Wedderburn decomposition $R = J \oplus S$. (Hint: R is separable over F, so $H^2(R,\_) = 0$ and Exercise B6 is applicable.)

The role of derivations for separable algebras

Assume throughout that R is a C-algebra and M is an R^e-module. Write $\mathrm{Deriv}_C(R,M)$ for the set of C-derivations from R to M, and $\mathrm{InnDer}(R,M)$ for the set of inner derivations on M, a C-submodule of $\mathrm{Deriv}_C(R,M)$. Recall $J = \ker p$; cf. Proposition 25B.3.

8. Verify the derivation $d\colon R \to J$ given by $d(r) = r\otimes 1 - 1\otimes r$.
9. Verify the isomorphism $\Phi_M\colon \mathrm{Hom}_{R^e}(J,M) \to \mathrm{Deriv}_C(R,M)$ of C-modules, given by $f \mapsto f\circ d$; f corresponds to an inner derivation iff f extends to a map in $\mathrm{Hom}_{R^e}(R^e,M)$. (Hint: Φ_M is a C-module map. If $f\circ\delta = 0$, then $f(r\otimes 1 - 1\otimes r) = 0$ for each $r \in R$, implying that $f = 0$ on J. Hence Φ_M is 1:1.

Next, prove that Φ_M is onto. Given a derivation $\delta\colon R \to M$, define a balanced map $g\colon R \times R^{\mathrm{op}} \to M$ by $g(a,b) = -a\delta(b)$. This yields a C-module map $R^e \to M$ whose restriction $f\colon J \to M$ is an R^e-module map.

For the last assertion, if $\delta = \mathrm{ad}_a$, then f extends to the map given by $1\otimes 1 \mapsto a$; conversely, f corresponds to $\delta = \mathrm{ad}_{\hat{f}(1\otimes 1)}$.)

10. Prove that every derivation $\delta\colon R \to M$ is inner iff the derivation $d\colon R \to J$ is inner. (Hint: If $d = \mathrm{ad}_a$, then $f\circ d = \mathrm{ad}_{f(a)}$ is inner.)
11. Show that the short exact sequence $0 \to J \to R^e \to R \to 0$ gives rise to an exact sequence

$$0 \to \mathrm{Hom}_{R^e}(R,M) \to \mathrm{Hom}_{R^e}(R^e,M) \to \mathrm{Hom}_{R^e}(J,M) \to \mathrm{Ext}^1_{R^e}(R,M) \to \mathrm{Ext}^1_{R^e}(R^e,M) = 0,$$

which can be reinterpreted as the exact sequence

$$0 \to M^R \to M \to \mathrm{Deriv}_C(R,M) \to \mathrm{Ext}^1_{R^e}(R,M) \to 0.$$

(Hint: $B^1(R,M)$ is the set of inner derivations $R \to M$, whereas $Z^1(R,M)$ is the set of derivations $R \to M$.)

12. Show that the following conditions are equivalent for R, notation as above:

    (i) R is separable over C.

    (ii) $\mathrm{Ext}^1_{R^e}(R, M) = 0$ for every R^e-module M.

    (iii) Every C-derivation of R into an R^e-module is inner.

    (iv) The derivation $d\colon R \to J$ of Exercise B8 is inner. Conclude that if R is separable and commutative over C, then R has no C-derivations.

Azumaya algebras

13. Show that any Azumaya algebra R over C can be defined as a tensor extension of an Azumaya algebra over a finitely generated (Noetherian) subring of C. (Hint: Use the structure constants affiliated with the separability idempotent.)
14. (Braun's criterion.) Show that a C-algebra R is Azumaya, iff there are $a_i, b_i \in R$ such that $\sum a_i b_i = 1$ and $\sum a_i R b_i \subseteq C$. (Hint: Remark 25B.7.)
15. By Braun's criterion, find an onto map $R \to C$ sending $1 \mapsto 1$. (This is not quite the trace map.)
16. Show that R is separable over C if $R \otimes_C H$ is separable over H and H is faithfully projective over C.
17. Show that any Azumaya algebra is a finite direct product of algebras of constant rank when the base ring C has no nontrivial idempotents. (Hint: Exercise 25.16.)
18. Show that a commutative separable subalgebra K of an Azumaya algebra of rank n^2 is maximal, iff $\operatorname{rank} K = n$. (Hint: Localize, pass to a simple homomorphic image, and apply Nakayama's Lemma.)

Appendix 25C

The category of representation theory

1. A subcategory $\mathcal{C}_0$ of $\mathcal{C}$ is called **dense** if any object of $\mathcal{C}$ is isomorphic to a suitable object in $\mathcal{C}_0$. A category is called **skeletally small** if it has a small dense subcategory. Show that Definition 1A.17 of Volume 1 makes sense if the category $\mathcal{C}$ is skeletally small.
2. Define R-**Fimod** to be the full subcategory of R-**Mod**whose objects are the f.g. modules. Show that R-**Fimod** is skeletally small. (Hint: Use Propositions 2.23 and 2.24 of Volume 1.)
3. Using Exercises C1 and C2, given rings R and T, build a category **Fun**(R-**Fimod**, T-**Mod**). This category is fundamental in the abstract representation theory of rings and modules.

4. What is the quiver of the algebra of upper triangular $n \times n$ matrices?
5. What is the quiver of the group algebra $F[C_2 \times C_2]$, where F is the algebraic closure of $\mathbb{Z}/2$? (See Exercise 19.8.)

Chapter 26

Hopf algebras

1. Define a **coideal** of a coalgebra C to be a submodule I such that $\Delta I \subseteq I \otimes C + C \otimes I$ and $\epsilon(I) = 0$. Show then that C/I is a coalgebra with comultiplication and counit inherited from Δ and ϵ. Conversely, the kernel of a coalgebra morphism is a coideal.
2. Define the **coopposite coalgebra** C^{cop} to be the coalgebra of the opposite algebra, and show that the new comultiplication is $\tau \circ \Delta$.
3. For any algebra A and coalgebra C, verify that $\mathrm{Hom}(C, A)$ becomes an algebra under the convolution product ($*$) of Definition 26.3. If H is a Hopf algebra, then its antipode S is the inverse to 1_H in $\mathrm{Hom}(H, H)$ under the convolution product (taking $C = A = H$). In particular, the antipode is uniquely defined.
4. Suppose A is both an algebra (A, μ, ι) and a coalgebra (A, Δ, ϵ). Prove that Δ and ϵ are algebra homomorphisms, iff μ and ι are coalgebra morphisms.
5. Verify the following equations in a Hopf algebra H: $S(ab) = S(b)S(a)$, $\Delta \circ S = \tau \circ (S \otimes S) \circ \Delta$, and $\epsilon \circ S = \epsilon$. (Hint for the first assertion: $\mathrm{Hom}(H \otimes_F H, H)$ is an algebra, given the convolution product of Exercise 3. The multiplication map $\mu\colon H \otimes_F H \to H$ has two inverses under convolution: the right inverse $a \otimes b \mapsto S(b)S(a)$ and the left inverse $a \otimes b \mapsto S(ab)$. Hence these two maps are equal.)
6. Verify that the antipode S of a cocommutative Hopf algebra is an algebra involution; i.e., $S^2 = 1$.
7. Show that the grouplike elements of a Hopf algebra are all linearly independent over F.
8. Construct a three-dimensional non-cocommutative coalgebra, by taking one grouplike element and one primitive element.
9. Given a monoid M and commutative ring C, let H denote the set of C-valued functions from M. Show that H is a bialgebra, under pointwise multiplication, with respect to $\Delta(f)(a, b) = f(ab)$, and $\epsilon(f) = f(1)$; when M is a group, H becomes a Hopf algebra $S(f)(a) = f(a^{-1})$. If $F = \mathbb{R}$ and M is a compact topological group with Haar measure η, define T by $T(f) = \int_M f\, d\eta$ for each $f \in H$. How does T resemble an integral as defined in the text? How is T different?

10. Given an algebra A over a field F, show that the **finite dual** $A^o = \{f \in \mathrm{Hom}_F(A,F) : f \text{ vanishes on an ideal of } A \text{ of finite codimension}\}$ is a coalgebra. If A is commutative, then A^o is cocommutative.
11. Show that the dual H^* of a f.d. Hopf algebra is also a Hopf algebra, obviously of the same dimension. When H is infinite-dimensional, H^* is no longer a coalgebra, but the finite dual H^0 is a Hopf algebra.
12. Taking A as in Example 26.9(ii), let Λ be the matrix $(\lambda_{ij}) \in A$. A is not a Hopf algebra, since $\det(\Lambda)$ is a group-like element that is not invertible. Nevertheless, the coordinate algebra of $\mathrm{SL}(n,F)$ is $A/\langle \det(\Lambda)-1\rangle$, and the coordinate algebra of $\mathrm{GL}(n,F)$ is $A[\det(\Lambda)^{-1}]$. What are the respective antipodes?
13. Write the enveloping algebra $U(\mathrm{sl}(2,F))$ explicitly as a Hopf algebra.

Comodules and Hopf modules

14. Show that any Hopf algebra H is also an H-module via the **left adjoint action** $\mathrm{ad}_h(a) = \sum h_1 a S(h_2)$. Note for h grouplike that $\mathrm{ad}_h(a) = hah^{-1}$; for h primitive, note that $\mathrm{ad}_h(a) = ha - ah$.

    Likewise, H is also a right H-module under the **right adjoint action**, defined analogously.

Sub-comodules and coalgebras

15. Prove that the sum of sub-comodules also is a sub-comodule.
16. (Fundamental Theorem of Comodules.) For any comodule M over a coalgebra C, show that any finite subset of M is contained in a finite-dimensional sub-comodule of M. Conclude that every simple A-comodule is f.d. (Hint, according to [**DasNR**]: Take a base $\{a_i : i \in I\}$ of C. It suffices to show that every $w \in M$ lies in a f.d. sub-comodule. Write $\rho(w) = \sum_i a_i \otimes w_i$ and $\Delta(a_i) = \sum_{j,k} \alpha_{ijk} a_j \otimes a_k$. Computing $(1 \otimes \rho)\rho(w) = \sum a_i \otimes \rho(w_i)$ and matching coefficients of the a_i, show that the subspace spanned by the w_i is a sub-comodule.)
17. (Fundamental Theorem of Coalgebras.) Prove that any finite subset of a coalgebra C is contained in a f.d. subcoalgebra of C. (Hint: The same sort of argument as in Exercise 16; write $r = \Delta^2(a) = \sum a_i \otimes b_{ij} \otimes c_j$ and let B be the space generated by the b_{ij}. Comparing $\Delta \otimes 1_C \otimes 1_C(r)$ and $1_C \otimes \Delta \otimes 1_C(r)$, show that $\Delta(B) \subset B \otimes C \cap C \otimes B = B \otimes B$.) Conclude that every simple subcoalgebra of C is f.d.
18. Show that any 1-dimensional subcoalgebra is spanned by a grouplike element.
19. A Hopf subalgebra N of H is called **normal** if N is invariant under both the left and right adjoint actions with respect to each element of H. Show in this case that $N^+H = H^+N$ is a Hopf ideal of H, where N^+ is defined as $N \cap \ker \epsilon$. H is called an **Abelian extension** of N if H/N^+H is commutative.

20. For $H = F$, show that the comodule algebras are just the F-algebras.
21. Show that any H^*-comodule algebra can be viewed as an H-module algebra. Interpret this for Example 26.10.

Tensor products of Hopf modules

22. Show that any graded vector space with respect to a group G becomes an $F[G]$-comodule, where $\rho(v) = g \otimes v$ for $v \in V_g$.
23. For any Hopf algebra H over a field F, show that any F-vector space V becomes a module over H under the **trivial action** $hv = \epsilon(h)v$, and then $H \otimes V$ becomes a Hopf module as in Remark 26.15, the action given by $h'(h \otimes v) = h'h \otimes v$.
24. Fill in the details of the proof of Theorem 26.21. (Hint: φ is a morphism of H-modules, so it remains to show that φ and ψ are inverses.

    The tricky part is to show that $\psi(M) \in H \otimes M^{\mathrm{co}\,H}$; i.e., that $\sum S(a_{-1,2})a_0 \in M^{\mathrm{co}\,H}$, by verifying that

    $$\rho\big(\sum S(a_{-1,2})a_0\big) = \sum S(a_{-1,2})a_0 \otimes 1.$$

    This relies heavily on Equation (26.8) in conjunction with Exercise 5. The proof that the φ and ψ are inverses is then straightforward.)
25. For $H = F$, show that the comodule algebras are just the F-algebras.
26. Write down the variants of the condition (26.9) for the other two possible placements of 1_C and verify them. Shnider-Sternberg [**ShnS, Chapter 6**] call these three conditions the **coherence conditions**, and discuss the categorical setting.
27. Show that the category H-**Mod** is monoidal for any Hopf algebra H; likewise, the category of Hopf modules f.d. over a given base field F is monoidal.

Quasitriangular Hopf algebras and the QYBE

28. Verify the following equations whenever (H,R) is quasitriangular with $R = \sum a_i \otimes b_i$:
    (i) $R^{-1} = \sum S(a_i) \otimes b_i$.
    (ii) $\sum \epsilon(a_i)b_i = \sum a_i \epsilon(b_i) = 1$.
    (iii) $(S \otimes S)(R) = R$.
29. Verify that Example 26.11 is quasitriangular for $n = 4$. (Hint: Take

    $$R = \frac{1}{2}\,(1 \otimes 1 + 1 \otimes g + g \otimes 1 - g \otimes g.)$$

    Also, find other possibilities for R.)
30. For any quasitriangular Hopf algebra (H,R) and H-modules U,V,W, show that $V \otimes_F W \cong W \otimes_F V$ under the map

    $$v \otimes w \mapsto \tau\,(R(v \otimes w)) = \sum b_i w \otimes a_i v,$$

where $R = \sum a_i \otimes b_i$. Show that the two natural chains of isomorphisms from $(U \otimes V) \otimes W$ to $V \otimes (W \otimes U)$ give the same result. (Such a monoidal category is called **braided**.)

31. Let $V = F^{(2)}$. Show that the transformation on $V \otimes V$ given by

$$R(e_1 \otimes e_1) = qe_1 \otimes e_1; \quad R(e_1 \otimes e_2) = e_1 \otimes e_2 + (q - q^{-1})e_2 \otimes e_1;$$
$$R(e_2 \otimes e_1) = e_2 \otimes e_1; \qquad R(e_2 \otimes e_2) = qe_2 \otimes e_2$$

is a solution to the QYBE.

32. Suppose H is almost cocommutative with antipode S, and let $R = \sum a_i \otimes b_i$ be as in Definition 26.23. Let $u = \sum_i S(b_i)a_i$. Verify that u is invertible in H, $uS(u)$ is central, and S^2 is the inner automorphism given by conjugation with respect to u. (Hint: First show that $uh = S^2(h)u$ for all $h \in H$. Next, writing $R^{-1} = \sum c_j \otimes d_j$ and $v = \sum_j S^{-1}(d_j)c_j$, check that $uv = 1$.)

If H is quasitriangular, then one can show that S^4 is given by conjugation by the grouplike element $uS(u)^{-1}$ of H; cf. Montgomery [**Mo1, Theorem 10.1.13**].

Hopf algebras of low dimension

33. For $\mathrm{char}(F) \neq 2$, define H_8 to be the Hopf algebra generated by elements x, y, z with respect to the relations $x^2 = y^2 = 1$, $xy = yx$, $zx = yz$, $xz = zy$, and $z^2 = \frac{1}{2}(1 + x + y - xy)$, with x, y grouplike and

$$\Delta(z) = \frac{1}{2}((1+y) \otimes 1 + (1-y) \otimes x)(z \otimes z).$$

Let $N = F[x, y]$, a normal Hopf subalgebra of H. Show that $H/N^+H = F[\bar{z}]$, where $\bar{z}^2 = 1$, so H is an Abelian extension of $F[C_4]$ by $F[C_2]$.

Smash products

Suppose H is a Hopf algebra and A is an H-module algebra. The **smash product** $A\#H$ is the tensor product $A \otimes H$ as a vector space, with multiplication given by

$$(a\#h)(b\#k) = \sum a(h_1 b)\#h_2 k.$$

34. Show that the smash product $A\#H$ is an associative algebra, with algebra isomorphisms $A \cong A\#F$ and $H \cong F\#H$. Also show that A is an $A\#H$-module, under the action $(a\#h)b = a(hb)$ for $a, b \in A$ and $h \in H$.
35. When H is the group algebra $F[G]$ and the group G acts as automorphisms on A, show that $A\#H$ is the skew group algebra; cf. Exercise 19.35. Describe $A\#H$ when H is the universal enveloping algebra of a Lie algebra L, and L acts as derivations on A.

36. Prove that $(A\#H)\#H^* \cong M_n(A)$ when $\dim_F H = n$.
37. Describe the Weyl algebra $\mathcal{A}_1(F)$ of Example 13A.1 as $F[\lambda]\#H$, where H is the enveloping algebra of the 1-dimensional Lie algebra $F\delta$, and δ is the derivative on $F[\lambda]$. Generalize this to $\mathcal{A}_n(F)$ of Example 13A.2.
38. Show that the smash product naturally gives rise to a Morita context

$$(A\#H, A^H, A, A', \tau, \tau'),$$

where $A' = A$ but has a new right $A\#H$-module action (see Montgomery [**Mo2, Theorem 4.5.3**] for more details). Often one can conclude that $A\#H$ is Morita equivalent to A^H, thereby enabling one to transfer the structure between two important constructions.

Quantum groups

39. Show that the quantized enveloping algebra $U_q(\mathrm{sl}(2, F))$ of Definition 21C.7 is a Hopf algebra, where k is grouplike, $\epsilon(e) = \epsilon(f) = 0$,

$$\Delta(e) = e \otimes 1 + k \otimes e, \quad \Delta(f) = f \otimes k^{-1} + 1 \otimes f,$$

$S(e) = -q^{-2}e$, $S(f) = -q^2 f$, and $S(k) = k^{-1}$. The actual verifications are done using q-binomial coefficients; cf. Exercises 21C.18ff.

40. Verify that Examples 16A.3 are Hopf algebras. (Hint: $\mathcal{O}_q(M_2(F))$ is a bialgebra with respect to the comultiplication and counit of Example 26.9(ii). Now the antipode S for $\mathcal{O}_q(\mathrm{GL}(2, F))$ is defined by $S(x_{11}) = \delta_q^{-1}x_{22}$, $S(x_{12}) = -q^{-1}\delta_q^{-1}x_{12}$, and $S(x_{21}) = -q\delta_q^{-1}x_{21}$. The bialgebra operations for $\mathcal{O}_q(\mathrm{SL}(2, F))$ are analogous, and the antipode is easier since δ_q is 1.)

Hopf duality

41. Given H-Hopf modules M, N, define a module structure on $\mathrm{Hom}(M, N)$ via

$$hf(a) = \sum h_1 f(S(h_2))a.$$

Show that the correspondence of Proposition 18.44 sends Hopf module morphisms to Hopf module morphisms, while the tensor product is taken as Hopf modules.

List of Major Results

The prefix E denotes that the referred result is an exercise, such as E0.5. Since the exercises do not necessarily follow a chapter immediately, their page numbers may be out of sequence.

Prerequisites.

Chapter 13.

Chapter 14.

E14.21. If R is simple and finite-dimensional over an algebraically closed field F, and R has an involution $(*)$, then $(R,*) \cong (M_n(F), J)$, where J is either the transpose or the canonical symplectic involution. 167

Chapter 15.

15.7. If a prime ring R has a minimal nonzero left ideal L, then R is primitive and every faithful simple R-module is isomorphic to L. 47

15.9, 15.10. The Wedderburn-Artin decomposition $R = M_n(D)$ of a simple Artinian ring is unique. Every semisimple ring has finitely many simple nonisomorphic modules. 48

15.18.–15.20. Any left Artinian ring R has only finitely many primitive ideals, and each primitive ideal is maximal. Their intersection is the Jacobson radical J, which is nilpotent, and R/J is a semisimple ring. Consequently, any prime left Artinian ring is simple Artinian; any semiprime left Artinian ring is semisimple Artinian. 50, 51

15.21 (Hopkins-Levitzki). Any left Artinian ring is also left Noetherian. 52

15.23. If R is left Artinian and N is a nil subset satisfying the condition that for any a_1, a_2 in N there is $\nu = \nu(a_1, a_2) \in \mathbb{Z}$ with $a_1a_2 + \nu a_2a_1 \in N$, then N is nilpotent. 52

15.26 (Wedderburn's Principal Theorem). If R is a f.d. algebra over an algebraically closed field F, then $R = S \oplus J$ where S is a subalgebra of R isomorphic to R/J. 54

15A.2 (Jacobson Density Theorem for simple modules). Suppose M is a simple R-module, and $D = \mathrm{End}_R M$. For any $n \in \mathbb{N}$, any D-independent elements $a_1, \ldots, a_n \in M$, and any elements $b_1, \ldots, b_n$ of M, there is r in R such that $ra_i = b_i$ for $1 \le i \le n$. 57

15A.4. If A is a subalgebra of $M_n(F) = \mathrm{End}\, F^{(n)}$ for F an algebraically closed field, and $F^{(n)}$ is simple as an A-module, then $A = M_n(F)$. 58

15A.5 (Amitsur). $\mathrm{Jac}(R[\lambda]) = 0$ whenever R has no nonzero nil ideals. 58

15A.8 (Amitsur). If R is a division algebra over a field F such that $\dim_F R < |F|$, then R is algebraic over F. 60

15B.4 (Kolchin). If S is a monoid of unipotent matrices of $M_n(F)$ with F algebraically closed field F, then S can be simultaneously triangularized via a suitable change of base. 61

E15.3. A ring R is primitive iff R has a left ideal comaximal with all prime ideals. 167

E15.6. Any prime ring having a faithful module of finite composition length is primitive. 167

E15.7. For $W = \mathrm{End}\, M_D$ and $f \in W$, the left ideal Wf is minimal iff f has rank 1. Also, the set of elements of W having finite rank is an ideal of W, which is precisely $\mathrm{soc}(W)$. 168

E15.21. For any semiprime ring, $\mathrm{soc}(R)$ is also the sum of the minimal right ideals of R. 169

E15.24. $\mathrm{Jac}(R)$ is a quasi-invertible ideal that contains every quasi-invertible left ideal of R. 169

E15.26. $\mathrm{Jac}(R)$ is the intersection of all maximal right ideals of R. 169

E15A.1. For any faithful simple R-module M that is infinite-dimensional over $D = \mathrm{End}_R M$, and each n, $M_n(D)$ is isomorphic to a homomorphic image of a subring of R. 170

E15A.3. If W is a finite normalizing extension of R, then any simple W-module is a finite direct sum of simple R-modules. 170

E15A.4. $\mathrm{Jac}(R) \subseteq \mathrm{Jac}(W)$ for any finite normalizing extension W of R. 170

E15A.6. $R \cap \mathrm{Jac}(W) \subseteq \mathrm{Jac}(R)$ whenever the ring R is a direct summand of W as an R-module. 171

E15A.8. For any algebra W over a field, every element of $\mathrm{Jac}(W)$ is either nilpotent or transcendental. 171

E15A.9 (Amitsur). $\mathrm{Jac}(R)$ is nil whenever R is an algebra over an infinite field F satisfying the condition $\dim_F R < |F|$. 171

E15B.9. Kolchin's Problem has an affirmative answer for locally solvable groups and for locally metabelian groups. 172

E15B.12. (Derakhshan). Kolchin's Problem has an affirmative answer in characteristic 2. 172

Chapter 16.

16.17. If $L < R$ and $Rs \cap L = 0$ with $s \in R$ left regular, then the left ideals $L, Ls, Ls^2, \ldots$ are independent. 70

16.23 (Goldie). A ring R has a semisimple left ring of fractions iff R satisfies the following two properties: (i) $Rs <_e R$ for each regular element s. (ii) Every large left ideal L of R contains a regular element. 72

16.24. Any ring R satisfying ACC(ideals) has only finitely many minimal prime ideals, and some finite product of them is 0. 74

16.26 (Levitzki). Any semiprime ring satisfying ACC on left ideals of the form $\{\ell(r) : r \in R\}$ has no nonzero nil right ideals and no nonzero nil left ideals. 75

16.29 (Goldie). Any semiprime left Noetherian ring has a semisimple left ring of fractions. Any prime left Noetherian ring R has a simple Artinian left ring of fractions. 75, 76

16.31. Generalization of Theorem 15.23 to left Noetherian rings. 77

16.35. Any left Noetherian ring R has IBN. 78

16.46 (Fitting's Lemma). If M has finite composition length n, then $M = f^n(M) \oplus \ker f^n$ for any map $f: M \to M$; furthermore, f restricts to an isomorphism on $f^n(M)$ and a nilpotent map on $\ker f^n$. 81

E16.4 (Levitzki). A ring R is semiprime iff $N(R) = 0$. 173

E16.6. The upper nilradical of R is the intersection of certain prime ideals, and is a nil ideal that contains all the nil ideals of R. 173

E16.8. If R is weakly primitive, then R is a primitive ring. 174

E16.12. The construction of the ring $S^{-1}R$, for any denominator set S of R. 174

E16.15 (Goldie's Second Theorem). A ring R has a semisimple left ring of fractions iff R is a semiprime left Goldie ring. 175

E16.16 (Goldie's First Theorem). The ring of fractions of any prime Goldie ring is simple Artinian. 175

E16.17. $ab = 1$ implies $ba = 1$ in a left Noetherian ring. 175

E16.25 (Martindale). If R is a prime ring and $a, b \in R$ with $arb = bra$ for all $r \in R$, then $a = cb$ for some c in the extended centroid. 177

E16.29. (Wedderburn-Krull-Schmidt-Azumaya-Beck). For any finite direct sum of LE-modules, every other decomposition as a direct sum of indecomposables is the same, up to isomorphism and permutation of summands. In particular, this is true for modules of finite composition length. 177

E16.30. Suppose the ring $R = Re_1 \oplus \cdots \oplus Re_t = Re'_1 \oplus \cdots \oplus Re'_{t'}$ is written in two ways as a direct sum of indecomposable left ideals. Then $t' = t$ and there is some invertible element $u \in R$ and permutation π such that $e'_{\pi(i)} = ue_iu^{-1}$ for each $1 \le i \le t$. 177

E16.33. A graded module M is gr-semisimple iff every graded submodule has a graded complement. 178

E16.34 (Graded Wedderburn-Artin.). Any gr-left Artinian, gr-simple ring has the form $\mathrm{END}(M)_D$, where M is f.g. over a gr-division ring D. 178

E16.36 (Graded First Goldie Theorem – Goodearl and Stafford). If R is graded by an Abelian group $\mathcal{G}$ and is gr-prime and left gr-Goldie, then R has a gr-simple left gr-Artinian graded ring of (left) fractions. 178

E16.40 (Bergman). $\mathrm{Jac}(R)$ is a graded ideal of any $\mathbb{Z}$-graded ring R. 179

E16A.4. The quantized matrix algebra, quantum affine space, and the quantum torus all are Noetherian domains. 180

Chapter 17.

17.12. Any domain R is either an Ore domain or contains a free algebra on two generators. 92

17.16 (The Pingpong Lemma). Suppose a group G acts on a set S, and $A, B \le G$. If S has disjoint subsets Γ_A and Γ_B satisfying $a\Gamma_B \subseteq \Gamma_A$, $b\Gamma_A \subseteq \Gamma_B$, and $b\Gamma_B \cap \Gamma_B \neq \emptyset$ for all $a \in A \setminus \{e\}$ and $b \in B \setminus \{e\}$, then A and B interact freely. 94

17.20. If $0 \to M_1 \stackrel{f_1}{\to} M_2 \stackrel{f_2}{\to} \cdots \to M_k \to 0$ is an exact sequence of f.g. modules over a left Artinian ring, then $\sum_{j=1}^{k} (-1)^j \ell(M_j) = 0$. 99

17.25 (König Graph Theorem). Any infinite connected, directed graph has an infinite path. 102

17.38. The Hilbert series of a commutative affine algebra is rational.107

17.49. Any commutative affine algebra has integral Gel′fand-Kirillov dimension, equal both to its Krull dimension and to its transcendence degree. For any algebra with filtration whose associated graded algebra is commutative affine, the Gel′fand-Kirillov dimension is an integer. 111

17.55 (Bergman Gap Theorem). The Gel′fand-Kirillov dimension cannot be between 1 and 2. 112

17.60. The growth rate of each nilpotent group is polynomially bounded. 115

17.61 (Milnor-Wolf). Any f.g. virtually solvable group of subexponential growth is virtually nilpotent. 117

17.66. Every f.g. linear group of subexponential growth is of polynomial growth. 120

17A.9 (Nielsen-Schreier). Every subgroup of a free group is free. 124

17B.5 (The Diamond Lemma). A reduction procedure is reduction-unique on A iff for each $r \in A$ and any reductions ρ, τ, the elements $\rho(r)$ and $\tau(r)$ have chains of reductions arriving at the same element. 128

17B.7. The word problem is solvable in any group satisfying Dehn's algorithm. 130

17B.13 (Bergman). Any set of relations can be expanded to a set of relations for which any given word h becomes reduction-unique. 133

17C.2. The generalized BP has a positive answer for solvable groups.134

E17.8. The free group on a countably infinite set can be embedded into the free group $\mathcal{G}$ on two letters. 181

E17.9. The free group $\mathcal{G}$ can be embedded into $\mathrm{GL}(2, F)$. 181

E17.13. $D[[M]]$ is a division ring, for any ordered group M and any division ring D. 181

E17.22. γ_t/γ_{t+1} is a free f.g. Abelian group, for every t. 182

E17.23 (Magnus-Witt). The free group $\mathcal{G}$ is an ordered group. 182

E17.24. $F[[\mathcal{G}]]$ is a division ring containing the free algebra $F\{X\}$. 182

E17.31 (Generalized Artin-Tate Lemma). If is an affine algebra is f.g. over a commutative (not necessarily central) subalgebra C, then C is affine. 183

E17.32. Any affine algebra that is f.g. over a commutative subalgebra has a rational Hilbert series with respect to a suitable generating set. 183

E17.37. $\mathrm{GK}(R/I) \leq \mathrm{GK}(R) - 1$ for any $I \lhd R$ containing a regular element of R. 183

E17.44. Under the hypotheses of Theorem 17.60, the nilpotent group N has polynomial growth of degree $\sum_j j d_j$. 184

E17A.1. The symmetric group S_n has the Coxeter presentation $\sigma_i^2 = 1$, $(\sigma_i\sigma_{i+1})^3 = 1$, and $(\sigma_i\sigma_j)^2 = 1$ for $|j - i| > 1$. 184

E17A.7. Any subgroup of index m in a free group of rank n is free of rank $mn - m + 1$. 185

E17A.9. Any group G is the fundamental group of a complex $\mathcal{K}$ of dimension 2. G is finitely presented iff $\mathcal{K}$ can be taken finite. 186

E17B.1. Any set of relations can be expanded to a Gröbner-Shirshov basis. 187

E17C.1. The Burnside group $B(m, 3)$ is finite for all m. 187

E17C.3. The Burnside group $B(m, 4)$ is finite for all m. 188

E17C.7, E17C.8. Grigorchuk's group is infinite but torsion; every element has order a power of 2. 189

Chapter 18.

18.4. Any balanced map $\psi\colon M \times N \to G$ yields a group homomorphism $\overline{\psi}\colon M \otimes N \to G$ given by $\overline{\psi}(a \otimes b) = \psi(a, b)$. 140

18.5. For any map $f\colon M \to M'$ of right R-modules and map $g\colon N \to N'$ of R-modules, there is a group homomorphism $f \otimes g\colon M \otimes_R N \to M' \otimes N'$ given by $(f \otimes g)(a \otimes b) = f(a) \otimes g(b)$. 140

18.11. $(M_1 \oplus \ldots \oplus M_t) \otimes N \cong (M_1 \otimes N) \oplus \cdots \oplus (M_t \otimes N)$. 142

18.12. Suppose M is a free right R-module with base $B = \{b_i : i \in I\}$, and N is an R-module. Then every element of $M \otimes N$ can be written uniquely in the form $\sum_{i \in I} b_i \otimes v_i$ for v_i in N. 143

18.13. $C^{(m)} \otimes_C C^{(n)} \cong C^{(mn)}$. 144

18.15. $M_1 \otimes_{R_2} (M_2 \otimes_{R_3} M_3) \cong (M_1 \otimes_{R_2} M_2) \otimes_{R_3} M_3$. 144

18.16. $\tau\colon A \otimes_C B \cong B \otimes_C A$. 145

18.21. If A and B are C-algebras, then $A \otimes_C B$ is also a C-algebra with multiplication $(a \otimes b)(a' \otimes b') = aa' \otimes bb'$ and $c(a \otimes b) = ca \otimes b$. 147

18.25. The following algebra isomorphisms hold for any C-algebras: $A \otimes_C C \cong C \otimes_C A \cong A$; $A_1 \otimes A_2 \cong A_2 \otimes A_1$; $A_1 \otimes (A_2 \otimes A_3) \cong (A_1 \otimes A_2) \otimes A_3$. 149

18.29′. A finite field extension $K \supseteq F$ is separable iff the ring $K \otimes_F K$ is semisimple. 151

18.31. Any splitting field K of an F-algebra R contains some subfield K_0 f.g. over F such that K_0 also splits R. 152

18.33. If R is simple with center a field F, and W is an F-algebra, then any nonzero ideal I of the tensor product $R \otimes_F W$ contains $1 \otimes w$ for some $w \in W$. In particular, if W is simple, then $R \otimes_F W$ is also simple. 152

18.36. $M_m(C) \otimes M_n(C) \cong M_{mn}(C)$. 153

18.41. The tensor product of two integral domains over an algebraically closed field F is an integral domain. 157

18.42. If X and Y are affine varieties over an algebraically closed field F, then $X \times Y$ is an affine variety, with $F[X] \otimes F[Y] \cong F[X \times Y]$. 157

18.44. $\Phi : \mathrm{Hom}_R(A \otimes_S B, C) \cong \mathrm{Hom}_S(B, \mathrm{Hom}_R(A, C))$. 158

E18.2. $(\bigoplus_{i \in I} M_i) \otimes N \cong \bigoplus_{i \in I} (M_i \otimes N)$. 189

E18.7. If $K \to N \to P \to 0$ is an exact sequence of right R-modules, then $K \otimes M \to N \otimes M \to P \otimes M \to 0$ is also exact. 190

E18.12. $C(V, Q)$ has an involution $(*)$ satisfying $v^* = v$, $\forall v \in V$. 191

E18.16. For any separable field extension K of F, $K \otimes_F K$ has a simple idempotent e with $(a \otimes b)\, e = (b \otimes a)\, e$ for all $a, b \in K$. 191

E18.18 (Wedderburn's Principal Theorem). Any finite-dimensional algebra R over a perfect field F has a Wedderburn decomposition $R = S \oplus J$ for a suitable semisimple subalgebra $S \cong R/J$ of R. 191

E18.19. The tensor product of two reduced algebras over an algebraically closed field is reduced. 191

E18.23 (Amitsur). If R is an algebra without nonzero nil ideals over a field F, then $\mathrm{Jac}(R \otimes_F F(\lambda)) = 0$. 192

E18.24. $K \otimes_F \mathrm{Jac}(R) \subseteq \mathrm{Jac}(K \otimes_F R)$ whenever $K \supseteq F$ are fields and R is an algebra over F, equality holding if K/F is separable. 192

Chapter 19.

19.18. For any vector space V over a field F, there is a 1:1 correspondence among: group representations $\rho : G \to \mathrm{GL}(V)$, algebra representations $F[G] \to \mathrm{End}_F V$, G-space structures on V, and $F[G]$-module structures on V. 206

19.22. A group representation ρ of degree n is reducible iff there is a representation τ equivalent to ρ for which each matrix $\tau(g)$, $g \in G$, has the form (19.4) (for suitable $1 \le m < n$). 208

19.26 (Maschke's Theorem). $F[G]$ is a semisimple ring, for any finite group G whose order is not divisible by $\mathrm{char}(F)$. 210

19.33. Any finite group G has a splitting field that is finite over $\mathbb{Q}$. 212

19.36. For any splitting field F of the group G, a representation ρ of degree n is irreducible iff $\{\rho(g) : g \in G\}$ spans $M_n(F)$. 213

19.38. The following are equivalent, for F a splitting field of a finite group G: (i) G is Abelian; (ii) The group algebra $F[G]$ is commutative; (iii) $F[G] \cong F \times F \times \cdots \times F$; (iv) Every irreducible representation of G has degree 1. 213

19.42. $\mathrm{Cent}(C[G])$ is free as a C-module. 215

19.43. The following numbers are equal, for F a splitting field of a finite group G: (i) the number of conjugacy classes of G; (ii) the number of inequivalent irreducible representations of G; (iii) the number of simple components of $F[G]$; (iv) $\dim_F \mathrm{Cent}(F[G])$. 216

19.48. Any complex irreducible representation of G of degree n_i either is extended from a real irreducible representation or corresponds to a real irreducible representation of degree $2n_i$. 218

19.61. If $\mathrm{char}(F) = 0$ or $\mathrm{char}(F) > n$, then $I_\lambda = \bigoplus_{T_\lambda \text{ standard}} F[S_n] e_{T_\lambda}$. 223

19.64 (Frame, Robinson, and Thrall). $f^\lambda = \frac{n!}{\prod h_{i,j}}$. 226

19A.4. If $\{(a, b) : a \in A,\, b \in B\}$ is finite for $A, B \lhd G$, then the group (A, B) is finite. 229

19A.9 (Burnside, Schur). In characteristic 0, every linear group of finite exponent is finite, and any f.g. periodic linear group is finite. 231

19A.12. Every open subgroup of a quasicompact group is closed of finite index. 234

19A.16. Every continuous f.d. representation of a compact (Hausdorff) group is a finite direct sum of continuous irreducible representations. 235

19A.19. Any Lie homomorphism $\phi : G \to H$ of Lie groups (G connected) is uniquely determined by its tangent map $d_e\phi$. 236

19B.4. In any algebraic group G, each open subgroup of G is closed of finite index, each closed subgroup H of G of finite index is open, and G_e is clopen of finite index in G. 239

19B.11. If $H \leq G$, then $\overline{H} \leq G$; furthermore, if H contains a nonempty open subset U of $\overline{H}$, then H is closed. 240

19B.19. Every affine algebraic group is linear. 243

19B.21 (The Tits alternative). Every f.g. linear group either is virtually solvable or contains a free subgroup. 244

19B.24 (Breuillard-Gelanter). Any f.g. linear group contains either a free subgroup that is Zariski dense (in the relative topology), or a Zariski open solvable subgroup. 248

E19.6 (Schur's Lemma, representation-theoretic formulation). For F a splitting field for G, $\mathrm{End}_{F[G]}(L_i) \cong F$ and $\mathrm{Hom}_{F[G]}(L_i, L_j) = 0$ for all $i \neq j$, where L_i denotes the module corresponding to ρ_i. 355

E19.13. A representation ρ of finite degree ρ is completely reducible whenever its G-space has a G-invariant Hermitian form. 356

E19.31. $C[G]$ is semiprime, for any group G and any integral domain C of characteristic 0. 358

E19.34 (Herstein; Amitsur). $\mathrm{Jac}(F[G]) = 0$ for any uncountable field F of characteristic 0. 358

E19.42 (Schur's Double Centralizer Theorem.) Suppose V is any f.d. vector space over a field of characteristic 0. The diagonal action of $\mathrm{GL}(V)$ and the permutation action of S_n on $V^{\otimes n} = V \otimes \cdots \otimes V$ centralize each other, and provide algebra homomorphisms $\hat{\rho}\colon F[\mathrm{GL}(V)] \to \mathrm{End}_F V^{\otimes n}$ and $\hat{\tau}\colon F[S_n] \to \mathrm{End}_F V^{\otimes n}$. Their respective images are the centralizers of each other in $\mathrm{End}_F V^{\otimes n}$. 359

E19A.6 (Burnside). Any f.g. periodic linear group is finite. 361

E19A.8 (Schur). Each periodic subgroup of $\mathrm{GL}(n, \mathbb{C})$ consists of unitary matrices with respect to some positive definite Hermitian form. 361

E19A.11 (Jordan). Any unitary subgroup $G \subseteq \mathrm{GL}(n, \mathbb{C})$ has a normal Abelian subgroup of index bounded by $(\sqrt{8n}+1)^{2n^2} - (\sqrt{8n}-1)^{2n^2}$. 362

E19A.16. For any continuous complex representation of degree n of a compact topological group G, the vector space $\mathbb{C}^{(n)}$ has a positive definite G-invariant Hermitian form. 362

E19A.18. Any continuous f.d. representation of G having a positive definite G-invariant Hermitian form is completely reducible 363

E19A.30 (Artin's combing procedure). The kernel of the map $P_n \to P_{n-1}$ obtained by cutting the n-th strand is the free group of rank $n-1$. 364

E19A.34. The braid group B_n satisfies $B_n' = (B_n', B_n)$. 364

E19B.7. The Tits alternative also over fields of any characteristic. 366

E19B.14. The commutator group of two closed subgroups of an algebraic group G is closed. In particular, all the derived subgroups of G are closed, and all subgroups in its upper central series are closed. 367

E19B.16. For F algebraically closed, any connected solvable algebraic subgroup G of $\mathrm{GL}(n, F)$ is conjugate to a subgroup of $\mathrm{T}(n, F)$. 367

Chapter 20.

20.5. The characters $\chi_1, \ldots, \chi_t$ comprise an orthonormal base of $\mathcal{R}$ with respect to the Schur inner product. 251

20.10. $\sum_{g \in G} \chi_i(ga)\overline{\chi_j(g)} = \frac{\delta_{ij}|G|\chi_i(a)}{n_i}$ for each $a \in G$. 252

20.14 (Schur I). $\delta_{ik}|G| = \sum_{j=1}^{t} m_j \chi_{ij}\overline{\chi_{kj}}$. 256

20.15 (Schur II). $\sum_{i=1}^{t} \chi_{ij}\overline{\chi_{ik}} = \delta_{jk}\frac{|G|}{m_k}$. 256

20.18 (Frobenius). n_i divides $|G|$ for each i. 258

20.20. If $\gcd(n_i, m_j) = 1$, then either $\chi_{ij} = 0$ or $|\chi_{ij}| = n_i$. 259

20.22. In a finite nonabelian simple group, the size of a conjugacy class cannot be a power (other than 1) of a prime number. 259

20.24 (Burnside). Every group of order $p^u q^v$ (p, q prime) is solvable.260

20.32. The character table of $G \times H$ is the tensor product of the character tables of G and of H. 262

20.42 (Frobenius Reciprocity Theorem). For $F \subseteq \mathbb{C}$ a splitting field of a finite group G, if σ is an irreducible representation of a subgroup H and ρ is an irreducible representation of G, then the multiplicity of ρ in σ^G is the same as the multiplicity of σ in ρ_H. 267

20.43. For $H < K < G$ and a representation ρ of H, the representations $(\rho^K)^G$ and ρ^G are equivalent, $(\rho_1 \oplus \rho_2)^G$ and $\rho_1^G \oplus \rho_2^G$, are equivalent, and $\rho^G \otimes \sigma$ and $(\rho \otimes \sigma_H)^G$ are equivalent for any representation σ of G. 268

20.44 (Artin). Every complex character of a group is a linear combination (over $\mathbb{Q}$) of complex characters induced from cyclic subgroups. 269

E20.20. n_i divides $[G\colon Z_i]$ for each i. 370

E20.22. The degree of each irreducible character of G divides $[G\colon A]$ for any Abelian normal subgroup A. 370

E20.27. For any representation ρ of finite degree of a subgroup $H \subseteq G$, the contragredient $(\rho^G)^*$ of the induced representation is equivalent to the induced representation $(\rho^*)^G$. 371

Chapter 21.

21.21. For F algebraically closed, if L is a Lie subalgebra of $\subseteq gl(n,F)$ and $a = \mathbf{s} + \mathbf{n}$ is the Jordan decomposition of $a \in L$, then $\mathrm{ad}_a = \mathrm{ad}_\mathbf{s} + \mathrm{ad}_\mathbf{n}$ is the Jordan decomposition of ad_a. 281

21.27. If L is a Lie subalgebra of R^- and ad_a is nilpotent for every $a \in L$, then $\mathrm{ad}\, L$ is nilpotent under the multiplication of R, and L is a nilpotent Lie algebra. 283

21.29 (Engel). Any Lie algebra $L \subseteq gl(n,F)$ of nilpotent transformations becomes a Lie subalgebra of the algebra of strictly upper triangular matrices under a suitable choice of base. 284

21.32 (Lie). If a Lie subalgebra L of $gl(n,F)$ acts solvably on $F^{(n)}$, with F an algebraically closed field, then L acts in simultaneous upper triangular form with respect to a suitable base of $F^{(n)}$. 285

21.38. If $L \subseteq gl(n,F)$ in characteristic 0 such that $\mathrm{tr}(aL') = 0$ for all $a \in L$, then L' is a nilpotent Lie algebra. 287

21.41 (Cartan's first criterion). A f.d. Lie algebra L of characteristic 0 is solvable iff its Killing form vanishes identically on L'. 288

21.47 (Cartan's second criterion). A f.d. Lie algebra L of characteristic 0 is semisimple iff its Killing form is nondegenerate. 289

21.51. Any f.d. semisimple Lie algebra L of characteristic 0 is a direct sum $\bigoplus S_i$ of simple nonabelian Lie subalgebras S_i, with each $S_i \lhd L$, and any Lie ideal of L is a direct sum of some of the S_i. 290

21.53. The trace bilinear form of any representation ρ of a f.d. semisimple Lie algebra is nondegenerate. 290

21.54 (Zassenhaus). Every derivation of a f.d. semisimple Lie algebra L of characteristic 0 is inner. 291

21.57. The Casimir element satisfies $\mathrm{tr}(c_\rho) = n$ and $[\rho(L), c_\rho] = 0$. 292

21.58 (Weyl). Any f.d. representation of a f.d. semisimple Lie algebra L (of characteristic 0) is completely reducible. 292

21.61. For any given nilpotent Lie subalgebra N of a f.d. Lie algebra L, there exists a unique root space decomposition $L = \bigoplus_\mathbf{a} L_\mathbf{a}$. 295

21.64. $L_\mathbf{b} \perp L_\mathbf{a}$ for any roots $\mathbf{a} \neq -\mathbf{b}$. 296

21.71, 21.72. Any f.d. semisimple Lie algebra over an algebraically closed field of characteristic 0 has a Cartan subalgebra $\mathfrak{h}$, which is its own nullspace under the corresponding root space decomposition. The restriction of the Killing form to $\mathfrak{h}$ is nondegenerate. $\mathfrak{h}$ is Abelian, and ad_h is semisimple for all $h \in \mathfrak{h}$. 298

21.79. For any root $\mathbf{a}$, $\dim L_\mathbf{a} = \dim L_{-\mathbf{a}} = 1$, and $k\mathbf{a}$ is not a root whenever $1 < |k| \in \mathbb{N}$. 301

21.80. $\langle h_1, h_2\rangle = \sum_{\mathbf{a}\neq 0} \mathbf{a}(h_1)\mathbf{a}(h_2), \quad \forall h_1, h_2 \in \mathfrak{h}$. 301

21.84. Any simple $\hat{L}_\mathbf{a}$-module V has an eigenspace decomposition $V = V_m \oplus V_{m-2} \oplus \cdots \oplus V_{-(m-2)} \oplus V_{-m}$, where each component $V_{m-2j} = Fv_j$ is a one-dimensional eigenspace of $h_\mathbf{a}$ with eigenvalue $m - 2j$. In particular, V is determined up to isomorphism by its dimension $m + 1$. 303

21.88. $[L_\mathbf{a} L_\mathbf{b}] = L_{\mathbf{b}+\mathbf{a}}$ whenever $\mathbf{a}, \mathbf{b}$, and $\mathbf{b} + \mathbf{a}$ are roots. 305

21.91. $\langle \mathbf{a}, \mathbf{a}\rangle > 0$ and $\langle \mathbf{a}, \mathbf{b}\rangle \in \mathbb{Q}$ for all nonzero roots $\mathbf{a}, \mathbf{b}$. The bilinear form given by Equation (21.18) restricts to a positive form on ${\mathfrak{h}^*}_0$, the $\mathbb{Q}$-subspace of $\mathfrak{h}^*$ spanned by the roots, and $\mathfrak{h}^* = {\mathfrak{h}^*}_0 \otimes_\mathbb{Q} F$. 306

21.96. $\langle \mathbf{a}, \mathbf{b}\rangle \leq 0$ for all $\mathbf{a} \neq \mathbf{b} \in P$. 308

21.97. The set of simple roots is a base of the vector space V and is uniquely determined by the given order on V. 308

21.102. The Cartan numbers m_{ij} satisfy $m_{ij}m_{ji} < 4$. 310

21.103. The Cartan numbers are integers. 311

21.108. Suppose $S = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a simple root system for the semisimple Lie algebra L. Take $e_i \in L_{\mathbf{a}_i}$, $e_i' \in L_{-\mathbf{a}_i}$, and $h_i = [e_i f_i]$. Writing any positive root $\mathbf{a} = \mathbf{a}_{i_1} + \cdots + \mathbf{a}_{i_\ell}$, let $x_\mathbf{a} = [e_{i_1} e_{i_2} \cdots e_{i_\ell}]$ and $y_\mathbf{a} = [f_{i_1} f_{i_2} \cdots f_{i_\ell}]$. Then $\{h_1, \ldots, h_n\}$ together with the $x_\mathbf{a}$ and $y_\mathbf{a}$ comprise a base of L. 313

21.110. The Lie multiplication table of L (with respect to the base in Theorem 21.108) has rational coefficients. 314

21.111. The split f.d. semisimple Lie algebra L is simple iff its simple root system is indecomposable. 315

21.115, 21.116. Any indecomposable generalized Cartan matrix A is of finite, affine, or indefinite type. The symmetric bilinear defined by A is positive definite iff A has finite type, and is positive semidefinite (of corank 1) iff A has affine type. 317

21B.18, 21B.19. Suppose the composition algebra $(\mathcal{A}, *)$ is the ν-double of $(A, *)$. If A is associative, then $\mathcal{A}$ is alternative. If $\mathcal{A}$ is associative, then A must be commutative. 326

21B.22. (Herstein). If R is a simple associative algebra, with $\frac{1}{2} \in R$, then R^+ is simple as a Jordan algebra. 328

21C.5. (PBW Theorem). The map $\nu_L \colon L \to U(L)^-$ is 1:1. 333

E21.10. In characteristic $\neq 2$, the classical Lie algebra B_n is simple for each $n \geq 1$, and C_n and D_n are simple Lie algebras for all $n > 2$. 372

E21.27 (Herstein). For any associative simple ring R of characteristic $\neq 2$, the only proper Lie ideals of R' are central. 373

E21.28 (Herstein). If T is an additive subgroup of a simple ring R of characteristic $\neq 2$ such that $[T, R'] \subseteq T$, then either $T \supseteq R'$ or $T \subseteq Z$. 374

E21.41. The radical of a Lie algebra is contained in the radical of the trace bilinear form with respect to any representation. 375

E21.42. The Casimir element of an irreducible Lie representation is always invertible. 375

E21.44 (Whitehead's First Lemma). For any f.d. Lie module V and linear map $f \colon L \to V$ satisfying $f([ab]) = af(b) - bf(a)$, $\forall a, b \in L$, there is $v \in V$ such that $f(a) = av$, $\forall a \in L$. 376

E21.47 (Whitehead's Second Lemma). For any f.d. semisimple Lie algebra L of characteristic 0 and f.d. Lie module V with $f \colon L \times L \to V$ satisfying $f(a, a) = 0$ and $\sum_{i=1}^{3} f(a_i, [a_{i+1}, a_{i+2}]) + a_i f(a_{i+1}, a_{i+2}) = 0$, subscripts modulo 3, there is a map $g \colon L \to V$ with $f(a_1, a_2) = a_1 g(a_2) - a_2 g(a_1) - g([a_1 a_2])$. 376

E21.48 (Levi's Theorem). Any f.d. Lie algebra L of characteristic 0 can be decomposed as vector spaces $L = S \oplus I$, where $I = \operatorname{rad} L$ and $S \cong L/I$ is a semisimple Lie subalgebra. 377

E21.50. $L' \cap \operatorname{rad}(L)$ is Lie nilpotent, for any f.d. Lie algebra L of characteristic 0. 377

E21.60. $\langle \mathbf{a}, \mathbf{a} \rangle = \sum_{\mathbf{b}} \langle \mathbf{a}, \mathbf{b} \rangle^2$ for any root $\mathbf{a}$. 378

E21.63. The formulas $[e_{j_1} e_{j_2} \cdots e_{j_\ell} h_i] = -\sum_{u=1}^{\ell} m_{ij_u} [e_{j_1} e_{j_2} \cdots e_{j_\ell}]$ and $[f_{j_1} f_{j_2} \cdots f_{j_\ell} h_i] = \sum_{u=1}^{\ell} m_{ij_u} [f_{j_1} f_{j_2} \cdots f_{j_\ell}]$ hold in Theorem 21.108. 378

E21.67. Every root system of a simple Lie algebra L has a unique maximal root. 379

E21.70. The Weyl group acts transitively on simple root systems. 379

E21.71–E21.73. Construction of the Kac-Moody Lie algebra and its root space decomposition. 379, 380

E21.75. Equivalent conditions for an indecomposable, symmetric generalized Cartan matrix to have finite type. 380

E21.77, E21.78. Construction of the Witt and Virasoro algebras. 380

E21.79 (Farkas). For $\mathbf{a}_i = (\alpha_{i1}, \ldots, \alpha_{i\ell})$, $1 \leq i \leq k$. the system $\sum_j \alpha_{ij} \lambda_j > 0$ of linear inequalities for $1 \leq i \leq k$ has a simultaneous solution over $\mathbb{R}$ iff every non-negative, nontrivial, linear combination of the $\mathbf{a}_i$ is nonzero. 381

E21.80 (The Fundamental Theorem of Game Theory). If there does not exist $\mathbf{x} > 0$ in $\mathbb{R}^{(\ell)}$ with $A\mathbf{x} < 0$, then there exists $\mathbf{w} \geq 0$ (written as a row) in $\mathbb{R}^{(k)}$ with $\mathbf{w}A \geq 0$. 381

E21.81. The generalized Cartan matrix A^t has the same type as A. 381

E21.90, E21.91 (Farkas-Letzter). For any prime ring R with a Poisson bracket, there exists c in the extended centroid of R such that $[a, b] = c\{a, b\}$ for every $a, b \in R$. 383

E21A.3. The Lie product in $T(G)_e$ corresponds to the natural Lie product of derivations in $\operatorname{Lie}(G)$. 383

E21A.4. $d\varphi \colon T(G)_e \to T(H)_e$ preserves the Lie product. 383

E21A.6. Description of the classical simple Lie algebras as the Lie algebras of the algebraic groups SL, O, and Sp. 384

E21B.3. The base field $K \supset F$ of any algebra can be cut down to a field extension of finite transcendence degree over F. 385

E21B.9 (Moufang). Every alternative algebra satisfies the three identities $a\big(b(ac)\big) = (aba)c$, $c\big(a(ba)\big) = c(aba)$, and $(ab)(ca) = a(bc)a$. 386

E21B.11 (E. Artin). Any alternative algebra generated by two elements is associative. 386

E21B.19. Any composition F-algebra must be either F itself, the direct product of two copies of F (with the exchange involution), a quadratic field extension of F, a generalized quaternion algebra, or a generalized octonion algebra. 387

E21B.20 (Hurwitz). If $\mathbb{C}$ satisfies an identity $\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} z_i^2$, where z_i are forms of degree 2 in the x_i and y_j, then $n = 1, 2, 4$, or 8. 387

E21B.22 (Zorn). Every f.d. simple nonassociative, alternative algebra is a generalized octonion algebra. 388

E21B.26. The Peirce decomposition of an alternative algebra in terms of pairwise orthogonal idempotents. 388

E21B.29. Any simple alternative algebra A containing three pairwise orthogonal idempotents e_1, e_2, and e_3 is associative. 388

E21B.37 (Glennie). Any special Jordan algebra satisfies the Glennie identity. 389

E21B.39. (Herstein). $\mathcal{S}(R, *)$ is Jordan simple for any simple associative algebra with involution of characteristic $\neq 2$. 390

E21C.4. $U(L)$ is an Ore domain, for any Lie algebra L of subexponential growth. 391

E21C.13 (Ado). Any f.d. Lie algebra of characteristic 0 is linear. 392

E21C.17. $U_q(sl(2, F))$ is a skew polynomial ring. 393

E21C.21. $U_q(L)$ is a Noetherian domain, for any f.d. semisimple Lie algebra L of characteristic 0. 394

Chapter 22.

22.11. Any connected Dynkin diagram is either A_n, $B_n = C_n$, D_n, E_6, E_7, E_8, F_4, or G_2 of Example 22.2. 342

22.13. If any single vertex of the extended Dynkin diagram of a simple affine Lie algebra is erased, the remaining subdiagram is a disjoint union of Dynkin diagrams (of finite type). 345

22.22. For $i \neq j$, the Coxeter bilinear form restricts to a positive semidefinite form on the two-dimensional subspace $Fe_i + Fe_j$, which is positive definite iff $\circ(\sigma_i\sigma_j) < \infty$. 348

22.25. The only abstract Coxeter graphs whose quadratic forms are positive definite are A_n, D_n, E_6, E_7, and E_8. 350

22.28. (Bernstein, Gel′fand, and Ponomarev). If an abstract Coxeter graph $(\Gamma; \nu)$ has only finitely many nonisomorphic indecomposable representations, then its quadratic form is positive definite. 352

E22.1–E22.4. Construction of the classical Lie algebras from their Dynkin diagrams. 394

E22.10. For any generalized Cartan matrix A of affine type, any proper subdiagram of its Dynkin diagram is the disjoint union of Dynkin diagrams of simple f.d. Lie algebras. 396

E22.19. Any finite reflection group is Coxeter. 397

E22.20. Any two positive systems Φ_1 and Φ_2 are conjugate under some element of the Weyl group. 398

E22.23. Each $m_{i,j} \in \{2, 3, 4, 6\}$ for any crystallographic group. 398

E22.26. The bilinear form of any finite Coxeter group $\mathcal{W}$ is positive definite. 398

E22.36. Every finite Coxeter group is a reflection group. 400

Chapter 23.

23.11. Any t-alternating polynomial f is an identity for every algebra spanned by fewer than t elements over its center. 410

23.26 (Razmyslov). There is a 1:1 correspondence between multilinear central polynomials of $M_n(F)$ and multilinear 1-weak identities that are not identities. 415

23.31 (Kaplansky). Any primitive ring R satisfying a PI of degree d is simple of dimension n^2 over its center, for some $n \leq \left[\frac{d}{2}\right]$. 418

23.32. A semiprime PI-ring R has no nonzero left or right nil ideals.418

23.33. Any semiprime PI-ring R has some PI-class n, and every ideal A intersects the center nontrivially. 419

23.34 (Posner et al). The ring of central fractions of a prime PI-ring R is simple and f.d. over the field of fractions of $\mathrm{Cent}(R)$. 419

23.35. Extension of Theorem 15.23 to PI-rings. 419

23.39. Suppose R has PI-class n and center C, and $1 \in h_n(R)$. Then R is a free C-module of rank n; also, there is a natural 1:1 correspondence between {ideals of R} and {ideals of C}. 421

23.48. If R is an algebra over an infinite field F, and H is any commutative F-algebra, then R is PI-equivalent to $R \otimes_F H$. 425

23.51. The algebra of generic matrices is the relatively free PI-algebra with respect to $\mathcal{I} = \mathcal{M}_{n,C}$. 426

23.57. Suppose R satisfies a PI of degree d, and $\frac{1}{u} + \frac{1}{v} \leq \frac{2}{e(d-1)^4}$, where $e = 2.71828\cdots$. Then any multilinear polynomial of a Young tableau whose shape contains a $u \times v$ rectangle is an identity of R. 429

23A.3 (Shirshov's Dichotomy Lemma). For any ℓ, d, k, there is $\beta \in \mathbb{N}$ such that any word w of length $\geq \beta$ in ℓ letters is either d-decomposable or contains a repeating subword of the form u^k with $1 \leq |u| \leq d$. 431

23A.5. Any hyperword h is either d-decomposable or has the form vu^∞ for some initial subword v and some subword u with $|u| < d$. 431

23A.6 (Shirshov's First Theorem). If $R = C\{r_1, \ldots, r_\ell\}$ satisfies a PI, and each word in the r_i of length $\leq d$ is integral over C, then R is f.g. as a C-module. 432

23A.7. If R is affine without 1 and satisfies a PI of degree d, and if each word in the generators of length $\leq d$ is nilpotent, then R is nilpotent. 432

23A.10. Any prime PI-algebra and its characteristic closure have a common nonzero ideal. 433

23A.11 (Kemer). Any affine PI-algebra over a field F of characteristic 0 is PI-equivalent to a finite-dimensional algebra. 434

23A.19. For any PI algebra R, the following assertions are equivalent for any multilinear polynomial f of degree n: $f \in \mathrm{id}(R)$; ${f_I}^* \in \mathrm{id}_2(R \otimes G)$ for some subset $I \subseteq \{1, \ldots, n\}$; ${f_I}^* \in \mathrm{id}_2(R \otimes G)$ for every subset of $\{1, \ldots, n\}$. 436

23A.22 (Kemer). Let R be a PI-superalgebra, and $f = f(x_1, \ldots, x_n) = \sum_{\pi \in S_n} \alpha_\pi x_{\pi 1} \cdots x_{\pi n}$. Then $f \in \mathrm{id}(\mathcal{G}(R))$ iff ${f_I}^* \in \mathrm{id}_2(R)$ for every subset $I \subseteq \{1, \ldots, n\}$. 437

23A.23 (Kemer). There is a 1:1 correspondence from {varieties of superalgebras} to {varieties of algebras} given by $R \mapsto \mathcal{G}(R)$. 437

23B.5 (Kostrikin-Zelmanov). Over a field of characteristic p, any f.g. Lie algebra satisfying the Engel identity e_{p-1} is Lie nilpotent. 442

23B.6 (Zelmanov). If a f.g. restricted Lie algebra L over a field of characteristic p satisfies the Engel identity e_n and all of its partial linearizations, then L is Lie nilpotent. 442

23B.13. The Lie algebra L of a nilpotent group G is indeed a Lie algebra and is $\mathbb{N}$-graded in the sense that $[L_i L_j] \subseteq L_{i+j}$. L is Lie nilpotent of the same index t as the nilpotence class of the group G. 444

23B.16 (Kostrikin and Zelmanov). Any sandwich algebra is Lie nilpotent. 446

E23.4. Any algebra that is f.g. as a module over a commutative affine subalgebra is representable. 563

E23.6. The Jacobson radical of a representable affine algebra is nilpotent. 564

E23.16. Every identity of an algebra over a field of characteristic 0 is a consequence of its multilinearizations. 565

E23.17. Over an infinite field, every identity is a sum of completely homogeneous identities. 565

E23.22 (Amitsur-Levitzki. The standard polynomial s_{2n} is an identity of $M_n(C)$ for any commutative ring C. 566

E23.24. Every PI-algebra has IBN. 566

E23.26 (Bell). Every prime affine PI-algebra has a rational Hilbert series. 566

E23.30 (Amitsur). Any PI-algebra R satisfies an identity s_d^k. 566

E23.32. If algebras R_1 and R_2 are PI-equivalent, then so are $M_n(R_1)$ and $M_n(R_2)$. 567

E23.36 (Regev). In characteristic 0, the T-ideal $\mathrm{id}(G)$ is generated by the Grassmann identity. 567

E23.40 (Regev). $M_n(G(p))$ satisfies the identity $s_{2n}^{n^2p+1}$. 568

E23.42 (Kemer). In any F-algebra, a suitable finite product of T-prime T-ideals is 0. Any T-ideal has only finitely many T-prime T-ideals minimal over it. 568

E23B.1. Any simple alternative, nonassociative algebra satisfies the central polynomial $[x,y]^2$. 569

E23B.4. Any Lie algebra of characteristic 3 satisfying the Engel adidentity $e_2 = X^2$ is Lie nilpotent of class ≤ 3. 570

E23B.6. For any nilpotent p-group G of exponent $n{=}p^k$, the Lie algebra $L_{\hat{\gamma}}(G)$ satisfies the multilinearized n-Engel identity $\tilde{e}_n$ and some weak Engel condition $e_{S,2n}$. 570

E23B.14 (Key step in proving Theorem 23B.16). An enveloping algebra R of a Lie algebra L is nilpotent whenever R is generated by a finite set of 1-thick sandwiches. 571

E23B.17 (Zelmanov). If a f.g. restricted Lie algebra L satisfies various Engel-type conditions, then its associative enveloping algebra R (without 1) is nilpotent. 572

Chapter 24.

24.10. If R_1 and R_2 are csa's, then $R_1 \otimes_F R_2$ is also a csa. 452

24.14. If R is a csa, then $\Phi\colon R \otimes_F R^{\mathrm{op}} \to \mathrm{End}_F R$ is an isomorphism. 453

24.15. The Brauer group $\mathrm{Br}(F)$ is a group, where $[R]^{-1} = [R]^{\mathrm{op}}$. 453

24.23, 24.24. $\mathrm{End}_K R \cong C_R(K) \otimes_F R^{\mathrm{op}}$ as K-algebras, for any F-subfield K of R. $C_R(K)$ is a K-csa and $[C_R(K):F] = [R:K]$. 455, 456

24.25. $R \otimes_F K \sim C_R(K)$ in $\mathrm{Br}(K)$. 456

24.32 (Double Centralizer Theorem). $C_R(K) \cong A \otimes_K C_R(A)$ and $[A:F][C_R(A):F] = n^2$, for any simple F-subalgebra A of a csa R, where $K = \mathrm{Cent}(A)$, 458

24.34 (Index Reduction Theorem). The index reduction factor divides the g.c.d. of $\mathrm{ind}(R)$ and $m = [L:F]$. 459

24.40 (Skolem-Noether Theorem). Suppose A_1 and A_2 are isomorphic simple subalgebras of a csa R. Any F-algebra isomorphism $\varphi\colon A_1 \to A_2$ is given by conjugation by some $u \in R^\times$. 460

24.42 (Wedderburn). Every finite division ring is a field. 461

24.44. A csa R of degree n over an infinite field F is split iff R contains an element of degree n whose minimal polynomial has a linear factor. 462

24.48′. $(K,\sigma,\beta_1) \otimes (K,\sigma,\beta_2) \sim (K,\sigma,\beta_1\beta_2)$. 464

24.50. Any F-csa R is PI-equivalent to $M_n(F)$ for $n = \deg R$. 465

24.51 (Koethe-Noether-Jacobson). Any separable subfield L of a cda D is contained in a separable maximal subfield of D. 465

24.52. Every csa is similar to a crossed product. 466

24.54. $\mathrm{UD}(n,F)$ is a division algebra of degree n (over its center) for every n and every field F of characteristic prime to n. 467

24.57. If D is a cda of degree $p^u q$ with p prime, $p \nmid q$, then there is a field extension L of F with $p \nmid [L:F]$, as well as a splitting field $L_u \supseteq L$ of D together with a sequence of subfields $L_0 = L \subset L_1 \subset L_2 \subset \cdots \subset L_u$ for which $\mathrm{ind}(D \otimes_F L_i) = p^{u-i}$ for each $0 \leq i \leq u$, and each L_i/L_{i-1} is cyclic Galois of dimension p. 468

24.62. $\exp(R)$ divides $\mathrm{ind}(R)$. If a prime number p divides $\mathrm{ind}(R)$, then p divides $\exp(R)$. 469

24.66. Any cda D is isomorphic to the tensor product of cda's of prime power index. 470

24.68 (Wedderburn). Suppose D is a cda. If $a \in D$ is a root of a monic irreducible polynomial $f \in F[\lambda]$ of degree n, then $f = (\lambda - a_n)\cdots(\lambda - a_1)$ in $D[\lambda]$, where each a_i is a conjugate of a. 472

24.73, 24.74. For any Galois extension E of F, $\mathrm{cor}_{E/F}$ induces a homomorphism of Brauer groups, and $\mathrm{cor}_{E/F}\,\mathrm{res}_{E/F}\, R \cong R^{\otimes [E:F]}$. 475

24.82 (Cohn-Wadsworth). A cda D has a valuation extending a given valuation v on F, iff v extends uniquely to a valuation of each maximal subfield of D. 480

24.85 (Hasse). Any cda D of degree n over a local field is a cyclic algebra, having a maximal subfield K isomorphic to the unramified extension of F of dimension n. 481

E24.1 (Frobenius). The only $\mathbb{R}$-cda other than R is $\mathbb{H}$. 572

E24.8 (Wedderburn's criterion). A cyclic algebra (K,σ,β) of degree n has exponent n, if β^j is not a norm from K for all $1 \leq j < n$. 573

E24.25. $(K,G,(c_{\sigma,\tau})) \otimes (K,G,(d_{\sigma,\tau})) \sim (K,G,(c_{\sigma,\tau}d_{\sigma,\tau}))$. 575

E24.31. Any p-algebra is split by a purely inseparable, finite-dimensional field extension. 575

E24.32. If $\mathrm{UD}(n, F)$ is a crossed product with respect to a group G, then every F-csa of degree n is a crossed product with respect to G. 576

E24.38. Division algebras of all degrees exist in any characteristic. 577

E24.42. When $\deg D = 3$, any element of reduced norm 1 is a multiplicative commutator. 577

E24.43. When $\deg D = 3$ and $\mathrm{char}(F) \neq 3$, any element of reduced trace 0 is an additive commutator. 577

E24.48 (The Projection Formula). $\mathrm{cor}_{L/F}(a, b; L) \sim (a, N_{L/F}(b))$ when $a \in F$. 578

E24.49 (Rosset). Any cda D of degree p is similar to the corestriction of a symbol algebra. 578

E24.51. $\mathrm{Br}(F)$ is divisible whenever F has enough m-roots of 1. 578

E24.54. $e(D/F)f(D/F) \leq [D:F]$, equality holding when the valuation is discrete and the field F is complete. 579

E24.58. $D = (K, \sigma, \pi^n)$ in Theorem 24.85. 579

E24A.7. (Plücker). The Brauer-Severi variety is a projective variety. 580

E24A.8. A geometric criterion for an n-dimensional subspace of a csa of degree n to be a left ideal. 580

Chapter 25.

25.10. Equivalent conditions for an R-module to be projective. 494

25.11. A direct sum $\oplus P_i$ of modules is projective iff each of the P_i is projective. 494

25.12′. A ring R is semisimple iff every short exact sequence of R-modules splits, iff every R-module is projective. 495

25.13 (Dual Basis Lemma). An R-module $P = \sum Ra_i$ is projective iff there are R-module maps $h_i: P \to R$ satisfying $a = \sum_{i \in I} h_i(a)a_i,\ \forall a \in P$, where, for each a, $h_i(a) = 0$ for almost all i. 495

25.24. If P and Q are modules over a commutative ring C such that $P \otimes Q \cong C^{(n)}$, then P is projective. 501

25.38 (The Snake Lemma). Any commutative diagram

$$\begin{array}{ccccccc} & & A_1'' & \xrightarrow{f_1} & A_1 & \xrightarrow{g_1} & A_1' \longrightarrow 0 \\ & & \downarrow{\scriptstyle d''} & & \downarrow{\scriptstyle d} & & \downarrow{\scriptstyle d'} \\ 0 & \longrightarrow & A_2'' & \xrightarrow{f_2} & A_2 & \xrightarrow{g_2} & A_2' \end{array}$$

gives rise to an exact sequence $\ker d'' \to \ker d \to \ker d' \to \mathrm{coker}\, d'' \to \mathrm{coker}\, d \to \mathrm{coker}\, d'$. 506

25.44. For any exact sequence $0 \to M' \to M \to M'' \to 0$ of modules and respective projective resolutions $(\mathbb{P}', d')$ and $(\mathbb{P}'', d'')$ of M' and M'', there exists a projective resolution $(\mathbb{P}, d)$ of M, such that $P_n = P_n' \oplus P_n''$ for each n, and the three projective resolutions form a commutative diagram. 509

25.45. Any short exact sequence $0 \to (\mathbf{A}'', d'') \xrightarrow{f} (\mathbf{A}, d) \xrightarrow{g} (\mathbf{A}', d') \to 0$ of complexes gives rise to a long exact sequence of the homology groups $\cdots \to H_{n+1}(\mathbf{A}'') \xrightarrow{f_*} H_{n+1}(\mathbf{A}) \xrightarrow{g_*} H_{n+1}(\mathbf{A}') \xrightarrow{\partial_*} H_n(\mathbf{A}'') \xrightarrow{f_*} H_n(\mathbf{A}) \xrightarrow{g_*} \cdots$ where $(\partial_*)_{n+1}: H_{n+1}(\mathbf{A}') \to H_n(\mathbf{A}'')$ is obtained via the Snake Lemma. 510

25.50, 25.51. Given a map $f: M \to N$ of modules, a resolution $\mathbf{A}$ of N, and a projective resolution $\mathbf{P}$ of M, one can lift f to a chain map $\mathbf{f}: \mathbf{P} \to \mathbf{A}$ that is unique up to homotopy equivalence. Consequently, any two projective resolutions of a module M are homotopy equivalent. 513, 514

25.54. A right exact covariant functor F is exact iff $L_1 F = 0$, in which case $L_n F = 0$ for all n. 515

25.58. The direct sum $\bigoplus M_i$ of right modules is flat iff each M_i is flat. 516

25.59. Every projective module P is flat. 516

25.67 (Shapiro's Lemma). $H_n(G, M_L^G) \cong H_n(L, M)$ for each L-module M and all n; $H^n(G, \mathrm{Coind}_L^G(M)) \cong H^n(L, M)$ for all n. 521

25A.8. An R-module M is a generator in R-**Mod** iff $T(M) = R$. 527

25A.14. If R and R' are Morita equivalent rings, then there is an R-progenerator P such that $R' \cong (\mathrm{End}_R P)\,\mathrm{op}$. 529

25A.19 (Morita's Theorem). Two rings R, R' are Morita equivalent iff there is an R-progenerator M such that $R' \cong (\mathrm{End}_R M)^{\mathrm{op}}$; in this case the categorical equivalence R-**Mod** $\to R'$-**Mod** is given by $M^* \otimes_R \_$. 531

E25B.6. If $H^2(R, \_) = 0$ and R has a nilpotent ideal N such that R/N is separable, then R has a subalgebra $S \cong R/N$ that is a complement to N as a C-module. 591

E25B.11. $0 \to M^R \to M \to \mathrm{Deriv}_C(R, M) \to \mathrm{Ext}^1_{R^e}(R, M) \to 0$ is an exact sequence. 592

E25B.12. Equivalent conditions for an algebra to be separable, in terms of derivations. 593

E25B.14 (Braun). A C-algebra R is Azumaya, iff there are $a_i, b_i \in R$ such that $\sum a_i b_i = 1$ and $\sum a_i R b_i \subseteq C$. 593

E25B.17. Any Azumaya algebra is a finite direct product of algebras of constant rank when the base ring has no nontrivial idempotents. 593

Chapter 26.

26.21 (The Fundamental Theorem of Hopf Modules). Any Hopf module M is isomorphic to $H \otimes M^{\mathrm{co}\,H}$ as Hopf modules (the latter under the "trivial action" $h'(h \otimes a) = (h'h \otimes a)$). 558

26.28 (Nichols-Zoeller [**NiZ**]). If K is a Hopf subalgebra of a f.d. Hopf algebra H, then H is free as a K-module, and $\dim K \mid \dim H$. 561

26.30. A f.d. Hopf algebra H is semisimple iff $\varepsilon(\int_H^l) \neq 0$. 562

E26.3, E26.5. For any algebra A and coalgebra C, $\mathrm{Hom}(C, A)$ becomes an algebra under the convolution product $(*)$. If H is a Hopf algebra, then its antipode S is the inverse to 1_H in $\mathrm{Hom}(H, H)$ under the convolution product. $S(ab) = S(b)S(a)$, $\Delta \circ S = \tau \circ (S \otimes S) \circ \Delta$, and $\epsilon \circ S = \epsilon$. 594

E26.16 (Fundamental Theorem of Comodules). Any finite subset of a comodule M (over a coalgebra C) is contained in a finite-dimensional subcomodule of M. 595

E26.17 (Fundamental Theorem of Coalgebras). Any finite subset of a coalgebra C is contained in a f.d. subcoalgebra of C. 595

E26.28. The following equations hold for $R = \sum a_i \otimes b_i$ in a quasi-triangular Hopf algebra: $R^{-1} = \sum S(a_i) \otimes b_i$; $\sum \epsilon(a_i) b_i = \sum a_i \epsilon(b_i) = 1$; $(S \otimes S)(R) = R$. 596

E26.32. For any almost cocommutative Hopf algebra H with antipode S, there exists invertible $u \in H$ such that $uS(u)$ is central and S^2 is the inner automorphism given by conjugation with respect to u. 597

E26.38. The smash product naturally gives rise to a Morita context $(A\#H, A^H, A, A', \tau, \tau')$. 598

E26.40. The quantum groups of Examples 16A.3 are Hopf algebras. 598

Bibliography

[Ab] Abe E., *Hopf Algebras*, Cambridge University Press, 1980.

[Am1] Amitsur S.A., *Doctoral dissertation*, Hebrew University, 1950.

[Am2] ———, *Generic splitting fields of central simple algebras*, Ann. Math. **62** (1955), 8–43.

[Am3] ———, *Algebras over infinite fields*, Proc. Amer. Math. Soc. **7** (1956), 35–48.

[Am4] ———, *On central division algebras*, Israel J. Math. **12** (1972), 408–420.

[AmL] Amitsur S.A. and Levitzki J., *Minimal identities for algebras*, Proc. Amer. Math. Soc. **1** (1950), 449–463.

[AtM] Atiyah M.F. and MacDonald I.G., *Introduction to Commutative Algebra*, Addison-Wesley, 1969.

[Ba1] Bass H., *Big projective modules are free*, Ill. J. Math. **7** (1963), 24– 31.

[Ba2] ———, *Algebraic K-theory*, Benjamin, 1968.

[Ba3] ———, *The degree of polynomial growth of finitely generated nilpotent groups*, Proc. London Math Soc. **25(3)** (1972), 603–614.

[BeaCSS] Beauville A., Colliot-Thélène J.-L., Sansuc J.-J., and Swinnerton-Dyer P., *Variétés stablement rationnelles non rationnelles. (French) [Nonrational stably rational varieties]*, Ann. of Math. (2) **121** (1985), 283–318.

[BeiMM] Beidar K.I., Martindale W.S. III, and Mikhalev A.V., *Rings with Generalized Identities*, Pure and Applied Mathematics, vol. 196, Dekker, 1996.

[Bel] Belov-Kanel A., *Counterexamples to the Specht problem,*, Sb. Math. **191** (2000), 329–340.

[BelK] Belov-Kanel A. and Kontsevich M., *Automorphisms of the Weyl algebra*, preprint (2005).

[BelR] Belov-Kanel A. and Rowen L.H., *Polynomial Identities: A Combinatorial Approach*, A K Peters, Boston, 2005.

[Ben] Benkart G., *The Lie inner structure of associative rings*, J. Algebra **43** (1976), 561–584.

[Berg] Bergman G., *The diamond lemma for ring theory*, Advances Math. **29** (1978), 178–218.

[BernGP] Bernstein I.N., Gel′fand I.M., and Ponomarev V.A., *Coxeter functors and Gabriel's theorem*, Russian Math. Survey **28** (1973), 17–32.

[Bl] Block R.E., *The irreducible representations of the Lie algebra sl(2) and of the Weyl algebra*, Advances Math. **39** (1981), 69–110.

[Boo] Boone W.W., *The word problem*, Ann. of Math. **70** (1959), 207–265.

[BooCL] Boone W.W., Cannonito F.B., and Lyndon R.C., *Word Problems: Decision Problems and the Burnside Problem in Group Theory*, Studies in Logic and the Foundations of Mathematics, vol. 71, North-Holland, Amsterdam, 1973.

[Bor] Borel A., *Linear Algebraic Groups*, Mathematics Lecture Notes, Benjamin, New York, 1969.

[Bou] Bourbaki N., *Commutative Algebra*, Hermann, 1972.

[BreG] Breuillard E. and Gelanter T., *A topological Tits alternative*, Ann. of Math. (2006).

[Bridg] Bridges D.S., *Computability – A Mathematical Sketchbook*, Springer, Berlin, 1994.

[BridsH] Bridson M.R. and Haefliger A., *Metric Spaces of Non-positive Curvature*, Grundlehren der mathematischen Wissenschaften, vol. 319, Springer, 1999.

[BrokaG] Brown K.A. and Goodearl K.R., *Lectures on Algebraic Quantum Groups*, Advanced Courses in Mathematics, CRM Barcelona, Birkhäuser, 2002.

[Broks] Brown K.S., *Cohomology of Groups*, Graduate Texts in Mathematics, vol. 87, Springer, 1982.

[Bur1] Burnside W., *On an unsettled question in the theory of discontinuous groups*, Quart. J. Pure Appl. Math. **33** (1902), 230–238.

[Bur2] ———, *On criteria for the finiteness of the order of a group*, Proc. London Math. Soc. **3** (1905), 435–448.

[CarE] Cartan H. and Eilenberg S., *Homological Algebra*, Princeton University Press, Princeton, New Jersey, 1956.

[Ch] Chatelet F., *Variations sur un thème de H. Poincaré*, Ann. Ecol. Norm. **59** (1944), 249–300.

[CoheGW] Cohen M., Gelaki S., and Westreich S., *Hopf Algebras*, Handbook of Algebra (M. Hazewinkel, ed.), vol. 4, Elsevier, 2007, pp. 173–239.

[Cohn1] Cohn P.M., *Algebra I*, Wiley, 1974.

[Cohn2] ———, *Algebra II, III*, Wiley, 1988.

[CuR] Curtis C. and Reiner I., *Representation Theory of Finite Groups and Associative Algebras*, Interscience, 1962.

[DasNR] Dascalescu S., Nastasescu C., and Raianu S., *Hopf algebras – An Introduction*, Pure and Applied Mathematics, vol. 235, Dekker, 2001.

[Deh] Dehn M., *Über unendliche diskontinuierliche Gruppen*, Math. Ann. **71** (1912), 116–144.

[Di] Dixon J.D., *The Tits alternative*, preprint (1989).

[DonF] Donovan P. and Freislich M.R., *The representation theory of finite graphs and associated algebras*, vol. 5, Carleton Math. Lecture Notes, 1973.

[DorSS] Dorfman R., Samuelson P., and Solow R., *Linear Programming and Economic Analysis*, McGraw-Hill, 1958.

[Dr1] Drinfeld V.G., *Quantum groups*, Proc. Int. Cong. Math. Berkeley **1** (1986), 789–820.

[Dr2] Drinfeld V.G., *On almost cocommutative Hopf algebras,*, Leningrad Math. J. **1** (1990), 321-342.

[EtG1] Etingof P. and Gelaki S., *Semisimple Hopf algebras of dimension pq are trivial*, J. Algebra **210** (1998), 664–669.

[EtG2] ———, *The classification of finite-dimensional triangular Hopf algebras over an algebraically closed field of characteristic 0. (English. English, Russian summary)*, Mosc. Math. J. **3** (2003), 37–43, 258.

[FaL] Farkas D. and Letzter G., *Ring theory from symplectic geometry*, J. Pure Appl. Algebra **125** (1998), 155–190.

[Fo] Formanek E., *The Polynomial Identities and Invariants of $n \times n$ Matrices*, CBMS, vol. 78, Amer. Math. Soc., Providence, R.I., 1991.

[FoP] Formanek E. and Procesi C., *Mumford's conjecture for the general linear group*, Advances Math. **19** (1976), 292–305.

[Ga1] Gabriel P., *Des categories abeliennes*, Bull. Soc. Math. France **90** (1962), 323–448.

[Ga2] ———, *Unzerlegbare Darstellungen I.*, Manuscripta Math. **6** (1972), 71–103, 309.

[GaR] Gabriel P. and Roiter A.V., *Representations of Finite Dimensional Algebras*, Springer, 1997.

[GeS] Gerstenhaber M. and Schack S., *Algebraic cohomology and deformation theory*, Deformation Theory of Algebras and Structures and Applications, vol. 247, Kluwer, 1988, pp. 11–265.

[Go] Golod E.S., *On nil algebras and residually finite p-groups*, Izv. Akad. Nauk. SSR **28** (1964), 273–276.

[GreNW] Greene C., Nijenhuis A., and Wilf H., *A probabilistic proof of a formula for the number of Young tableaux of a given shape*, Adv. in Math. **31** (1979), 104–109.

[Gri1] Grigorchuk R.I., *The Burnside problem on periodic groups*, Funct. Anal. Appl. **14(1)** (1980), 53–54.

[Gri2] ———, *On Milnor's problem of group growth*, Soviet Math. Dokl. **28** (1983), 23–26.

[Grob] Gröbner W., *Uber die algebraischen Eigenschaften der Integrale von linearen Differentialgleichungen mit konstanten Koeffizienten*, Monatsh. der Math. **47** (1939), 247–284.

[Grom1] Gromov M., *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Pub. Math. IHES **53** (1981), 53–78.

[Grom2] ———, *Hyperbolic groups*, Essays in Group Theory, MSRI Publications (S.M. Gersten, ed.), vol. 8, Springer-Verlag, 1987, pp. 75–265.

[Haa] Haar A., *Der Massbegriff in der Theorie der kontinuierlichen Gruppen*, Annals of Math. **34** (1933), 147–169.

[Halm] Hall M., *The Theory of Groups*, Macmillan, New York, 1959.

[HalmT] Hall M. and Tibor R., *On Schreier systems in free groups*, Trans. Amer. Math. Soc. **64** (1948), 386–408.

[Halp] Hall P., *Some word problems*, J. London Math. Soc. **33** (1958), 482–496.

[HalpH] Hall P. and Higman G., *On the p-length of p-soluble groups, and reduction theorems for Burnside's problem*, Proc. London Math. Soc. **6** (1956), 1–40.

[Har] Hartshorne R., *Algebraic Geometry*, Graduate Texts in Mathematics, vol. 52, Springer, New York, 1993.

[Haz] Hazewinkel M. (ed.), *Handbook of Algebra*, vol. 4, Elsevier, 2007.

[He] Helling H., *Eine Kennzeichnung von Charakteren auf Gruppen und Assoziativen Algebren*, Comm. Alg. **1** (1974), 491–501.

[Hers1] Herstein I., *Topics in Algebra*, Chicago Lectures in Mathematics, Xerox, 1964.

[Hers2] ———, *Topics in Ring Theory*, Chicago Lectures in Mathematics, Univ. Chicago Press, 1969.

[Hers3] ———, *Rings with Involution*, Chicago Lectures in Mathematics, Univ. Chicago Press, 1976.

[Hof] Hofstadter D.R., *Bach, Goedel, and Escher – An Eternal Golden Braid*, Vintage Books, 1980.

[Hum1] Humphreys J., *Introduction to Lie Algebras and Representation Theory*, Springer, New York, 1972.

[Hum2] ———, *Linear Algebraic Groups*, Springer Lecture Notes in Mathematics, vol. 21, Springer, New York, 1975.

[Hum3] ———, *Reflection Groups and Coxeter Groups*, Cambridge Studies in Advanced Mathematics, vol. 29, Cambridge, 1990.

[Jac1] Jacobson N., *Lie Algebras*, Wiley, New York, 1962.

[Jac2] ———, *Structure and Representations of Jordan Algebras*, Colloq. Pub., vol. 39, Amer. Math. Soc., Providence RI, 1968.

[Jac3] ———, *Exceptional Lie Algebras*, Lecture Notes in Pure and Applied Math., vol. 1, Marcel Dekker, New York, 1971.

[Jac4] ———, *Basic Algebras I*, Freeman, San Francisco, 1974.

[Jac5] ———, *Basic Algebras II*, Freeman, San Francisco, 1980.

[Jac6] ———, *Structure Theory of Jordan Algebras*, Lecture Notes in Math., vol. 5, U. of Arkansas, Fayettesville, 1981.

[Jac7] ———, *Finite-dimensional division algebras over fields.*, Springer-Verlag, Berlin, 1996.

[Jan] Jantzen J.C., *Lectures on Quantum Groups*, Graduate Studies in Mathematics, vol. 6, Amer. Math. Soc., Providence, RI, 1996.

[Ji] Jimbo M., *A q-difference analog of $U(\mathfrak{g})$ and the Yang-Baxter equation*, Lett. Mat. Phys. **10** (1985), 63–69.

[Kacg] Kac G.I., *Certain arithmetic properties of ring groups*, Funct. Anal. Appl. **6** (1972), 158–160.

[Kacv1] Kac V.G., *Infinite-dimensional Lie algebras and Dedekind's η-function*, Funct. Anal. Appl. **8** (1974), 68–70.

[Kacv2] ———, *Infinite root systems, representations of graphs, and invariant theory*, Invent. Math. **56** (1980), 57–92.

[Kacv3] ———, *Infinite root systems, representations of graphs, and invariant theory*, J. Algebra **78** (1982), 141–162.

[Kacv4] ———, *Infinite-dimensional Lie Algebras, Third edition*, Cambridge University Press, 1990.

[Kan] Kantor I., *Graded Lie algebras. (Russian)*, Trudy Sem. Vektor. Tenzor. Anal. **15** (1970), 227–266.

[Kapl1] Kaplansky I., *Lie Algebras and Locally Compact Groups*, Chicago Lectures in Mathematics, Univ. Chicago Press, 1974.

[Kapl2] Kaplansky I., *Bialgebras*, University of Chicago Lecture Notes, 1975.

[KapoM] Kapovich I. and Myasnikov A., *Stallings foldings and subgroups of free groups*, J. Algebra **248** (2002), 608–668.

[Ke] Kemer A.R., *Finite basability of identities of associative algebras (English translation)*, Algebra and Logic **26** (1987), 362–397.

[Ki] Kirillov A.A., *Elements of the Theory of Representations*, Springer, 1976.

[KnuMRT] Knus M., Merkurjev A., Rost M., and Tignol J.-P., *The Book of Involutions*, American Mathematical Society Colloquium Publications, vol. 44, Amer. Math. Soc., Providence, RI, 1998.

[KnuO] Knus M. and Ojanguren M., *Théorie de la descente et algèbres d'Azumaya*, Lecture Notes in Mathematics, vol. 389, Springer, Berlin, 1974.

[Ko] Kostrikin A.I., *Around Burnside*; transl. by J. Wiegold, Springer, 1990.

[KrafR] Kraft H. and Riedtmann C., *Geometry of representations of quivers*, Representations of Algebras, London Math. Soc. Lecture Notes in Math., vol. 116, 1985, pp. 109–145.

[KrauL] Krause G.R. and Lenagan T.H., *Growth of Algebras and Gelfand-Kirillov Dimension*, Graduate Studies in Mathematics, vol. 22, Amer. Math. Soc., 2000.

[Lam] ———, *A First Course in Noncommutative Rings. Second edition*, Graduate Texts in Mathematics, vol. 131, Springer, 2001.

[Lan] ———, *Algebra*, Addison-Wesley, 1965.

[LenS] Lenagan, T. H. and Smoktunowicz, Agata, *An infinite dimensional affine nil algebra with finite Gelfand-Kirillov dimension*, J. Amer. Math. Soc. **20** (2007), 989–1001.

[Li] Lin V.Ya., *Artin braids and the groups and spaces connected with them*, J. Soviet Math. **18** (1982), 736–788.

[LuS] Lubotzky A. and Segal D., *Subgroup Growth*, Progress in Mathematics, vol. 212, Birkhauser, 2003.

[Mal] Mal′cev A.I., *Algorithms and Recursive Functions*, Wolters–Noordhorff, 1970.

[McCR] McConnell J.C. and Robson J.C., *Noncommutative Noetherian Rings*, Graduate Studies in Math., vol. 30, Amer. Math. Soc., 2001.

[McK] McKinnie K., *Prime to p extensions of the generic abelian crossed product.*, J. Algebra **317** (2007), 813–832.

[MeS] Merkurjev A.S. and Suslin A.A., *K-cohomology of Severi-Brauer varieties and norm residue homomorphisms.*, Izv. Akad. Nauk. USSR **46** (1982), 1011–1046.

[Mi] Milnor J., *Growth of finitely generated solvable groups*, J. Diff. Geom. **2** (1968), 447-449.

[Mo1] Montgomery S., *Hopf Algebras and their Actions on Rings*, CBMS Regional Conference Series in Mathematics, vol. 82, Amer. Math. Soc., 1993.

[Mo2] ———, *Classifying finite dimensional Hopf algebras*, Contemp. Math. **229** (1998), Amer. Math. Soc., 265–279.

[Naz] Nazarova L.A., *Representations of quivers of infinite type*, Math. USSR Izvestija **37** (1973), 752–791.

[Ne] Newman, M. H. A., *On theories with a combinatorial definition of "equivalence"*, Ann. of Math. (2) **43** (1942), 223–243.

[Ng] Ng Siu-Hung, *Non-semisimple Hopf algebras of dimension p^2*, J. Algebra **255** (2002), 182–197.

[NiZ] Nichols W.D. and Zoeller M.B., *A Hopf algebra freeness theorem*, Amer. J. Math. **1111** (1989), 381–385.

[Nie] Nielsen J., *Om Regning med ikke-kommutative Faktorer og dens Anvendelse i Gruppeteorien*, Matematisk Tidsskrift B (1921), 77–94.

[Pak] Pak I., *Hook length formula and geometric combinatorics*, Sem. Lothar. Combin. **46** (2001), 1–4.

[Pas] Passman D., *The Algebraic Structure of Group Algebras*, Wiley, 1977.

[Pie] Pierce C., *Associative Algebras*, Springer, 1982.

[Pr] Procesi C., *The invariant theory of $n \times n$ matrices*, Advances in Math. **19** (1976), 306–381.

[Q] Quillen D., *Projective modules over polynomial rings*, Inv. Math. **36** (1976), 167–171.

[Ra1] Razmyslov Yu.P., *On a problem of Kaplansky*, Math. USSR Izv. **7** (1972), 479–496.

[Ra2] Razmyslov Yu.P., *Trace identities of full matrix algebras over a field of characteristic zero*, Math. USSR Izv. **8** (1974), 724–760.

[Re] Reiten I., *Dynkin diagrams and the representation theory of algebras*, Notices Amer. Math. Soc. **44** (1997), 546–556.

[Ro] Rogers H. Jr., *Theory of recursive functions and effective Computability, Second edition*, MIT Press, Cambridge, 1987.

[Rol] Rolfson D., *New developments in the theory of Artin's braid groups*, Topology Appl. **127** (2003), 77–90.

[Rot1] Rotman J., *The Theory of Groups: An Introduction (second edition)*, Allyn and Bacon, 1973.

[Rot2] Rotman J., *An Introduction to Homological Algebra*, Academic Press, 1979.

[Row1] Rowen L.H., *Polynomial Identities in Ring Theory*, Pure and Applied Math, vol. 84, Academic Press, 1980.

[Row2] ———, *Ring Theory I, II*, Pure and Applied Math, vols. 127,128, Academic Press, 1988.

[Row3] ———, *Algebra: Group, Rings, and Fields*, A K Peters, 1994.

[Sag] Sagan B., *The Symmetric Group. Representations, Combinatorial Algorithms, and Symmetric Functions. Second edition*, Springer Graduate Texts in Mathematics,
Springer, 2001.

[Sal1] Saltman D., *Noncrossed product p-algebras and Galois extensions*, J. Algebra **52** (1978), 302–314.

[Sal2] ———, *The Brauer group is torsion*, Proc. Amer. Math. Soc. **81** (1981), 385–387.

[Sal3] ———, *Generic Galois extensions and problems in field theory*, J. Algebra **43** (1982), 250–283.

[Sal4] ———, *Division algebras over p-adic curves*, J. Ramanujan Math. Soc. **12** (1997), 25–47.

[Sal5] ———, *Lectures on Division Algebras*, CBMS Regional Conference Series in Mathematics, vol. 94, Amer. Math. Soc., 1999.

[Sam] Samelson H., *Notes on Lie Algebras*, Mathematical studies, vol. 23, van Nostrand, 1969.

[Scha] Schafer R., *An Introduction to Nonassociative Algebras*, Pure and Applied Math, vol. 22, Academic Press, 1966.

[Scho] Schofield A., *Representations of Rings over Skew Fields*, London Math. Soc. Lecture Notes, vol. 92, Cambridge U. Press, 1985.

[Ser1] Serre J.-P., *A Course in Arithmetic*, Springer, 1973.

[Ser2] ———, *Linear Representations of Finite Groups*, Graduate Texts in Mathematics, vol. 42, Springer, 1973.

[Ser3] ———, *Galois Cohomology*, Springer; transl. by Patrick Ion, 1997.

[ShiW] Shirvani M. and Wehrfritz B., *Skew Linear Groups*, Graduate Texts in Mathematics, vol. 118, Cambridge Univ. Press, 1986.

[ShnS] Shnider S. and Sternberg S., *Quantum Groups: From Coalgebras to Drinfeld Algebras, A Guided Tour*, International Press, 1993.

[SmaW] Small L. and Warfield R.B., *Affine algebras of Gelfand-Kirillov dimension one are PI*, J. Algebra **91** (1984), 386–389.

[SmoV] Smoktunowicz A. and Vishne U., *An affine prime non-semiprimitive algebra with quadratic growth*, preprint.

[Sp] Springer T.A., *Linear Algebraic Groups (Second edition)*, Progress in Mathematics, vol. 9, Birkhäuser, Boston, 1998.

[St] Stefan D., *The set of types of n-dimensional semisimple and cosemisimple Hopf algebras*, J. Algebra **193** (1997), 571–580.

[Su] Suslin A., *Projective modules over polynomial rings are free*, Dokl. Akad. Nauk USSR **229** (1976), 1063–1066.

[Swa] ———, *Gubeladze's proof of Anderson's conjecture*, Cont. Math **124** (1992), 215–250.

[Swe] Sweedler M.E., *Hopf Algebras*, Benjamin, 1969.

[Tig] Tignol J.P., *Galois Theory of Algebraic Equations*, World Scientific, 1988.

[Tit] Tits J., *Free subgroups in linear groups*, J. Algebra **20** (1972), 250–270.

[Tse] Tsen C., *Zur Stufentheorie der Quasi-algebraisch-Abgeschlossenheit kommutativer Korper*, J. Chinese Math. Soc. **1** (1936).

[Tsu] Tsuchimoto Y., *Endomorphisms of Weyl algebras and p-curvature*, Osaka J. Math. **42** (2005), 435–452.

[vdDW] van den Dries L. and Wilkie A.J., *On Gromov's theorem concerning groups of polynomial growth and elementary logic*, J. Algebra **89** (1984), 349–374.

[Vinb] Vinberg E.B., *Linear Representations of Groups*, Birkhäuser, 1989.

[VN] Von Neumann J., *Invariant Measures*, Amer. Math. Soc., 1999.

[Wei] Weibel C., *An Introduction to Homological Algebra*, Cambridge Studies in Advanced Mathematics, vol. 38, Cambridge, 1994.

[Wo] Wolf J., *Growth of finitely generated solvable groups and curvature of Riemannian manifolds*, J. Diff. Geom. **2** (1968), 421–446.

[Ze1] Zelmanov Y., *The solution of the restricted Burnside problem for groups of odd exponent*, Math. USSR Izv. **36** (1991), 41–60.

[Ze2] ———, *The solution of the restricted Burnside problem for groups of prime power*, mimeographed notes, Yale University (1991).

[Ze3] ———, *The solution of the restricted Burnside problem for 2-groups*, Math. USSR Sb. **72** (1992), 543–565.

[ZheSSS] Zhevlakov K.A., Slin'ko A.M., Shestakov I.P., and Shirshov A.I., *Rings that are Nearly Associative*, Pure and Applied Mathematics, vol. 104, Academic Press, 1982.

[Zhu] Zhu Y., *Quantum double construction of quasitriangular Hopf algebras and Kaplansky's conjecture*, preprint (1997).

List of Names

Index

Titles in This Series

97 **David C. Ullrich,** Complex made simple, 2008

96 **N. V. Krylov,** Lectures on elliptic and parabolic equations in Sobolev spaces, 2008

95 **Leon A. Takhtajan,** Quantum mechanics for mathematicians, 2008

94 **James E. Humphreys,** Representations of semisimple Lie algebras in the BGG category $\mathcal{O}$, 2008

93 **Peter W. Michor,** Topics in differential geometry, 2008

92 **I. Martin Isaacs,** Finite group theory, 2008

91 **Louis Halle Rowen,** Graduate algebra: Noncommutative view, 2008

90 **Larry J. Gerstein,** Basic quadratic forms, 2008

89 **Anthony Bonato,** A course on the web graph, 2008

88 **Nathanial P. Brown and Narutaka Ozawa,** C*-algebras and finite-dimensional approximations, 2008

87 **Srikanth B. Iyengar, Graham J. Leuschke, Anton Leykin, Claudia Miller, Ezra Miller, Anurag K. Singh, and Uli Walther,** Twenty-four hours of local cohomology, 2007

86 **Yulij Ilyashenko and Sergei Yakovenko,** Lectures on analytic differential equations, 2007

85 **John M. Alongi and Gail S. Nelson,** Recurrence and topology, 2007

84 **Charalambos D. Aliprantis and Rabee Tourky,** Cones and duality, 2007

83 **Wolfgang Ebeling,** Functions of several complex variables and their singularities (translated by Philip G. Spain), 2007

82 **Serge Alinhac and Patrick Gérard,** Pseudo-differential operators and the Nash–Moser theorem (translated by Stephen S. Wilson), 2007

81 **V. V. Prasolov,** Elements of homology theory, 2007

80 **Davar Khoshnevisan,** Probability, 2007

79 **William Stein,** Modular forms, a computational approach (with an appendix by Paul E. Gunnells), 2007

78 **Harry Dym,** Linear algebra in action, 2007

77 **Bennett Chow, Peng Lu, and Lei Ni,** Hamilton's Ricci flow, 2006

76 **Michael E. Taylor,** Measure theory and integration, 2006

75 **Peter D. Miller,** Applied asymptotic analysis, 2006

74 **V. V. Prasolov,** Elements of combinatorial and differential topology, 2006

73 **Louis Halle Rowen,** Graduate algebra: Commutative view, 2006

72 **R. J. Williams,** Introduction the the mathematics of finance, 2006

71 **S. P. Novikov and I. A. Taimanov,** Modern geometric structures and fields, 2006

70 **Seán Dineen,** Probability theory in finance, 2005

69 **Sebastián Montiel and Antonio Ros,** Curves and surfaces, 2005

68 **Luis Caffarelli and Sandro Salsa,** A geometric approach to free boundary problems, 2005

67 **T.Y. Lam,** Introduction to quadratic forms over fields, 2004

66 **Yuli Eidelman, Vitali Milman, and Antonis Tsolomitis,** Functional analysis, An introduction, 2004

65 **S. Ramanan,** Global calculus, 2004

64 **A. A. Kirillov,** Lectures on the orbit method, 2004

63 **Steven Dale Cutkosky,** Resolution of singularities, 2004

62 **T. W. Körner,** A companion to analysis: A second first and first second course in analysis, 2004

61 **Thomas A. Ivey and J. M. Landsberg,** Cartan for beginners: Differential geometry via moving frames and exterior differential systems, 2003

60 **Alberto Candel and Lawrence Conlon,** Foliations II, 2003

59 **Steven H. Weintraub,** Representation theory of finite groups: algebra and arithmetic, 2003

58 **Cédric Villani,** Topics in optimal transportation, 2003

57 **Robert Plato,** Concise numerical mathematics, 2003

56 **E. B. Vinberg,** A course in algebra, 2003

55 **C. Herbert Clemens,** A scrapbook of complex curve theory, second edition, 2003

54 **Alexander Barvinok,** A course in convexity, 2002

53 **Henryk Iwaniec,** Spectral methods of automorphic forms, 2002

52 **Ilka Agricola and Thomas Friedrich,** Global analysis: Differential forms in analysis, geometry and physics, 2002

51 **Y. A. Abramovich and C. D. Aliprantis,** Problems in operator theory, 2002

50 **Y. A. Abramovich and C. D. Aliprantis,** An invitation to operator theory, 2002

49 **John R. Harper,** Secondary cohomology operations, 2002

48 **Y. Eliashberg and N. Mishachev,** Introduction to the h-principle, 2002

47 **A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi,** Classical and quantum computation, 2002

46 **Joseph L. Taylor,** Several complex variables with connections to algebraic geometry and Lie groups, 2002

45 **Inder K. Rana,** An introduction to measure and integration, second edition, 2002

44 **Jim Agler and John E. M^{c}Carthy,** Pick interpolation and Hilbert function spaces, 2002

43 **N. V. Krylov,** Introduction to the theory of random processes, 2002

42 **Jin Hong and Seok-Jin Kang,** Introduction to quantum groups and crystal bases, 2002

41 **Georgi V. Smirnov,** Introduction to the theory of differential inclusions, 2002

40 **Robert E. Greene and Steven G. Krantz,** Function theory of one complex variable, third edition, 2006

39 **Larry C. Grove,** Classical groups and geometric algebra, 2002

38 **Elton P. Hsu,** Stochastic analysis on manifolds, 2002

37 **Hershel M. Farkas and Irwin Kra,** Theta constants, Riemann surfaces and the modular group, 2001

36 **Martin Schechter,** Principles of functional analysis, second edition, 2002

35 **James F. Davis and Paul Kirk,** Lecture notes in algebraic topology, 2001

34 **Sigurdur Helgason,** Differential geometry, Lie groups, and symmetric spaces, 2001

33 **Dmitri Burago, Yuri Burago, and Sergei Ivanov,** A course in metric geometry, 2001

32 **Robert G. Bartle,** A modern theory of integration, 2001

31 **Ralf Korn and Elke Korn,** Option pricing and portfolio optimization: Modern methods of financial mathematics, 2001

30 **J. C. McConnell and J. C. Robson,** Noncommutative Noetherian rings, 2001

29 **Javier Duoandikoetxea,** Fourier analysis, 2001

28 **Liviu I. Nicolaescu,** Notes on Seiberg-Witten theory, 2000

27 **Thierry Aubin,** A course in differential geometry, 2001

26 **Rolf Berndt,** An introduction to symplectic geometry, 2001

For a complete list of titles in this series, visit the
AMS Bookstore at **www.ams.org/bookstore/**.

Photo by Elihu Walber

This book is a companion volume to *Graduate Algebra: Commutative View* (published as volume 73 in this series). The main and most important feature of the book is that it presents a unified approach to many important topics, such as group theory, ring theory, Lie algebras, and gives conceptual proofs of many basic results of noncommutative algebra. There are also a number of major results in noncommutative algebra that are usually found only in technical works, such as Zelmanov's proof of the restricted Burnside problem in group theory, word problems in groups, Tits's alternative in algebraic groups, PI algebras, and many of the roles that Coxeter diagrams play in algebra.

The first half of the book can serve as a one-semester course on noncommutative algebra, whereas the remaining part of the book describes some of the major directions of research in the past 100 years. The main text is extended through several appendices, which permits the inclusion of more advanced material, and numerous exercises. The only prerequisite for using the book is an undergraduate course in algebra; whenever necessary, results are quoted from *Graduate Algebra: Commutative View*.

ISBN 978-0-8218-4153-2

9 780821 841532

GSM/91

For additional information and updates on this book, visit

www.ams.org/bookpages/gsm-91